

FUJITSU Storage
ETERNUS AX series All-Flash Arrays,
ETERNUS HX series Hybrid Arrays

ONTAP FlexGroup Volumes Best Practices and Implementation Guide



Table of Contents

| | |
|--|-----------|
| 1. The Evolution of NAS in ONTAP | 14 |
| Flexible volumes: A tried-and-true solution | 14 |
| FlexGroup volumes: An evolution of NAS | 15 |
| 2. Terminology | 16 |
| What are large files? | 17 |
| 3. ONTAP FlexGroup Advantages..... | 18 |
| Massive capacity and predictable low latency for high-metadata workloads | 18 |
| Efficient use of all cluster hardware | 18 |
| Simple, easy-to-manage architecture and balancing | 18 |
| Superior density for big data | 18 |
| 4. Use Cases..... | 19 |
| Ideal use cases | 19 |
| Nonideal cases | 19 |
| FlexGroup volume use case examples | 19 |
| FlexGroup use case example #1: Active IQ infrastructure | 19 |
| FlexGroup use case example #2: Back up repository for SQL Server | 20 |
| Conclusion | 22 |
| 5. FlexGroup Feature Support and Maximums..... | 23 |
| Behavior of unsupported SMB features | 25 |
| Maximums and minimums | 26 |
| 6. Deciding whether FlexGroup Volumes Are the Right Fit | 27 |
| Scale-out performance | 27 |
| Feature compatibility limitations | 28 |
| Simplifying performance | 29 |
| AFF A700 testing | 29 |
| FlexGroup performance with big data workloads | 31 |
| Automatic workload adaptation | 32 |
| Ingest algorithm improvements | 32 |

| | |
|--|-----------|
| Performance features | 33 |
| Workloads and behaviors | 35 |
| Optimal workloads | 36 |
| Good workloads | 37 |
| Nonideal workloads – large files | 37 |
| Performance expectations: read-heavy workloads | 41 |
| Data imbalances in FlexGroup volumes | 41 |
| 7. Initial FlexGroup Design Considerations..... | 43 |
| Cluster considerations | 43 |
| ONTAP version considerations | 43 |
| Failure domains | 44 |
| Aggregate layout considerations | 44 |
| Deploying a FlexGroup volume on aggregates with existing FlexVol volumes | 45 |
| Flash Cache and Flash Pool | 46 |
| Advanced disk partitioning | 47 |
| SyncMirror (mirrored aggregates) | 47 |
| MetroCluster | 47 |
| Cloud Volumes ONTAP | 47 |
| Capacity considerations | 48 |
| Maximums and minimums | 48 |
| Theoretical or absolute maximums | 49 |
| FlexVol member volume layout considerations | 50 |
| When would clients experience out of space errors? | 55 |
| When would I need to manually create a FlexGroup volume? | 56 |
| Aggregate free space considerations | 61 |
| Initial volume size considerations | 61 |
| Snapshot copies and snapshot reserve | 62 |
| Volume autosize (autogrow and autoshrink) | 64 |
| Elastic sizing | 67 |
| Proactive resizing | 71 |
| Networking considerations | 76 |
| LACP considerations | 77 |
| DNS load-balancing considerations | 78 |
| Border Gateway Protocol (BGP) | 78 |
| Security and access control list style considerations | 78 |
| Basic volume security style guidance | 79 |
| 8. FlexGroup Administration Considerations..... | 81 |
| Viewing FlexGroup volumes | 81 |
| ONTAP System Manager | 81 |

| | |
|--|------------|
| Active IQ Unified Manager | 82 |
| Command Line | 85 |
| Viewing FlexGroup volume capacity | 86 |
| Total FlexGroup capacity | 86 |
| Overprovisioning or Thin Provisioning in a FlexGroup Volume | 86 |
| Adding capacity to a FlexGroup volume | 87 |
| Recommendations for adding capacity | 87 |
| Volume Expand | 90 |
| Adding disks, aggregates, and nodes | 91 |
| Removing or evacuating nodes from a cluster | 92 |
| Nondisruptive volume move considerations | 93 |
| When to use nondisruptive volume moves | 94 |
| Using nondisruptive volume moves | 94 |
| Considerations when deleting FlexGroup volumes | 95 |
| Volume rename considerations | 95 |
| 9. Qtrees | 96 |
| Qtrees and file moves | 96 |
| Qtree IDs and rename behavior | 96 |
| File handle effects for qtree exports | 97 |
| Managing quotas with FlexGroup | 97 |
| User and group quota considerations | 98 |
| Creating a user or group quota | 98 |
| Creating a tree reporting quota from the command line | 100 |
| Quota enforcement example | 101 |
| Performance effect of using quotas | 103 |
| Quota scan completion times | 104 |
| User-mapping considerations with quotas | 105 |
| Tree quota considerations | 105 |
| How clients see space when quotas are enabled | 107 |
| 10. General NAS and High-File-Count Considerations | 108 |
| High file count considerations | 108 |
| Default and maximum inode counts | 108 |
| Increasing maximum files: considerations | 108 |
| Default and maximum inode counts – FlexGroup volume considerations | 109 |
| High file counts, low-capacity needs | 110 |
| Planning for high file counts in ONTAP | 110 |
| Viewing used and total inodes | 111 |
| What happens when you run out of inodes | 111 |
| Async delete | 111 |
| 64-bit file identifiers | 112 |
| What happens when I modify this option? | 113 |
| NFSv3 versus NFSv4.x: File IDs | 113 |

| | |
|--|------------|
| Using quota enforcement to limit file count | 114 |
| ONTAP System Manager: 9.7 | 115 |
| ONTAP System Manager: 9.8 and later | 115 |
| Effect of file ID collision | 115 |
| Effects of file system ID changes in ONTAP | 116 |
| How FSIDs operate with SVMs in high-file-count environments | 117 |
| How FSIDs operate with Snapshot copies | 117 |
| Directory size considerations: maxdirsize | 117 |
| What directory structures can affect maxdirsize? | 118 |
| How flat directory structures can affect FlexGroup volumes | 119 |
| Querying for used maxdirsize values | 119 |
| Number of files that can fit into a single directory with the default maxdirsize | 121 |
| Event management system messages sent when maxdirsize is exceeded | 121 |
| Effect of increasing the maxdirsize value | 121 |
| Do FlexGroup volumes bypass maxdirsize limitations? | 121 |
| Effect of exceeding maxdirsize | 122 |
| File system analytics | 122 |
| Special character considerations | 123 |
| Support for utf8mb4 volume language | 124 |
| Managing slow directory listings via NFS in high-file-count environments | 125 |
| File deletions/FlexGroup member volume balancing | 126 |
| Rebalancing data within a FlexGroup volume | 126 |
| Listing files when a member volume is out of space | 128 |
| File rename considerations | 128 |
| Symlink considerations | 128 |
| NFS version considerations | 128 |
| Network connection concurrency: NFSv3 | 129 |
| NFS write appends | 131 |
| Nconnect | 131 |
| Mapping NFS connected clients to volume names | 131 |
| Enabling and using NFSv4.x with FlexGroup volumes | 132 |
| NAS metadata effect in a FlexGroup volume | 136 |
| CIFS/SMB considerations | 136 |
| SMB version considerations | 136 |
| Use of change notifications with SMB | 136 |
| Large MTU | 137 |
| SMB multichannel | 137 |
| Continuously available shares (CA shares) | 137 |
| Other considerations | 138 |
| Virtualization workload considerations | 138 |
| ONTAP tools for VMware vSphere support (formerly Virtual Storage Console) | 138 |
| Copy offload | 139 |
| Considerations | 139 |

| | |
|---|------------|
| Databases on FlexGroup volumes | 140 |
| FlexCache volume considerations | 141 |
| FlexClone | 141 |
| FlexClone to different storage virtual machine (SVM) | 142 |
| Volume rehost | 143 |
| FlexClone deletion | 143 |
| 11. Encryption At-Rest..... | 144 |
| Rekeying a FlexGroup volume or encrypting existing FlexGroup volumes | 144 |
| Drive-level encryption (NSE and SED) | 144 |
| 12. FlexGroup Sample Designs | 145 |
| Volume affinity and CPU saturation | 145 |
| FlexGroup sample design 1: FlexGroup volume, entire cluster (24 nodes) | 146 |
| FlexGroup sample design 2: multiple nodes, aggregates, partial cluster | 147 |
| FlexGroup sample design 3: FlexGroup, single node | 148 |
| FlexGroup sample design 4: FlexGroup volumes mounted to FlexGroup volumes | 153 |
| FlexVol Volumes Mounted to FlexGroup Volumes | 153 |
| 13. General Troubleshooting and Remediation | 155 |
| Failure scenarios | 155 |
| Storage failovers | 155 |
| Network failures | 155 |
| Snapshot failures | 155 |
| Hardware failures | 155 |
| Time synchronization | 156 |
| 14. Capacity Monitoring and Alerting..... | 157 |
| Capacity monitoring and alerting with the command line | 157 |
| Event management system messages | 157 |
| Client-side capacity considerations with thin provisioning | 159 |
| Windows capacity reporting | 160 |
| Viewing FlexVol member capacity from the ONTAP command line | 160 |
| FlexGroup capacity viewer | 161 |
| Logical space accounting | 163 |
| Monitoring FlexGroup performance | 163 |
| Monitoring performance from the command line | 163 |
| flexgroup show | 166 |
| Performance archiver | 167 |
| Monitoring performance (Active IQ Unified Manager) | 167 |

| | |
|--|------------|
| 15. FlexGroup Data Protection Best Practices..... | 168 |
| 16. Migrating to ONTAP FlexGroup | 169 |
| Migration using NDMP | 169 |
| FlexVol to FlexGroup conversion | 170 |
| Why convert a FlexVol volume to a FlexGroup volume? | 170 |
| When not to convert a FlexVol volume | 170 |
| How it works | 172 |
| Other considerations and caveats | 172 |
| Migrating from third-party storage to FlexGroup | 175 |
| Migrating from Data ONTAP operating in 7-Mode | 175 |
| Migrating from SAN LUNs in ONTAP | 175 |
| XCP Migration Tool | 175 |
| Using XCP to scan files before migration | 176 |
| Using XCP to run disk usage (du) scans | 179 |
| 17. Examples | 180 |
| Thin provisioning example | 180 |
| Volume Autosize example | 181 |
| Snapshot spill example | 182 |
| Capacity monitoring and alerting examples in Active IQ Unified Manager | 185 |
| Editing volume thresholds in Active IQ Unified Manager | 188 |
| Inode monitoring | 189 |
| Active IQ "Fix It" | 190 |
| Sample FlexVol to FlexGroup conversion | 190 |
| Converting FlexVols in existing SnapMirror relationships – example | 195 |
| Sample FlexVol to FlexGroup conversion – 500 million files | 199 |
| Event management system examples | 203 |
| Inode-related EMS examples | 203 |
| Example of maxdirsize message | 204 |
| Examples of capacity-related event management system messages | 205 |
| 18. Command Examples..... | 207 |
| FlexGroup capacity commands | 207 |
| Example of statistics show-periodic command for entire cluster | 211 |
| Real-time SVM-level statistics show-periodic for NFSv3 read and write operations | 211 |
| Real-time FlexGroup local and remote statistics | 212 |

| | |
|---|-----|
| Example of creating a FlexGroup volume and specifying fewer member volumes than the default value | 212 |
| Sample REST API for creating a FlexGroup volume | 213 |
| Example of increasing a FlexGroup volume's size | 217 |
| Example of expanding a FlexGroup volume | 218 |
| Other command-line examples | 219 |
| Creating a FlexGroup volume by using flexgroup deploy | 219 |
| Creating a FlexGroup volume across multiple nodes by using volume create | 219 |
| Modifying the FlexGroup Snapshot policy | 219 |
| Applying storage QoS | 219 |
| Applying volume autogrow | 219 |

List of Figures

| | | |
|-----------|--|----|
| Figure 1 | FlexVol design with junctioned architecture for >100TB capacity | 14 |
| Figure 2 | What is a large file? | 17 |
| Figure 3 | SQL Server backup environment | 20 |
| Figure 4 | Throughput and total operations during test runs | 21 |
| Figure 5 | CPOC scale-out throughput results | 22 |
| Figure 6 | FlexGroup volume | 27 |
| Figure 7 | FlexVol volume versus FlexGroup volume—maximum throughput trends under increasing workload | 30 |
| Figure 8 | FlexVol volume versus FlexGroup volume—maximum throughput trends under increasing workload, detailed | 30 |
| Figure 9 | FlexVol volume versus FlexGroup volume—maximum average total IOPS | 31 |
| Figure 10 | TeraSort benchmark statistics summary on a FlexGroup volume | 31 |
| Figure 11 | Storage QoS on FlexGroup volumes—single-node connection | 33 |
| Figure 12 | Storage QoS on FlexGroup volumes: multinode connection | 34 |
| Figure 13 | Qtree QoS use cases | 34 |
| Figure 14 | Capacity imbalance and likelihood of remote placement | 36 |
| Figure 15 | FlexGroup volume with a few large files; why usage can be suboptimal | 38 |
| Figure 16 | Potential worst case scenario for large files; all land in the same member volume | 39 |
| Figure 17 | Capacity imbalance example | 40 |
| Figure 18 | How FlexVol capacity can affect FlexGroup load distribution | 45 |
| Figure 19 | FlexVol and FlexGroup architecture comparison | 50 |
| Figure 20 | ONTAP System Manager FlexGroup volume creation | 53 |
| Figure 21 | Error when creating a FlexGroup volume beyond the allowed maximum in System Manager | 57 |
| Figure 22 | How capacity is divided among member volumes | 58 |
| Figure 23 | Effect of larger files in a FlexGroup member volume | 59 |
| Figure 24 | Fewer, larger member volumes | 60 |
| Figure 25 | FlexGroup volumes—member sizes versus FlexGroup volume capacity | 62 |
| Figure 26 | Member volume size allocation after a volume autosize operation | 66 |
| Figure 27 | File write behavior before elastic sizing | 68 |
| Figure 28 | File write behavior after elastic sizing | 69 |
| Figure 29 | Initial FlexGroup data balance – proactive resize, autosize disabled | 72 |
| Figure 30 | FlexGroup data balance, ~68% used – proactive resize, autosize disabled | 72 |
| Figure 31 | FlexGroup data balance, job complete – proactive resize, autosize disabled | 72 |
| Figure 32 | FlexGroup data balance, new large file – proactive resize, autosize disabled | 73 |
| Figure 33 | FlexGroup data balance, 80GB file – proactive resize, autosize disabled | 73 |
| Figure 34 | FlexGroup data balance, out of space – proactive resize, autosize disabled | 73 |
| Figure 35 | Initial FlexGroup data balance – proactive resize, autosize enabled | 74 |
| Figure 36 | FlexGroup data balance, ~68% used – proactive resize, autosize enabled | 74 |
| Figure 37 | FlexGroup data balance, job complete – proactive resize, autosize enabled | 75 |
| Figure 38 | FlexGroup data balance, second test run – proactive resize, autosize enabled | 75 |
| Figure 39 | FlexGroup data balance, autosize limit – proactive resize, autosize enabled | 76 |
| Figure 40 | Modifying FlexGroup volume security styles in ONTAP System Manager | 80 |
| Figure 41 | ONTAP System Manager FlexGroup volume view | 82 |
| Figure 42 | Active IQ Unified Manager; FlexGroup capacity view | 82 |
| Figure 43 | Active IQ Unified Manager Capacity Trend | 83 |
| Figure 44 | Active IQ Performance Manager FlexGroup volume view | 84 |
| Figure 45 | Member volume performance chart | 84 |
| Figure 46 | Member volume graphs | 85 |
| Figure 47 | Capacity effect when thin-provisioned FlexGroup volumes exist with space-guaranteed FlexVol volumes | 87 |
| Figure 48 | Adding aggregates with FlexGroup volumes | 91 |

| | | |
|-----------|---|-----|
| Figure 49 | Adding nodes and expanding the FlexGroup volume | 92 |
| Figure 50 | Removing nodes that contain FlexGroup member volumes | 93 |
| Figure 51 | Quota reports – ONTAP System Manager | 98 |
| Figure 52 | Quota volume status – ONTAP System Manager | 99 |
| Figure 53 | Quota rules – ONTAP System Manager | 99 |
| Figure 54 | ONTAP 9.5 performance (operations/sec)–quotas on and off | 103 |
| Figure 55 | ONTAP 9.5 performance (MBps)–quotas on and off | 104 |
| Figure 56 | 64-bit File IDs in ONTAP System Manager 9.8 | 115 |
| Figure 57 | File System Analytics – enable | 122 |
| Figure 58 | File System Analytics – directory and file information | 123 |
| Figure 59 | File System Analytics – inactive and active data | 123 |
| Figure 60 | Capacity imbalance after deletion of larger files | 126 |
| Figure 61 | Effect of RPC slot tables on NFSv3 performance | 130 |
| Figure 62 | pNFS diagram | 133 |
| Figure 63 | pNFS operations diagram | 134 |
| Figure 64 | pNFS operations diagram–FlexGroup volumes | 135 |
| Figure 65 | ONTAP tools for VMware vSphere Support – FlexGroup Datastores | 139 |
| Figure 66 | FlexGroup volume, entire cluster (24 nodes) | 146 |
| Figure 67 | Multiple nodes, partial cluster | 147 |
| Figure 68 | Git clone completion times comparison | 149 |
| Figure 69 | Average and maximum throughput comparison | 149 |
| Figure 70 | Maximum read throughput comparison | 150 |
| Figure 71 | Maximum write throughput comparison | 150 |
| Figure 72 | Total average IOPS comparison | 151 |
| Figure 73 | Average CPU utilization, throughput, and IOPS for a FlexGroup volume–ETERNUS AX4100 HA pair, 128 threads | 151 |
| Figure 74 | Average CPU utilization, throughput, and IOPS for a FlexGroup volume–single-node ETERNUS AX4100, 128 threads | 152 |
| Figure 75 | FlexGroup volume, single node | 152 |
| Figure 76 | FlexGroup volume mounted to FlexGroup volume | 153 |
| Figure 77 | Google Sheet – FlexGroup capacity view | 161 |
| Figure 78 | Google Sheet – FlexGroup member capacity view | 161 |
| Figure 79 | Google Sheet – Average inode size | 162 |
| Figure 80 | How logical space accounting works | 163 |
| Figure 81 | Active IQ Performance Manager graphs | 167 |
| Figure 82 | Converting a FlexVol volume that is nearly full and at maximum capacity | 171 |
| Figure 83 | Converting a FlexVol volume to a FlexGroup and adding member volumes | 172 |
| Figure 84 | XCP reporting graphs | 176 |
| Figure 85 | XCP report | 178 |
| Figure 86 | FlexGroup capacity breakdown–Active IQ Unified Manager | 180 |
| Figure 87 | Editing volume thresholds | 188 |
| Figure 88 | Active IQ Unified Manager – fix out of inodes | 190 |
| Figure 89 | Sample statistics from conversion process | 200 |
| Figure 90 | Sample statistics during conversion process – Adding member volumes | 201 |
| Figure 91 | Sample statistics of conversion process – two times performance | 202 |

List of Tables

| | | |
|----------|--|-----|
| Table 1 | General ONTAP feature support | 23 |
| Table 2 | General NAS protocol version support | 24 |
| Table 3 | Unsupported SMB2.x and 3.x features | 25 |
| Table 4 | How unsupported SMB features behave with FlexGroup volumes | 25 |
| Table 5 | FlexGroup maximums | 26 |
| Table 6 | FlexGroup minimums | 26 |
| Table 7 | ONTAP volume family comparison | 28 |
| Table 8 | Best practices for aggregate layout with FlexGroup volumes..... | 44 |
| Table 9 | FlexGroup maximums | 48 |
| Table 10 | FlexGroup minimums | 49 |
| Table 11 | Theoretical maximums for FlexGroup based on allowed volume count in ONTAP..... | 50 |
| Table 12 | Situations in which you see out of space errors | 55 |
| Table 13 | Capacity Management Decision Matrix | 60 |
| Table 14 | Autosize maximum size examples | 76 |
| Table 15 | Inode defaults and maximums according to FlexVol size | 108 |
| Table 16 | Inode defaults resulting from FlexGroup member sizes and member volume counts..... | 109 |
| Table 17 | High-file-count/small capacity footprint examples—increasing member volume counts..... | 110 |
| Table 18 | Async-delete performance | 111 |
| Table 19 | nconnect performance results | 131 |
| Table 20 | flexgroup show output column definitions | 166 |

Preface

This document provides a brief overview of ONTAP FlexGroup and a set of best practices and implementation tips to use with this feature. The FlexGroup feature is an evolution of scale-out NAS containers that blends nearly infinite capacity with predictable, low- latency performance in metadata-heavy workloads. For information about FlexGroup volumes that is not covered in this document, contact Fujitsu Support, and we will add information to this document as necessary.

Copyright 2021 FUJITSU LIMITED

First Edition
September 2021

Trademarks

Third-party trademark information related to this product is available at:
<https://www.fujitsu.com/global/products/computing/storage/eternus/trademarks.html>

Trademark symbols such as ™ and ® are omitted in this document.

About This Manual

Intended Audience

This manual is intended for system administrators who configure and manage operations of the ETERNUS AX/HX, or field engineers who perform maintenance. Refer to this manual as required.

Related Information and Documents

The latest information for the ETERNUS AX/HX is available at:
<https://www.fujitsu.com/global/support/products/computing/storage/manuals-list.html>

Document Conventions

■ Notice Symbols

The following notice symbols are used in this manual:

**Caution**

Indicates information that you need to observe when using the ETERNUS AX/HX. Make sure to read the information.

**Note**

Indicates information and suggestions that supplement the descriptions included in this manual.

1. The Evolution of NAS in ONTAP

As hard-drive costs are driven down and flash hard-drive capacity grows exponentially, file systems are following suit. The days of file systems that number in the tens of gigabytes or even terabytes are over. Storage administrators face increasing demands from application owners for large buckets of capacity with enterprise-level performance.

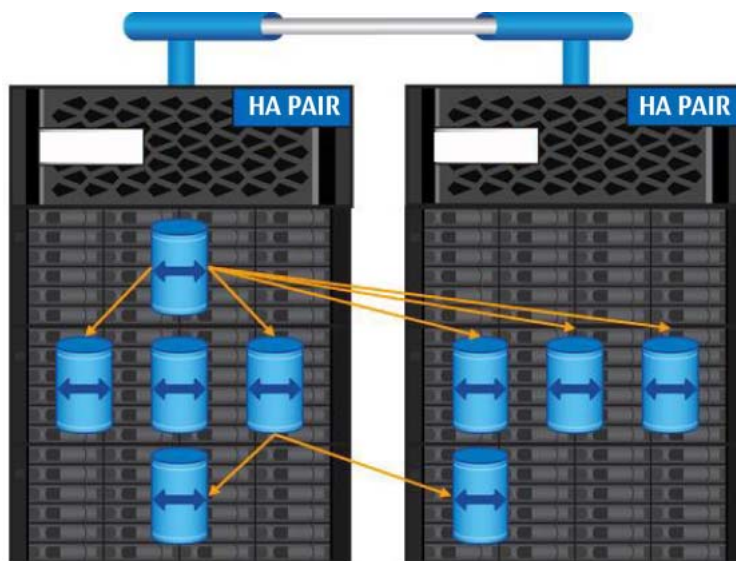
Machine learning and artificial intelligence workloads involve storage needs for a single namespace that can extend into the petabyte range (with billions of files). With the rise in these technologies, along with the advent of big data frameworks such as Hadoop, the evolution of NAS file systems is overdue. ONTAP FlexGroup is the ideal solution for these types of architectures.

Flexible volumes: A tried-and-true solution

The concept of FlexVol is to take a storage file system and virtualize it across a hardware construct to provide flexible storage administration in an ever-changing data center.

FlexVol volumes could be grown or shrunk nondisruptively and be allocated to the storage operating system as thin-provisioned containers to enable overprovisioning of storage systems. Doing so allowed storage administrators the freedom to allocate space as consumers demanded it.

Figure 1 FlexVol design with junctioned architecture for >100TB capacity



However, as data grew, file systems needed to grow. FlexVol can handle most storage needs with its 100TB capacity, and Data ONTAP provided a clustered architecture that those volumes could work with.

Before FlexGroup, ONTAP administrators could create junction paths to attach FlexVol volumes to one another. In this way, they created a file system on the cluster that could act as a single namespace.

[Figure 1](#) shows an example of what a FlexVol volume junction design for a large namespace would look like.

Although this architecture worked for many environments, it was awkward to manage and did not give a "single-bucket" approach to the namespace, where the FlexVol volume's capacity and file count constraints are limiting factors.

FlexGroup volumes: An evolution of NAS

With FlexGroup volumes, a storage administrator can easily provision a massive single namespace in a matter of seconds. FlexGroup volumes have virtually no capacity or file count constraints outside of the physical limits of hardware or the total volume limits of ONTAP. Limits are determined by the overall number of constituent member volumes that work in collaboration to dynamically balance load and space allocation evenly across all members. There is no required maintenance or management overhead with a FlexGroup volume. You simply create the FlexGroup volume and share it with your NAS clients. ONTAP does the rest.

2. Terminology

Terminology specific to ONTAP FlexGroup is covered in the following list.

- **Constituent/member volumes**

In a FlexGroup context, "constituent volume" and "member volume" are interchangeable terms. They refer to the underlying FlexVol volumes that make up a FlexGroup volume and provide the capacity and performance gains that are achieved only with a FlexGroup volume.

- **FlexGroup volume**

A FlexGroup volume is a single namespace that is made up of multiple constituent/member volumes. It is managed by storage administrators, and it acts like a FlexVol volume. Files in a FlexGroup volume are allocated to individual member volumes and are not striped across volumes or nodes.

- **Affinity**

Affinity describes the tying of a specific operation to a single thread.

- **Automated Incremental Recovery (AIR)**

Automated Incremental Recovery is an ONTAP subsystem that repairs FlexGroup inconsistencies dynamically, with no outage or administrator intervention required.

- **Ingest**

Ingest is the consumption of data by way of file or folder creations.

- **Junction paths**

Junction paths were used to provide capacity beyond a FlexVol volume's 100TB limit prior to the simplicity and scale-out of FlexGroup. Junction paths join multiple FlexVol volumes together to scale out across a cluster and provide multiple volume affinities. The use of a junction path in ONTAP is known as "mounting" the volume within the ONTAP namespace.

- **Large files**

Refer to ["What are large files?"](#).

- **Overprovisioning and thin provisioning**

Overprovisioning (or thin provisioning) storage is the practice of disabling a volume's space guarantee (`guarantee = none`). This practice allows the virtual space allocation of the FlexVol volume to exceed the physical limits of the aggregate that it resides on. For example, with overprovisioning, a FlexVol volume can be 100TB on an aggregate that has a physical size of only 10TB. Overprovisioning allows storage administrators to grow volumes to large sizes to avoid the need to grow them later, but it does present the management overhead of needing to monitor available space closely.

In overprovisioned volumes, the available space reflects the actual physical available space in the aggregate. Therefore, the usage percentage and capacity available values might seem off a bit. However, they simply reflect a calculation of the actual space that is available when compared with the virtual space that is available in the FlexVol volume. For a more accurate portrayal of space allocation when using overprovisioning, use the `aggregate show-space` command.

- **Remote access layer (RAL)**

The remote access layer (RAL) is a feature in the WAFL system that allows a FlexGroup volume to balance ingest workloads across multiple FlexGroup constituents or members.

- **Remote hard links**

Remote hard links are the building blocks of FlexGroup. These links act as normal hard links but are unique to ONTAP. The links allow a FlexGroup volume to balance workloads across multiple remote members or constituents. In this case, "remote" simply means "not in the parent volume." A remote hard link can be another FlexVol member on the same aggregate or node.

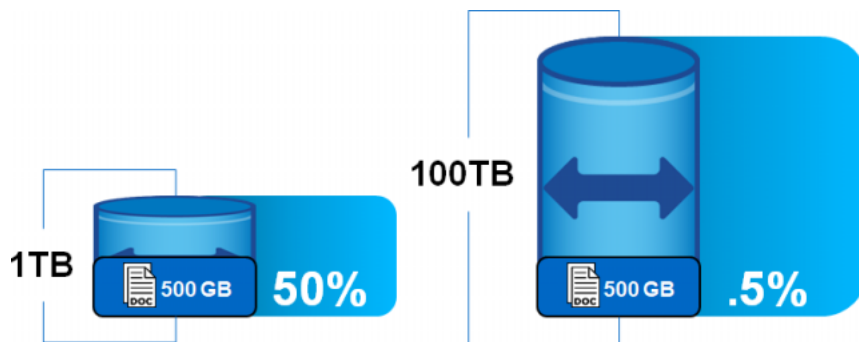
What are large files?

This document uses the term "large file" liberally. Therefore, it's important to define exactly what a "large file" is in the context of FlexGroup.

A FlexGroup volume operates optimally when a workload is ingesting numerous small files, because FlexGroup volumes maximize the system resources to address those specific workloads that might bottleneck because of serial processing in a FlexVol volume. FlexGroup volumes also work well with various other workloads (as defined in "[Use Cases](#)"). One type of workload that historically created problems in a FlexGroup volume, however, is a workload with larger files or files that grow over time, such as database files.

In a FlexGroup volume, a large file is a product of the percentage of allocated space, not of any specific file size. Thus, in some FlexGroup configurations—for example, in which the member volume size is only 1TB—a "large file" might be 500GB (50% of the member volume size). In other configurations, for example, in which the member volume size is 100TB, that same 500GB file size would only take up 0.5% of the volume capacity.

Figure 2 What is a large file?



This type of file could be large enough to throw off the ingest heuristics in the FlexGroup volume, or it could potentially create problems later when the member volume gets closer to full. ONTAP has no visibility into how large a file will get when it's created, so it doesn't know how to prioritize larger files. Instead, it works reactively to balance data to other member volumes with free space discrepancies.

Each ONTAP release is constantly improving the approach to large files in FlexGroup volumes.

- ONTAP 9.7 introduced ingest algorithm changes to help balance large files and/or datasets with mixed file sizes.
- ONTAP 9.8 brings capacity management simplicity by way of Proactive resizing.

3. ONTAP FlexGroup Advantages

ONTAP FlexGroup provides various advantages for different workloads. The advantages are described in the following sections.

Massive capacity and predictable low latency for high-metadata workloads

FlexGroup volumes offer a way for storage administrators to easily provision massive amounts of capacity with the ability to nondisruptively scale out that capacity. FlexGroup also enables parallel performance for high metadata workloads that can increase throughput and total operations while still providing low latency for mission-critical workloads.

Efficient use of all cluster hardware

FlexGroup volumes allow storage administrators to easily span multiple physical aggregates and nodes with member FlexVol volumes, while maintaining a true single namespace for applications and users to dump data into. Although clients and users see the space as monolithic, ONTAP is working behind the scenes to distribute the incoming file creations evenly across the FlexGroup volume to provide efficient CPU and disk utilization.

Simple, easy-to-manage architecture and balancing

To make massive capacity easy to deploy, Fujitsu lets you manage FlexGroup volumes like FlexVol volumes. ONTAP handles the underlying member volume creation and balance across the cluster nodes and provides a single access point for NAS shares.

Superior density for big data

A FlexGroup volume enables you to condense large amounts of data into smaller data center footprints by way of the superb storage efficiency features of ONTAP, including the following:

- Thin provisioning
- Data compaction
- Data compression
- Deduplication

In addition, ONTAP supports large SSDs, which can deliver massive amounts of raw capacity in a single 24-drive shelf enclosure. It is possible to get petabytes of raw capacity in just 10U of rack space, which cuts costs on cooling, power consumption, and rack rental space and offers excellent density in the storage environment. These features, combined with a FlexGroup volume's ability to efficiently use that capacity and balance performance across a cluster, give you a solution that was made for big data.

4. Use Cases

The ONTAP FlexGroup design is most beneficial in specific use cases (electronic design and automation, software development, and so on).

Ideal use cases

A FlexGroup volume works best with workloads that are heavy on ingest (a high level of new data creation), heavily concurrent, and evenly distributed among subdirectories:

- Electronic design automation (EDA)
- Artificial intelligence and machine learning log file repositories
- Software build/test environments (such as GIT)
- Seismic/oil and gas
- Media asset or HIPAA archives
- File streaming workflows (such as video surveillance)
- Unstructured NAS data (such as home directories)
- Big data and data lakes (Hadoop with the NFS connector)
- Virtualized workloads (ONTAP 9.8 and later)

Nonideal cases

Some workloads are currently not recommended for FlexGroup volumes. These workloads include:

- Workloads that require file striping (large files spanning multiple nodes or volumes)
- Workloads that require specific control over the layout of the relationships of data to FlexVol volumes
- Workloads with a large amount of file renames
- Workloads with millions of files in a single directory that require frequent scans of all of the files
- Workloads with thousands of symlinks
- Workloads that require specific features and functionality that are not currently available with FlexGroup volumes

If you have questions, feel free to contact Fujitsu Support.

FlexGroup volume use case examples

The following sections describe two examples of real-world use cases.

FlexGroup use case example #1: Active IQ infrastructure

For details about the solution, contact Fujitsu Support.

FlexGroup use case example #2: Back up repository for SQL Server

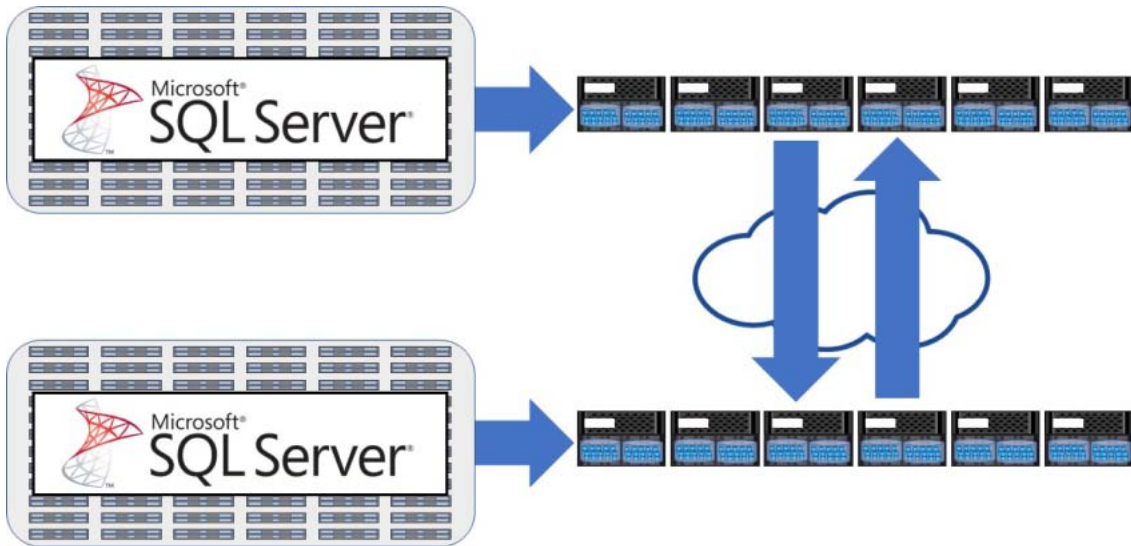
In this environment, the customer wanted to perform compressed backups of 5,000 Microsoft SQL Servers over SMB. This test was done with approximately 200 servers to vet out the solution, with a slow ramp up over the course of a few months.

But this database isn't only a backup target—it will also be replicated to a disaster recovery site by using SnapMirror for extra data protection.

Each site has a six-node ETERNUS HX series cluster running ONTAP using 6TB near-line SAS (NL-SAS) encrypted drives. Each cluster holds 3PB of usable capacity. The clusters use 30 FlexGroup volumes and qtrees within the volumes for data organization.

The FlexGroup volumes are 64TB each and the member volumes are 2.6TB each, with four members per node across six nodes (24 total members per FlexGroup volume).

Figure 3 SQL Server backup environment



■ The results

This customer needed a single namespace that could collect ~150TB worth of MSSQL backup data over a 12-hour period. That's ~12TB per hour at ~3.5GB per second.

During testing, we used 222 servers at site A and 171 servers at site B. During the test, each cluster's CPU was at 95% utilization and the backup jobs (sequential writes) were able to accomplish 8.4GB per second, which is ~2.4 times the amount of throughput the job needed. At this rate, the backups could complete in approximately 5 hours, rather than the 12-hour window. Also, this SMB workload performed approximately 120,000 IOPS. When more clients are added to this workload, we expect the throughput to max out at around 9GBps.

Figure 4 Throughput and total operations during test runs

| cpu avg | cpu busy | total ops | nfs-ops | cifs-ops | fscache ops | spin-ops | total recv | total data sent | data busy | data recv | cluster sent | cluster busy | cluster recv | cluster sent | disk read | disk write | pkts recv | pkts sent |
|------------|-------------|--------------|---------|----------|----------------|----------|---------------|--------------------|--------------|--------------|-----------------|-----------------|-----------------|-----------------|--------------|---------------|--------------|--------------|
| 56% | 81% | 54530 | 0 | 54530 | 0 | 54420 | 6.16GB | 2.65GB | 44% | 3.34GB | 28.3MB | 22% | 2.82GB | 2.62GB | 128MB | 3.31GB | 968237 | 898917 |
| 65% | 78% | 70482 | 0 | 70482 | 0 | 70467 | 8.03GB | 3.44GB | 47% | 4.33GB | 30.9MB | 24% | 3.70GB | 3.41GB | 114MB | 4.79GB | 1178768 | 1102812 |
| 74% | 87% | 88725 | 0 | 88725 | 0 | 88195 | 10.2GB | 4.30GB | 49% | 5.44GB | 37.1MB | 26% | 4.78GB | 4.26GB | 157MB | 5.56GB | 1389743 | 1324559 |
| 86% | 92% | 111577 | 0 | 111577 | 0 | 110569 | 12.8GB | 5.88GB | 53% | 6.84GB | 41.9MB | 31% | 6.00GB | 5.84GB | 153MB | 6.77GB | 1726469 | 1679506 |
| 88% | 92% | 115036 | 0 | 115036 | 0 | 113599 | 13.2GB | 6.44GB | 51% | 7.06GB | 45.9MB | 29% | 6.14GB | 6.40GB | 142MB | 7.65GB | 1845740 | 1814549 |
| 92% | 95% | 118148 | 0 | 118148 | 0 | 117104 | 13.6GB | 6.11GB | 45% | 7.26GB | 49.9MB | 42% | 6.34GB | 6.07GB | 149MB | 8.11GB | 1802929 | 1769902 |
| 95% | 98% | 122953 | 0 | 122953 | 0 | 122123 | 14.3GB | 7.10GB | 47% | 7.54GB | 45.9MB | 43% | 6.75GB | 7.06GB | 134MB | 8.29GB | 1978205 | 1952416 |
| 96% | 99% | 126241 | 0 | 126241 | 0 | 125104 | 14.6GB | 6.49GB | 53% | 7.75GB | 54.3MB | 44% | 6.80GB | 6.37GB | 138MB | 8.28GB | 1865375 | 1849777 |
| 95% | 97% | 121948 | 0 | 121948 | 0 | 120719 | 13.9GB | 7.29GB | 44% | 7.47GB | 47.3MB | 40% | 6.41GB | 7.20GB | 109MB | 8.30GB | 1995998 | 1947271 |
| 95% | 98% | 123079 | 0 | 123079 | 0 | 121113 | 13.9GB | 5.71GB | 41% | 7.56GB | 49.0MB | 38% | 6.37GB | 5.66GB | 129MB | 8.40GB | 1761097 | 1712061 |
| 95% | 97% | 120567 | 0 | 120567 | 0 | 120493 | 13.7GB | 7.01GB | 42% | 7.41GB | 47.6MB | 36% | 6.34GB | 6.96GB | 116MB | 8.49GB | 1888934 | 1882711 |
| 95% | 98% | 119573 | 0 | 119573 | 0 | 119458 | 13.6GB | 5.74GB | 37% | 7.33GB | 44.4MB | 35% | 6.28GB | 5.69GB | 111MB | 8.19GB | 1702969 | 1671363 |
| 95% | 97% | 119538 | 0 | 119538 | 0 | 119829 | 13.5GB | 6.98GB | 41% | 7.34GB | 46.2MB | 35% | 6.17GB | 6.93GB | 120MB | 8.44GB | 1880298 | 1873821 |
| 95% | 98% | 118119 | 0 | 118119 | 0 | 118373 | 13.4GB | 5.56GB | 37% | 7.25GB | 45.4MB | 37% | 6.17GB | 5.52GB | 118MB | 8.42GB | 1666066 | 1630785 |
| 95% | 98% | 118862 | 0 | 118862 | 0 | 118327 | 13.6GB | 6.29GB | 39% | 7.29GB | 47.1MB | 33% | 6.30GB | 6.24GB | 114MB | 8.31GB | 1784134 | 1759266 |
| 96% | 99% | 121039 | 0 | 121039 | 0 | 121136 | 13.7GB | 6.67GB | 38% | 7.44GB | 44.5MB | 34% | 6.21GB | 6.63GB | 120MB | 8.35GB | 1832520 | 1827158 |
| 96% | 99% | 120852 | 0 | 120852 | 0 | 120920 | 13.7GB | 5.77GB | 39% | 7.42GB | 47.5MB | 33% | 6.24GB | 5.72GB | 111MB | 8.31GB | 1706939 | 1678778 |
| 94% | 97% | 119819 | 0 | 119819 | 0 | 120129 | 13.7GB | 7.05GB | 41% | 7.36GB | 42.6MB | 35% | 6.29GB | 7.01GB | 118MB | 8.49GB | 1882656 | 1877381 |

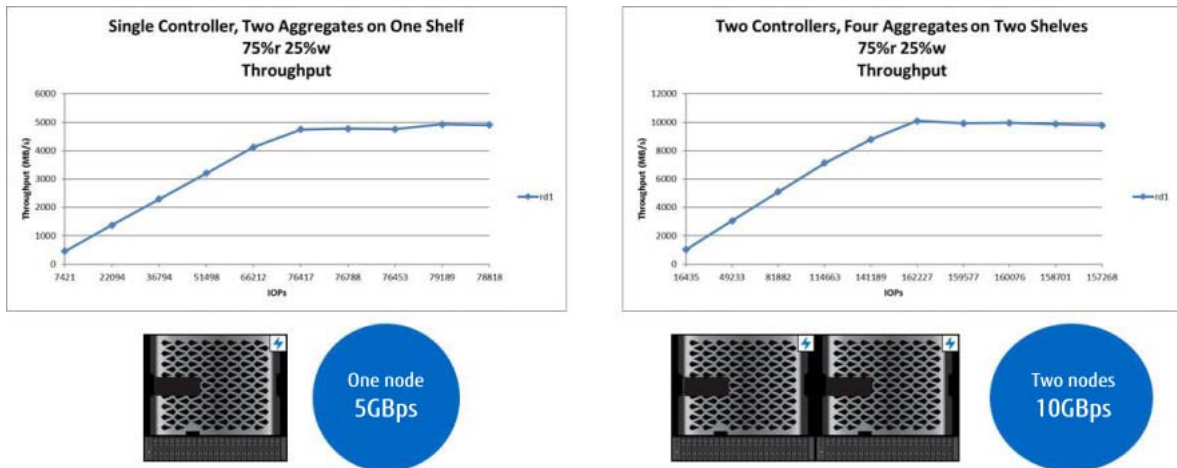
■ Data protection

In addition to the performance seen on the FlexGroup volume for the production workload, this customer was also able to achieve a high rate of transfer for the SnapMirror relationships between sites—8.4GB per second for the SnapMirror transfer. This rate means that the replication window for a 150TB dataset would be about 5.5 hours for the initial transfer. After that, the deltas should be able to complete well within the required transfer window, providing a solid disaster recovery plan for these MSSQL backups.

■ Scale-out performance

This six-node cluster was able to push over 8.4GB per second to a FlexGroup volume. In Customer Proof of Concept (CPOC) labs, we've seen near-linear performance gains by adding nodes to a cluster. The following graphs show throughput results for a single-node AFF A700 all-flash storage system and a two-node AFF A700.

Figure 5 CPOC scale-out throughput results



Note

If you want to add more performance to your backup workload, you can add more nodes.

Conclusion

Not only is a FlexGroup volume great for small or high-file-count workloads such as EDA and software builds, it also can handle high throughput requirements for larger streaming files. It also reduces backup windows by scaling out storage across multiple nodes and applies all your cluster resources while maintaining performance even with spinning drives.

5. FlexGroup Feature Support and Maximums

This chapter shows which ONTAP features are supported for use with FlexGroup volumes; it also notes the ONTAP version in which feature support was added. If a feature is not listed in this chapter, contact Fujitsu Support for information. For features specific to data protection, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays Data Protection and Backup for ONTAP FlexGroup Volumes](#).

Table 1 General ONTAP feature support

| Supported Feature | Version of ONTAP First Supported |
|---|----------------------------------|
| Snapshot copy technology | ONTAP 9.0 |
| SnapRestore software (FlexGroup level) | ONTAP 9.0 |
| Hybrid aggregates | ONTAP 9.0 |
| Constituent or member volume move | ONTAP 9.0 |
| Postprocess deduplication | ONTAP 9.0 |
| RAID-TEC technology | ONTAP 9.0 |
| Per-aggregate consistency point | ONTAP 9.0 |
| Sharing FlexGroup with FlexVol in the same SVM | ONTAP 9.0 |
| Active IQ Unified Manager | ONTAP 9.1 |
| Inline adaptive compression | ONTAP 9.1 |
| Inline deduplication | ONTAP 9.1 |
| Inline data compaction | ONTAP 9.1 |
| Thin provisioning | ONTAP 9.1 |
| AFF | ONTAP 9.1 |
| Quota reporting | ONTAP 9.1 |
| SnapMirror technology | ONTAP 9.1 |
| User and group quota reporting (no enforcement) | ONTAP 9.1 |
| Access Based Enumeration (ABE) for SMB shares | ONTAP 9.1 |
| Aggregate inline deduplication (cross-volume deduplication) | ONTAP 9.2 |
| Volume Encryption (VE) | ONTAP 9.2 |
| SnapVault technology | ONTAP 9.3 |
| Qtrees | ONTAP 9.3 |
| Automated deduplication schedules | ONTAP 9.3 |
| Version-independent SnapMirror and unified replication | ONTAP 9.3 |
| Antivirus scanning for SMB | ONTAP 9.3 |
| Volume autogrow | ONTAP 9.3 |
| QoS maximums/ceilings | ONTAP 9.3 |
| FlexGroup expansion without SnapMirror rebaseline | ONTAP 9.3 |
| Improved ingest heuristics | ONTAP 9.3 |
| SMB change/notify | ONTAP 9.3 |
| File audit | ONTAP 9.4 |
| FPolicy | ONTAP 9.4 |
| Adaptive QoS | ONTAP 9.4 |
| QoS minimums (AFF only) | ONTAP 9.4 |
| Relaxed SnapMirror limits | ONTAP 9.4 |
| SMB 3.x Multichannel | ONTAP 9.4 |
| FabricPool | ONTAP 9.5 |
| Quota enforcement | ONTAP 9.5 |

5. FlexGroup Feature Support and Maximums

| Supported Feature | Version of ONTAP First Supported |
|---|----------------------------------|
| Qtree statistics | ONTAP 9.5 |
| Inherited SMB watches and change notifications | ONTAP 9.5 |
| SMB copy offload (offloaded data transfer) | ONTAP 9.5 |
| Storage-Level Access Guard | ONTAP 9.5 |
| FlexCache (cache only; FlexGroup as origin supported in ONTAP 9.7) | ONTAP 9.5 |
| Volume rename | ONTAP 9.6 |
| Volume shrink | ONTAP 9.6 |
| MetroCluster | ONTAP 9.6 |
| Elastic sizing | ONTAP 9.6 |
| Continuously Available Shares (SMB)* * SQL Server and Hyper-V workloads only | ONTAP 9.6 |
| Aggregate Encryption (AE) | ONTAP 9.6 |
| Cloud Volumes ONTAP | ONTAP 9.6 |
| FlexClone | ONTAP 9.7 |
| In-place conversion of FlexVol to FlexGroup (refer to " Deploying a FlexGroup volume on aggregates with existing FlexVol volumes ".) | ONTAP 9.7 |
| vStorage APIs for Array Integration (VAAI) | ONTAP 9.7 |
| NDMP | ONTAP 9.7 |
| NFSv4.0 and NFSv4.1 (including parallel NFS, or pNFS) | ONTAP 9.7 |
| FlexGroup volumes as FlexCache origin volumes | ONTAP 9.7 |
| File cloning | ONTAP 9.8 |
| Proactive resizing | ONTAP 9.8 |
| NFSv4.2 (base protocol support) | ONTAP 9.8 |
| NDMP enhancements: EXCLUDE, RBE (Restartable Backup Extension), MULTI_SUBTREE_NAMES | ONTAP 9.8 |
| 1,023 Snapshots | ONTAP 9.8 |
| Qtree QoS | ONTAP 9.8 |

Table 2 General NAS protocol version support

| Supported NAS Protocol Version | Version of ONTAP First Supported |
|---------------------------------|----------------------------------|
| NFSv3 | ONTAP 9.0 |
| SMB2.1, SMB3.x | ONTAP 9.1 RC2 |
| NFSv4.0, NFSv4.1, pNFS | ONTAP 9.7 |
| NFSv4.2 (base protocol support) | ONTAP 9.8 |

Table 3 Unsupported SMB2.x and 3.x features

| Unsupported SMB2.x Features | Unsupported SMB 3.x Features |
|---|---|
| SMB Remote Volume Shadow Copy Service (VSS) | <ul style="list-style-type: none"> • VSS for SMB file shares. • SMB directory leasing • SMB direct or remote direct memory access (RDMA) |
| | <p>Note</p> <p>SMB 3.0 encryption is supported.</p> |

Note

Remote VSS is not the same as the SMB Previous Versions tab. Remote VSS is application-aware Snapshot functionality and is most commonly used with Hyper-V workloads. FlexGroup volumes have supported the SMB Previous Versions tab since it was introduced.

Behavior of unsupported SMB features

Usually, if an SMB feature is unsupported in ONTAP, it simply does not work. With ONTAP FlexGroup, there are some considerations regarding unsupported SMB features and functionality.

Table 4 How unsupported SMB features behave with FlexGroup volumes

| Feature | Behavior with FlexGroup Volumes |
|---|---|
| SMBv1.0 | Access fails or is denied for any shares accessing with SMB 1.0. This can affect Windows XP, Windows 2003, and office equipment such as scanners or copiers that attempt to connect to the NAS with SMB. |
| Change notification/SMB watches | <p>For details on change notifications, refer to "Use of change notifications with SMB".</p> <p>Note</p> <p>ONTAP 9.7 and later offers parallel processing of change notification operations for a better overall experience. If you wish to use change notifications, use ONTAP 9.7 and later for the best results.</p> |
| Offloaded data transfer (ODX) | For more information about ODX, see this TechNet article . |
| Remote Volume Shadow Copy Service (VSS) | There is no warning; Remote VSS just does not work. The effect should be low because the primary use case for Remote VSS is with Hyper-V, which is not a recommended workload for FlexGroup volumes. For more information about Remote VSS, see this TechNet article . |

Maximums and minimums

This section covers the maximums and minimums that are specific to ONTAP FlexGroup volumes. [Table 5](#) lists the maximum values and shows whether the maximum is hard-coded/enforced or a recommended or tested value.

Table 5 FlexGroup maximums

| | Value | Value Type |
|---|----------------------|---------------------|
| FlexGroup volume size | 20PB | Tested/recommended* |
| FlexGroup total file count | 400 billion | Tested/recommended* |
| Cluster node count | 24 (12 HA pairs) | Hard-coded/enforced |
| FlexVol member volume size | 100TB | Hard-coded/enforced |
| FlexVol member volume file count | 2 billion | Hard-coded/enforced |
| SnapMirror volume count (member per FlexGroup) | 200 | Hard-coded/enforced |
| SnapMirror volume count (FlexGroup total per cluster) | 6,000 | Hard-coded/enforced |
| File size | 16TB | Hard-coded/enforced |
| FlexVol member constituent count | 200 | Tested/recommended* |
| Aggregate size/count | Same as ONTAP limits | Hard-coded/enforced |

Table 6 FlexGroup minimums

| | Value | Value Type |
|------------------------|------------|---------------------|
| FlexVol member size | 100GB | Tested/recommended* |
| Data aggregate count | 1 | Hard-coded/enforced |
| SnapMirror schedule | 30 minutes | Tested/recommended* |
| Snapshot copy schedule | 30 minutes | Tested/recommended* |

Note

*Limits described as tested/recommended are tested limits based on a 10-node cluster. If allowed by the platform, actual limits are not hard-coded and can extend beyond these limits up to 24 nodes. For more information, refer to ["Theoretical or absolute maximums"](#). However, official support for the number of member volumes is 200. If you need to exceed this limit, contact Fujitsu Support to start the qualification process for more member volumes.

6. Deciding whether FlexGroup Volumes Are the Right Fit

ONTAP FlexGroup volumes are an ideal fit for many use cases—particularly the ones that are listed in ["Ideal use cases"](#).

However, not all use cases make sense for FlexGroup volumes. This chapter provides information to help you decide whether FlexGroup volumes are the right fit for your workloads.

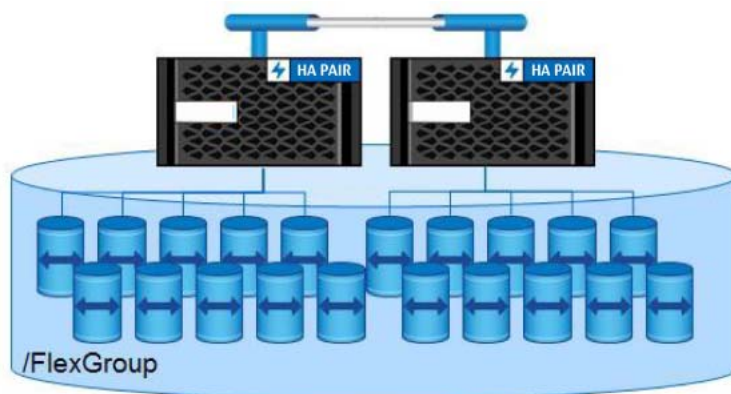
Scale-out performance

FlexGroup volumes distribute their data and load among the multiple constituents that make up the collective FlexGroup volume. This model allows a FlexGroup volume to use more of the resources within each node (CPU, network adapters, disks, and so on) and to use more nodes within a cluster to address a workload.

In addition, the concept ties in nicely with the ONTAP clustered architecture, which allows the nondisruptive addition of nodes and disks to increase performance without negatively affecting applications. With a FlexGroup volume, you can simply expand the FlexGroup to add more members or use nondisruptive volume move technology to redistribute the member volumes across the new nodes.

A single FlexGroup volume internally comprises multiple separate FlexVol volumes, which in turn can be stored on any aggregates and can span multiple nodes in your cluster.

Figure 6 FlexGroup volume



In addition, the concept ties in nicely with the ONTAP clustered architecture, which allows the nondisruptive addition of nodes and disks to increase performance without negatively affecting applications. With a FlexGroup volume, you can simply expand the FlexGroup to add more members or use nondisruptive volume move technology to redistribute the member volumes across the new nodes.

A single FlexGroup volume internally comprises multiple separate FlexVol volumes, which in turn can be stored on any aggregates and can span multiple nodes in your cluster.

When clients add files and subdirectories to the FlexGroup volume, ONTAP automatically determines the best FlexVol member to use for storing each new file and subdirectory. The FlexGroup volume attempts to organize your data, both for best performance and for data and load distribution.

Because of this workload distribution, FlexGroup volumes can handle much more metadata traffic than a FlexVol volume. Thus, FlexGroup volumes can be useful for a variety of workloads that are metadata-intensive or that require a large amount of throughput.

Feature compatibility limitations

FlexGroup volumes support some common NAS protocols, such as NFSv3, NFSv4.0, v4.1 and v4.2, SMB2.x, and SMB3. For details about the support of those protocols and which ONTAP release they were supported, see [Table 1](#), [Table 2](#), and [Table 3](#) in "[FlexGroup Feature Support and Maximums](#)".

Additionally, FlexGroup volumes are what makes up FlexCache volumes and ONTAP S3 object buckets. However, only FlexCache volumes are able to leverage NAS protocol interaction. S3 buckets are only accessible via the S3 protocol. ONTAP S3 is available for general use in ONTAP 9.8.

SMB 1.0 is not supported for use with FlexGroup volumes. FlexGroup volumes do not support block protocol/SAN access (iSCSI, FCP, NVMe).

[Table 7](#) provides information for deciding whether FlexGroup volumes are the right fit for an environment by comparing the currently available container types in ONTAP.

Table 7 ONTAP volume family comparison

| | FlexVol Volumes | FlexGroup Volumes |
|---|---|--|
| Client access protocols (current support) | SAN (FCP, iSCSI, NVMe) NAS <ul style="list-style-type: none"> • SMB1.0, 2.1, 3.x • NFSv3 • NFSv4.0, NFSv4.1, NFSv4.2 | S3 NAS <ul style="list-style-type: none"> • SMB2.x, 3.x • NFSv3 • NFSv4.0, NFSv4.1, NFSv4.2 |
| Capacity scaling | <ul style="list-style-type: none"> • Single FlexVol volume • Can be mounted to FlexVol or FlexGroup volumes in the namespace • 100TB, 2 billion file limit | <ul style="list-style-type: none"> • Can be mounted to FlexGroup or FlexVol volumes in the namespace • 20PB* • 400 billion files* • Nondisruptive capacity increases *Current tested limits on 10-node cluster; can extend beyond these values |
| Metadata scaling | FlexVol volumes are limited to a single node for metadata processing and serial processing of metadata, which does not take full advantage of the node's CPU threads. | FlexGroup volumes can use multiple nodes (and their resources) and multiple aggregates. In addition, FlexGroup can use multiple volume affinities to maximize CPU thread utilization potential. |
| ONTAP feature compatibility | Compatible with all ONTAP features | Supports most ONTAP features. <ul style="list-style-type: none"> • For information, refer to Table 1. |
| Throughput scaling | Limited to: <ul style="list-style-type: none"> • One node (set of CPU, RAM, network ports, connection limits, and so on) • One aggregate | FlexGroup volumes can use the resources of an entire cluster in service of I/O, providing much higher throughput than a single FlexVol volume can, with linear scale of performance as you add nodes to the FlexGroup. |
| Cloud support | All cloud integration, such as: <ul style="list-style-type: none"> • Cloud Volumes ONTAP • Cloud Volumes Services • SnapMirror Cloud | Cloud Volumes ONTAP (CVO) – CLI only <ul style="list-style-type: none"> • Capacity limitations of CVO apply |

| | FlexVol Volumes | FlexGroup Volumes |
|----------------------------|---|-------------------|
| ONTAP upgrades and reverts | Data stored in any volume family is safely retained during ONTAP version changes. | |
| GUI compatibility | <ul style="list-style-type: none">• ONTAP System Manager• Active IQ Performance Manager• Active IQ Unified Manager• Cloud Insights | |

Simplifying performance

A single FlexGroup volume can consist of multiple FlexVol member volumes, which in turn can reside on any aggregate and on any node in your cluster. As clients drive traffic against that FlexGroup volume, ONTAP automatically breaks that traffic into tasks for different constituent FlexVol volumes to perform. This approach provides for a concurrency of operations that a single FlexVol volume is incapable of handling.

The benefit of this scale-out behavior is a dramatic increase in processing power that can scale linearly as you add nodes to your ONTAP cluster. A single FlexGroup volume can service much heavier workloads than a single FlexVol volume can at more predictable latencies.

AFF A700 testing

In a simple workload benchmark using a software build tool (Git), a Linux kernel was compiled on a two-node A700 cluster.

The following configuration was used:

- Two-node AFF A700 cluster
- A single aggregate of 800GB SSDs per node
- FlexVol volume: single node, 100% local
- FlexGroup volume: spans high-availability (HA) pair, eight members per node (16 members total)

The workload was as follows:

- GCC library compile
- Clone operations only (these operations showed the highest maximum throughput for both FlexVol and FlexGroup)
- Four physical servers
- User workloads/threads on the clients that ranged from 4 to 224

[Figure 7](#) compares the maximum achieved throughput (read + write) on Git clone operations on a single FlexVol volume versus a single FlexGroup volume that spanned two nodes.

Note

The maximum throughput of the FlexGroup volume reaches nearly five times the amount of the FlexVol volume without the same degradation of the FlexVol volume as the workload reaches 64 threads.

Figure 7 FlexVol volume versus FlexGroup volume—maximum throughput trends under increasing workload

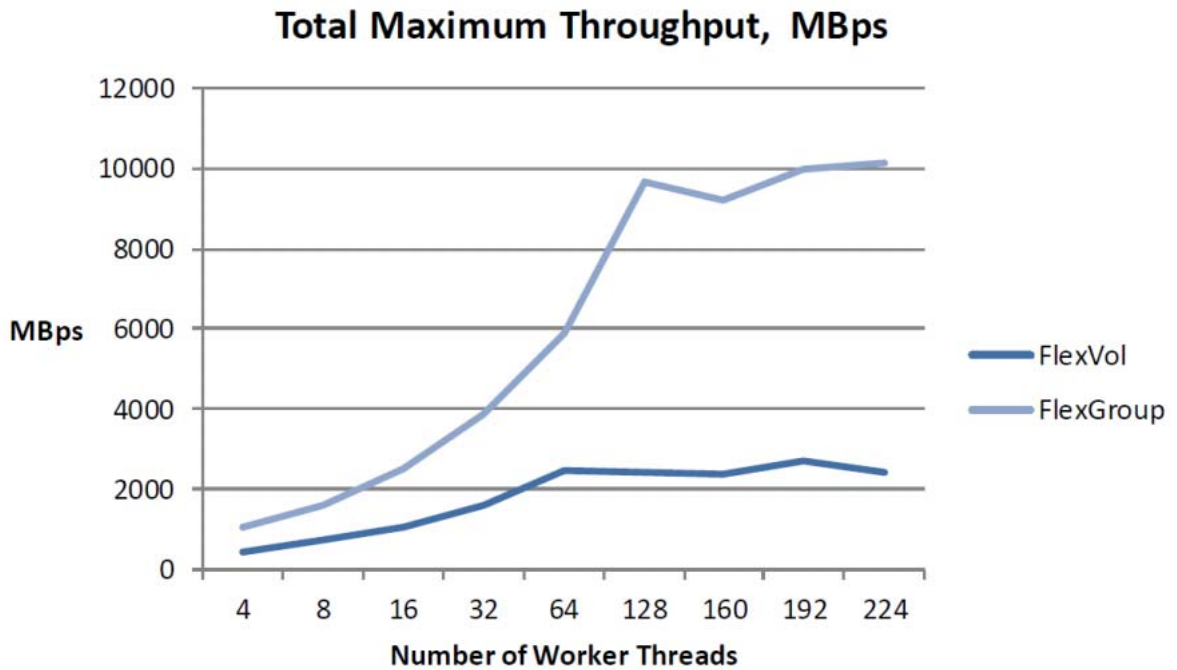


Figure 8 compares a FlexVol volume and a FlexGroup volume in the same configurations. This time, we break down the maximum read and write throughput individually, as well as comparing that against the average throughput for the FlexVol volume and the FlexGroup volume.

Figure 8 FlexVol volume versus FlexGroup volume—maximum throughput trends under increasing workload, detailed

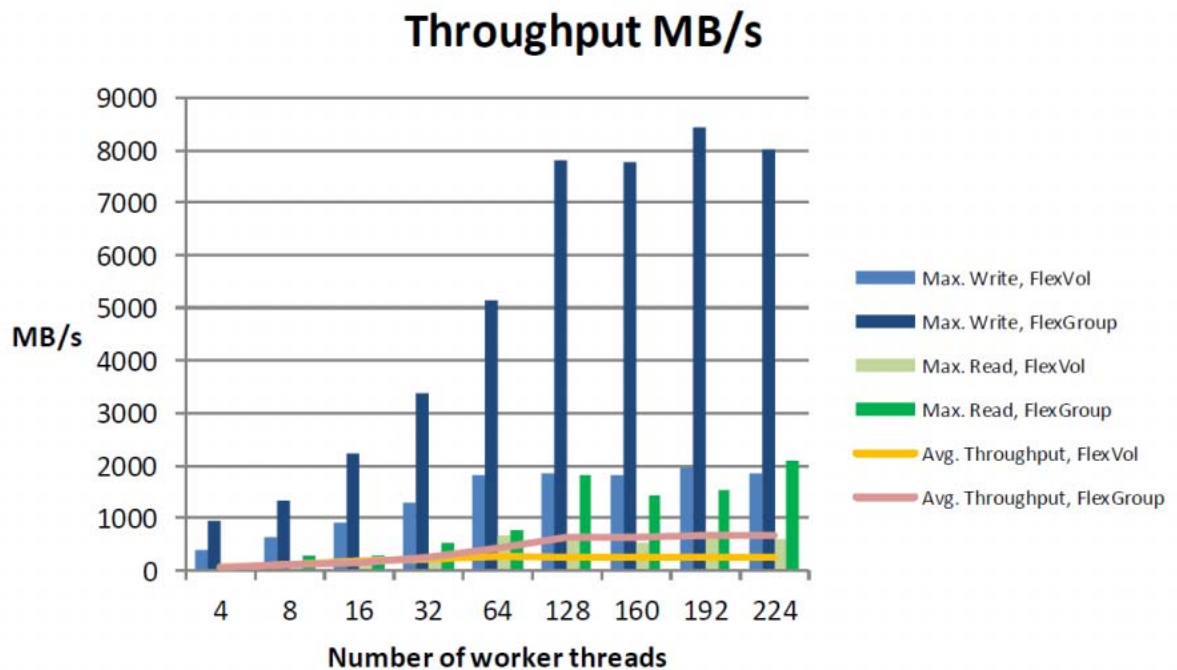
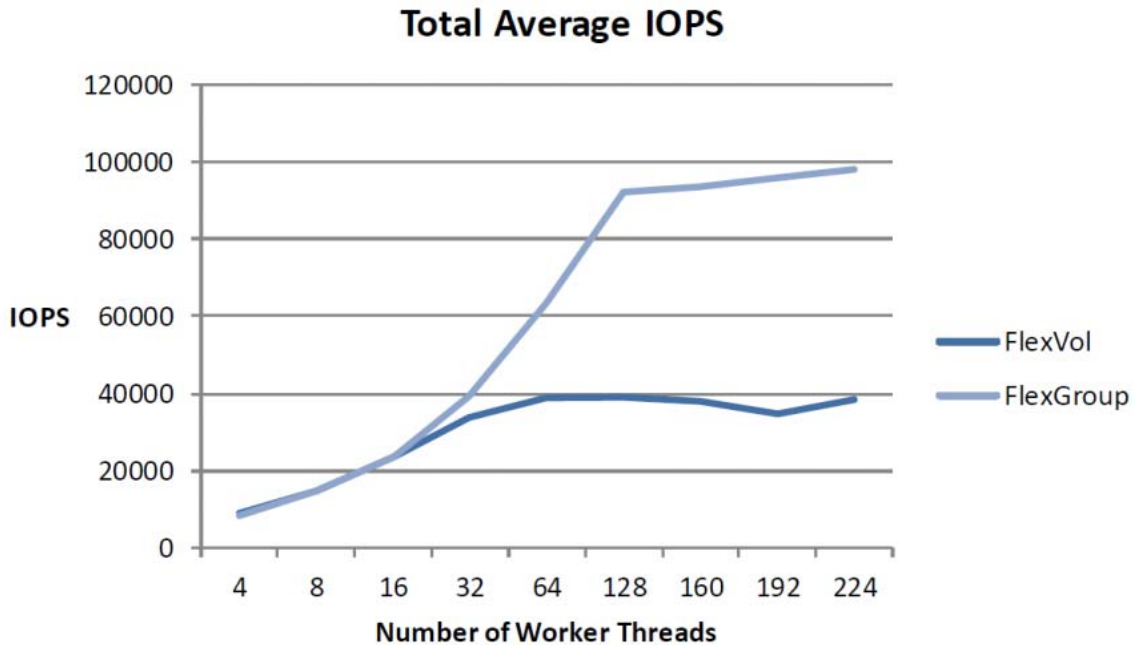


Figure 9 shows the maximum total average IOPS for a FlexGroup volume versus a FlexVol volume on the AFF A700. Again, note the dramatic increase of IOPS for the FlexGroup volume versus the degradation of IOPS at 64 threads for the FlexVol volume.

Figure 9 FlexVol volume versus FlexGroup volume—maximum average total IOPS



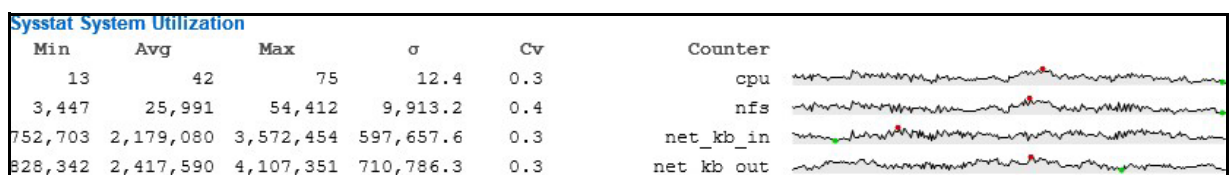
FlexGroup performance with big data workloads

Due to the FlexGroup volume’s capacity and ability to scale a single namespace across multiple compute nodes in a cluster, it provides an interesting use case for big data workloads, such as [Apache Hadoop](#), [Splunk](#), and [Apache Spark](#). These applications generally expect only one or two directories to dump large amounts of data and high file counts, requiring high throughput at a low latency. FlexVol volumes were able to accomplish this performance, but not without some tweaks to the application to make it aware of multiple volumes.

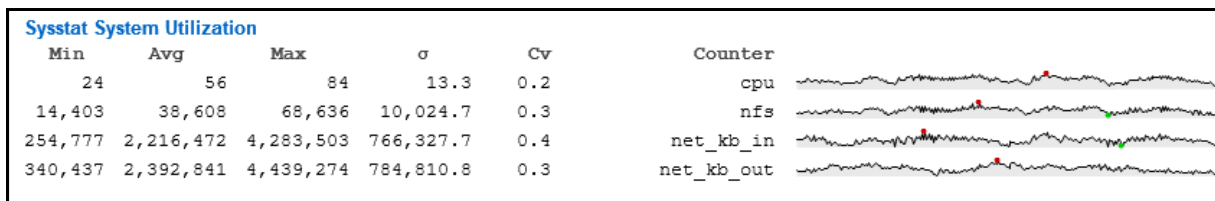
Also, the Customer Proof-of-Concept (CPOC) lab conducted some performance testing using the [TeraSort](#) benchmark, which is used to test Apache big data workloads. In this testing, a two-node AFF A700 cluster running ONTAP 9.2 was used to push a maximum of 8GBps in and out of the cluster at an average read latency from ~3ms to 5ms and an average write latency from ~4ms to 8ms, while keeping the average CPU utilization around 55% on both nodes. Using a FlexGroup volume with big data workloads allows all available hardware to be used and provides a way to nondisruptively scale the capacity and performance by adding nodes for the workload as needed.

Figure 10 TeraSort benchmark statistics summary on a FlexGroup volume

Node1



Node2



As a bonus, big data workloads running on ONTAP FlexGroup volumes have shown a space savings of nearly 50% with storage efficiency features such as inline aggregate deduplication, inline data compaction, and inline compression.

Automatic workload adaptation

The FlexGroup volume continually adapts to the current conditions in the cluster, changing behavior constantly to keep usage evenly consumed and to keep dynamic load evenly balanced. Trade-offs are implicit in this continual balancing act. The cost of this automatic balancing is that a FlexGroup volume cannot attain the same theoretical maximum performance that a perfectly balanced and manually organized collection of FlexVol volumes could otherwise attain. However, the FlexGroup volume can get very close to that maximum, and it requires no fore-knowledge of the workload to accomplish its work. In addition, a FlexGroup volume adds a simplicity aspect to large data layouts that a multiple FlexVol architecture cannot.

FlexGroup volumes perform better—balancing load and usage more smoothly—when faced with a broad variety of workloads and high data-creation rates. Thus, a single FlexGroup volume that performs many different roles can be a more effective use of your cluster's resources than if you use different FlexGroup volumes for different workloads. You can, however, junction multiple FlexVol volumes and FlexGroup volumes together in the same ONTAP SVM if you require greater control and flexibility over your data.

If a workload is creating a high number of small files, then the FlexGroup volume places those files to balance them evenly across volumes while favoring folder locality to increase performance. If the workload is a smaller number of large files, then ONTAP recognizes that difference. Rather than favoring local folder placement (which could result in multiple large files ending up on the same member volume and creating an artificial imbalance of data), ONTAP instead places files in a more round-robin fashion to ensure even space allocation. This allows for a wider variety of workloads to perform optimally on FlexGroup volumes, preventing space imbalance scenarios and reducing the need for administrator intervention.

ONTAP 9.8 introduces a change to how capacity is managed called Proactive resizing. This change effectively maintains a free space buffer across all member volumes when a capacity threshold is reached to automatically help protect against member volumes getting too full and guards against volumes having disparate free space. Additionally, prior to 9.8, if a FlexGroup member volume reached a 90% capacity threshold, performance of new file creations would suffer, as ONTAP would start to create more remote hardlinks for new files. ONTAP 9.8 removes that 90% threshold since proactive resizing should maintain enough free space to avoid the need to redirect traffic.

Ingest algorithm improvements

Every ONTAP release further improves the ingest algorithms for FlexGroup volumes that help ONTAP make better decisions about how new data is placed in FlexGroup volumes. The algorithms also improve the way FlexGroup volumes respond when member volumes approach "nearly full" status.

Best Practice 1: Always run the latest ONTAP version

Fujitsu strongly recommends that you run the latest patched ONTAP version when using FlexGroup volumes for the best ingest results. You can download the latest release at [Fujitsu download site](#).

Some of the changes to ingest that have taken place in various releases include:

- Better handling of mixed workload types in ONTAP 9.7
- Proactive resizing and adjustment of remote placement triggers in ONTAP 9.8

Performance features

ONTAP provides various features to better control and monitor performance.

Quality of service (QoS)

You can apply maximum storage QoS policies to help prevent a FlexGroup volume's workload from overrunning other volume workloads. ONTAP storage QoS can help you manage risks around meeting your performance objectives.

ONTAP supports FlexGroup volumes for QoS minimums (also referred to as guarantees or floors), which provide a set threshold of performance that is allocated to a specified object.

You use storage QoS to limit the throughput to workloads, provide guaranteed performance to workloads, and to monitor workload performance. You can reactively limit workloads to address performance problems, and you can proactively manage workload performance to prevent problems.

■ How storage QoS policies work with FlexGroup

With FlexGroup, storage QoS policies are applied to the entire FlexGroup volume – not at the member volume level. Because a FlexGroup volume contains multiple FlexVol member volumes and can span multiple nodes, the QoS policy gets shared evenly across nodes as clients connect to the storage system. [Figure 11](#) and [Figure 12](#) show how storage QoS gets applied to a FlexGroup volume.

Figure 11 Storage QoS on FlexGroup volumes—single-node connection

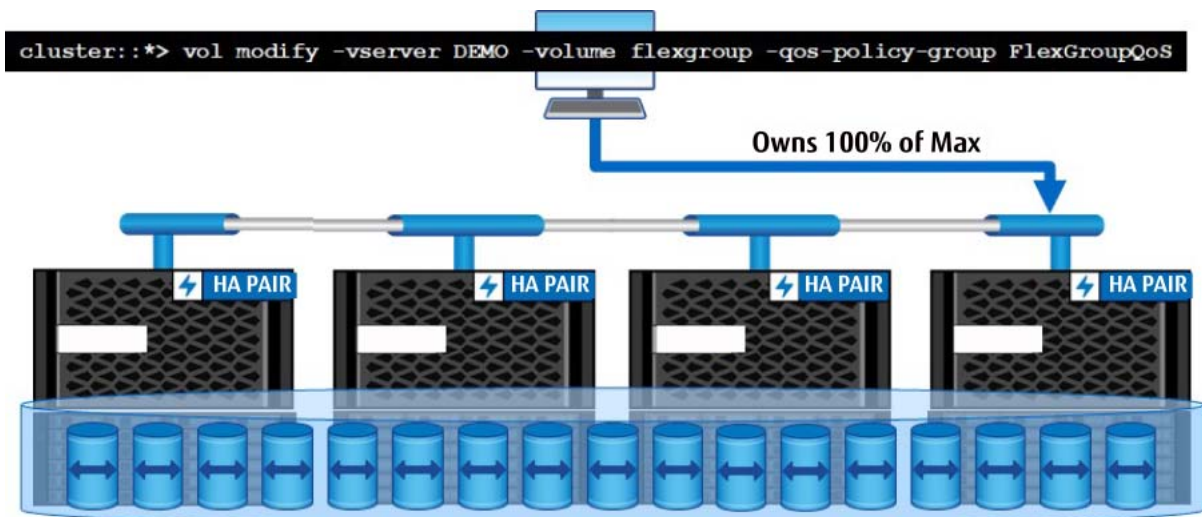
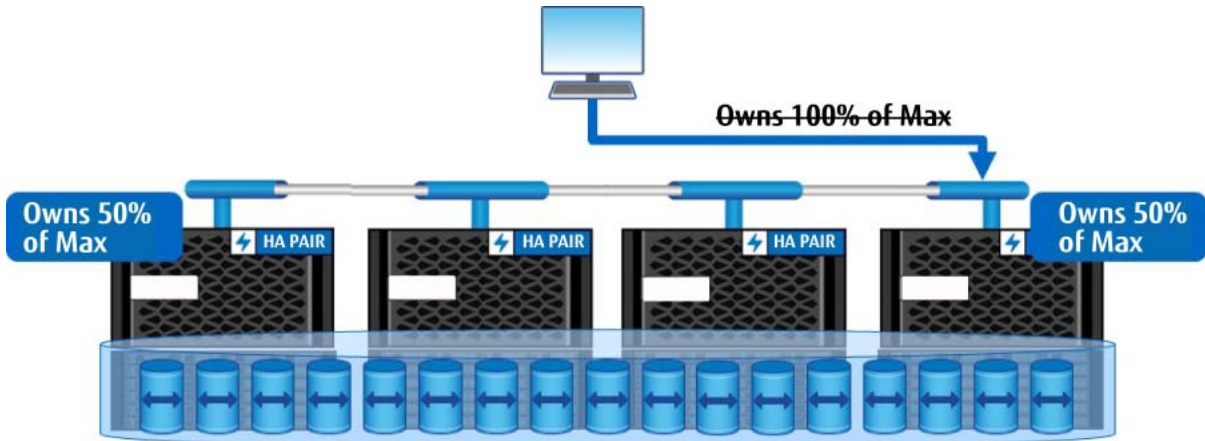


Figure 12 Storage QoS on FlexGroup volumes: multinode connection

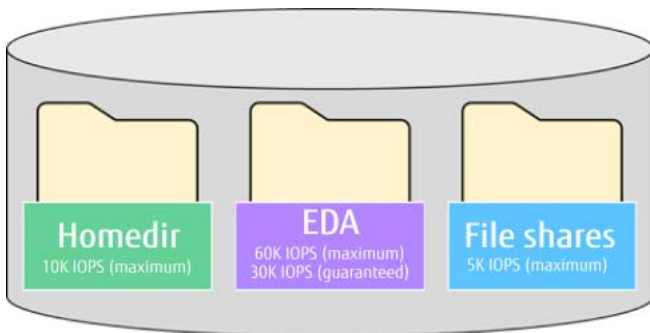


Note
Nested policies are currently not supported with FlexGroup volumes.

Qtree QoS

ONTAP 9.8 introduces the ability to apply QoS policies at the qtree level.

Figure 13 Qtree QoS use cases



This means you can provision a FlexGroup volume and manage performance in that volume with qtrees, rather than creating multiple FlexVol or FlexGroup volumes to divide that workload up.

Qtree QoS also provides a more granular level of statistics for the qtree than previous Qtree Statistics offered.

Qtree QoS in ONTAP 9.8 can be used with FlexGroup volumes and FlexVol volumes, but it has the following limitations:

- NFS only
- CLI/REST API only; no current GUI support
- No adaptive QoS support

Qtree QoS also provides some enhanced statistics for performance monitoring, which aids in understanding specific workloads.

| Policy Group | IOPS | Throughput | Latency |
|--------------|------|------------|---------|
| qtree | 113 | 113.00MB/s | 2.82ms |

Adaptive QoS

ONTAP introduced adaptive QoS support for FlexGroup volumes, which allows ONTAP to adjust the IOPS and TB values of a QoS policy as the volume capacity is adjusted.

Note

Adaptive QoS is not supported with qtree QoS because qtrees are not objects you can grow or shrink.

Qtree statistics

Qtree statistics are made available for FlexGroup volumes. These statistics provide granular performance information about FlexGroup volumes and their qtrees. The following example shows a statistics capture for a FlexGroup volume running a large NFS workload.

```
cluster::> statistics qtree show -interval 5 -iterations 1 -max 25 -vserver DEMO -volume
flexgroup_local

cluster : 11/7/2018 15:19:15

      Qtree Vserver      Volume      NFS CIFS Internal *Total
      -----
DEMO:flexgroup_local/  DEMO flexgroup_local 22396    0        0 22396
DEMO:flexgroup_local/qtrees
                        DEMO flexgroup_local    0    0        0
```

Workloads and behaviors

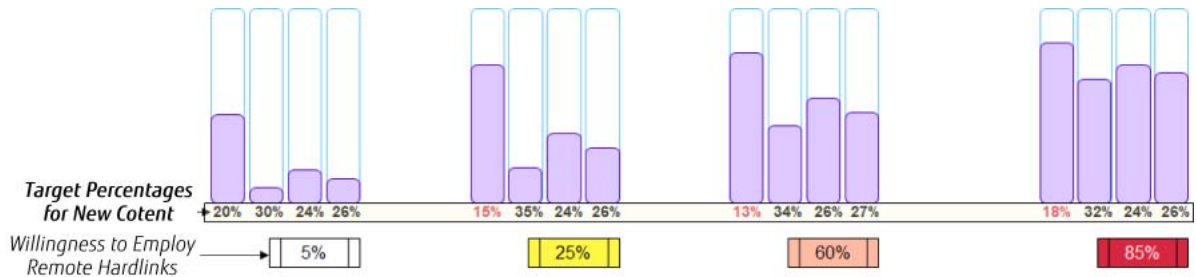
In an optimally balanced FlexGroup volume, all constituents have roughly the same amount of data and load, and the volume can maintain that state while using a high frequency of local placement for best performance. A workload with a good balance of folders and similarly sized files would be able to maintain local parent folder placement while also keeping a relatively even balance of capacity.

A less optimal FlexGroup volume might have some constituents that hold more or less data than their peers, or that are receiving much more or much less traffic. Workloads that have only a few folders with many files per folder, or workloads with highly variant file sizes can experience data usage imbalances in a FlexGroup.

Capacity balance, however, is not the most important function of a FlexGroup volume. Instead, a FlexGroup volume functions best when there is a mix of local placement for performance along with capacity and inode count balance. We don't want to sacrifice performance for the sake of perfectly balanced capacities across member volumes.

In the figures below, you can see several examples of capacity imbalances in a FlexGroup volume and the varying degrees of remote placement based on the overall fullness of the FlexGroup volume. For example, a relatively empty FlexGroup with data imbalances is less likely to use remote hardlinks than a FlexGroup that is much closer to being full.

Figure 14 Capacity imbalance and likelihood of remote placement



ONTAP constantly monitors the ongoing state of the member volumes and adjust placement decisions based on the current state of the FlexGroup volume. If one member volume is a little out of balance from the others, then it's likely that no adjustments will be made to ingest. But if that member volume starts to approach 90% capacity, or has a capacity discrepancy that exceeds 512GB, then ONTAP makes more aggressive placement choices for new data to correct the imbalance. This adjustment means more remote file placement to other member volumes, which can have a negative (but potentially unnoticeable—roughly 5% to 10%) effect on FlexGroup volume performance.

In some cases, a FlexGroup volume might appear to be perfectly balanced in usage and load, but it has had to resort to more frequent remote placement frequently to maintain that state. This situation can occur when FlexGroup member volumes get closer to being 100% used.

Best Practice 2: Stop worrying about capacity imbalances.

A FlexGroup with a capacity imbalance across member volumes is not in itself a problem and should not be treated as such. Instead, look at capacity imbalances as a potential cause if a FlexGroup is not performing as expected or if the capacity imbalances are so extreme that you're running out of space on your cluster. Be sure to engage Fujitsu Support if you feel that your FlexGroup volume capacity imbalance is the source of a performance issue.

Workloads determine the degree to which a FlexGroup volume behaves optimally. Most workloads can be used with FlexGroup volumes in ONTAP 9.8, but some workloads (such as EDA/software development) perform more optimally than others.

Optimal workloads

A FlexGroup volume works optimally when it is under heavy ingest load—that is, when there is a high rate of file and directory creations. ONTAP makes its placement decisions as new files and directories are created, so the more often this action occurs, the more frequently ONTAP has an opportunity to correct existing imbalances in load and usage. If a workload is a heavy read or write-append to existing files, then the FlexGroup placement doesn't really factor in as much; once the files are placed, they remain where they landed. As mentioned in ["Ingest algorithm improvements"](#), each new ONTAP release adds improvements and adjustments to FlexGroup volumes that can address more variant workloads.

Generally speaking, the following represent attributes of the most optimal FlexGroup workloads.

- FlexGroup volumes work best with numerous small subdirectories**
This means dozens to hundreds of files per directory, because they allow the FlexGroup volume to place new child subdirectories remotely while keeping individual files local to their parent directories for best performance. Directories containing more files experience more remote placement to other member volumes in an attempt to balance capacity and file counts.
- A FlexGroup volume responds well to heavy concurrent traffic**
Bringing more workloads—especially traffic from multiple clients that are doing different things at the same time—to bear against a FlexGroup volume simultaneously can improve its overall performance. In other words, don't expect to push a FlexGroup volume to its limit and achieve the performance possibilities mentioned in this document with one to a few clients.

- **A FlexGroup volume works best when there is plenty of free space**
When constituents begin to fill up, the FlexGroup volume begins to employ remote placement more frequently so that no one constituent becomes full before its peers do. This increased usage of remote placement comes with a metadata performance penalty.
- **FlexGroup volumes work best with high rates of write metadata operations**
ONTAP FlexVol volumes already process read and write I/O in parallel, and metadata read operations (such as `GETATTR`). However, ONTAP processes write metadata (such as `SETATTR` and `CREATE`) serially, which can create bottlenecks on normal FlexVol volumes. FlexGroup volumes provide a parallel processing option for these types of workloads, which results in performance that is [two to six times better for these types of workloads](#).

Performance and capacity considerations

For best performance, keep plenty of free space on the FlexGroup volume (at least 10% free space available) when it is under heavy load.

To better manage FlexGroup volume capacity, use ONTAP 9.8 or later. That release contains a number of features to manage capacity, including volume autogrow, [elastic sizing](#) and, specifically for ONTAP 9.8, [proactive resizing](#).

Free space for FlexVol member or constituent volumes can be monitored at the admin privilege level with the following command:

```
cluster::> vol show -vserver SVM -volume [flexgroupname__]* -is-constituent true -fields available,percent-used
```

Note

You can also monitor free space by using GUI utilities such as Active IQ Unified Manager and/or by configuring ONTAP to generate alerts. Refer to ["14. Capacity Monitoring and Alerting" \(page 157\)](#) for more details.

Good workloads

Even if a workload does not conform to the preceding parameters, odds are good that a FlexGroup volume can accommodate it with ease. Remember that ["Optimal workloads"](#) describes situations that can help a FlexGroup volume perform optimally, but even a suboptimal one provides good throughput, scaling, and load distribution for most use cases.

Nonideal workloads – large files

A few activities can make a FlexGroup volume work harder to maintain its balance of load and usage among constituents. Most of these activities relate to large files in one way or another. Although these workloads are able to use FlexGroup volumes, you should strive to understand the average file size and largest file size of the workload before implementing. The following considerations should be made when deploying workloads that don't fit into the "ideal" or "good" workload definitions.

Consideration #1: ONTAP cannot predict the future size of your files

One of the key challenges of large file workloads is that storage systems are not aware of how large a file will become over time. Clients often do not have this information either; instead, a file starts out as a small inode in the storage system, and then data is written to it until the file creation/write is completed. This is exacerbated by the FlexGroup volume's tendency to keep file placement local to the parent folder for performance considerations. It's equally possible that a folder with 100 files that are 500MB will all land in the same member volume as it is that a folder with 100 files that are 4K in size, depending on how fast the files are written and how many clients are involved in creating the files. As a result, in that scenario, one member volume might end up with 50GB of used space and the other member volume might only have 400KB used.

As mentioned before, this isn't inherently a problem, but it is a noticeable discrepancy to the storage administrator and can present problems if the FlexGroup volume isn't sized appropriately. For example, what if the member volumes are all 100GB in size? Therefore, in this example, one member is 50% full, and the others have 0% capacity.

Generally speaking, workloads like this balance themselves out over time, and the FlexGroup volume maintains even distribution and good performance. The benchmark for concern with these workloads should not be "My member volume capacities are uneven," but rather, "My FlexGroup volume is not performing as well as I expect."

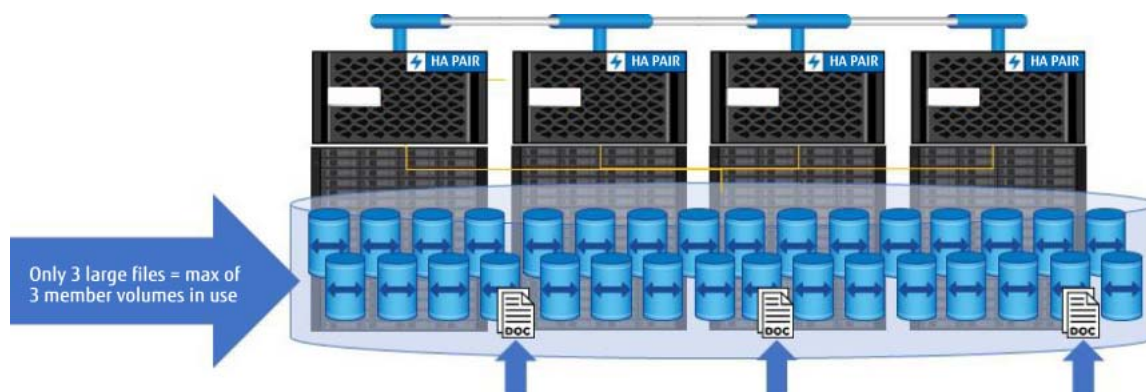
Consideration #2: Large-file workloads are generally low-file-count workloads

Large files are marginally more difficult for the FlexGroup volume to process than small files are, primarily because using large files typically means using fewer of them overall. As previously mentioned, the FlexGroup volume performs best when new files and directories are being created frequently. If the working set consists of many large files that are roughly the same size, the FlexGroup volume should not have trouble maintaining usage and load distribution among constituents. Performance with large file workloads act more like that of a FlexVol, as the benefits of parallel ingest don't come into play with workloads that are not workloads that ingest many files at a time.

Consideration #3: Large-file workloads are not guaranteed to distribute evenly

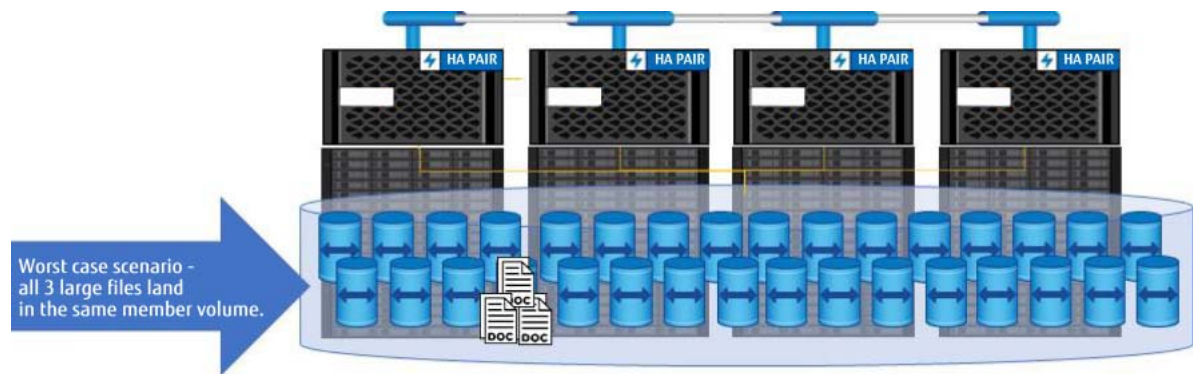
Large files also have the property of holding a great deal of information. Reading or writing that much information can take a long time. If the workload concentrates on only a few of those large files (say, reading or writing a large single-file database), then all that traffic is handled exclusively by the constituents that host those files. Because other constituents are not participating in the workload at the time, this situation can result in suboptimal usage of the FlexGroup volume. In general, expect roughly the same performance for large-file or streaming workloads in a FlexGroup volume as you would see in a FlexVol volume.

Figure 15 FlexGroup volume with a few large files; why usage can be suboptimal



In addition, there is no guarantee that the large files are all distributed evenly across the FlexGroup volume. In the above graphic, the three large files landed in three different member volumes. This scenario might happen if we write one of those files at a time, and they finish writing the entire file before the next file is written. But if all three files are written at the same time, then we run the possibility that all three of the files land in the same member volume.

Figure 16 Potential worst case scenario for large files; all land in the same member volume



Again, if there is enough capacity in the FlexGroup volume, it isn't necessarily a problem, but it does mean we have to factor in a few things when dealing with large file workloads.

- Total FlexGroup size
- Member volume count
- Member volume size (as compared to largest file size)

Having larger member volumes offsets potential issues large files might create, and features such as proactive resizing and elastic sizing mitigates potential capacity issues affecting data availability.

Consideration #4: Large files create imbalances that potentially affect performance

Another concern with large files is that a single file can consume enough space on the constituent to substantially affect the balance of usage among constituents. Sometimes a few files grow to a size that is orders of magnitude above the average file size. The result is that some constituents (the ones that happen to hold the aberrantly large files) end up with much more data than their peers have. In response, the FlexGroup volume begins to divert other new content creations onto the underused constituents. As a result, a subset of constituents can end up servicing most of the traffic. This problem is not typically severe; it simply represents suboptimal behavior. ONTAP 9.7 and later versions make substantial strides in handling the placement of these types of files and workloads so that they are better balanced across member volumes.

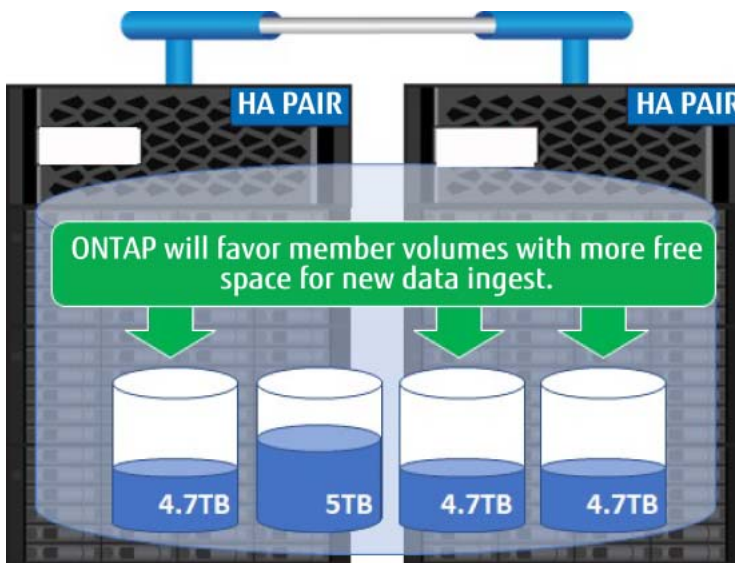
Best Practice 3: Large File Size Considerations

Before sizing a FlexGroup volume, perform an analysis to determine the largest possible file size in a workload. Then, the member volume sizes should reflect those large file sizes, so that a large file cannot consume more than 1% to 5% of a FlexGroup member volume. Following this best practice helps avoid "out of space" concerns. Also, running ONTAP can help avoid "out of space" concerns by way of the elastic sizing functionality. Running the latest patched version of ONTAP is always a good practice for FlexGroup volumes.

One other concern relates to running with the FlexGroup volume continually very close to full. As the FlexGroup volume becomes full, ONTAP becomes proactive in placing new content on those constituents that still have free space. If the working set consists primarily of small files, this behavior is adequate to prevent clients from receiving `Volume Full` errors until the collective FlexGroup volume is indeed full.

Elastic sizing provides a way for ONTAP to borrow space from less-full member volumes and allow file writes to complete in member volumes. ONTAP 9.8 also introduces proactive resizing, which further improves the capacity management for full member volumes.

Figure 17 Capacity imbalance example



Best practices when using large files with FlexGroup volumes

FlexGroup volumes operate best when dealing with lots of smaller files. However, they can also be effective when storing larger files as long as the FlexGroup volume is configured to account for that workload up front. When you're sizing a FlexGroup volume for large files, it's important to consider [what a large file is](#), and [what the largest and average file sizes in a workload are](#).

File sizes must be factored in when you design a FlexGroup volume so that [member volumes are sized appropriately](#). ONTAP 9.7 and later versions make this process unnecessary with the addition of [elastic sizing](#), and ONTAP 9.8 simplifies capacity management even more with [proactive resizing](#). In general, you can apply the following best practices for large file workloads:

- For large file sizes, consider deploying larger member volumes with fewer members per FlexGroup volume. See [Initial volume size considerations](#) for details.
- When running ONTAP 9.7, elastic sizing is enabled by default. Using volume autogrow disables elastic sizing for a FlexGroup volume in those releases, so decide how you want to manage capacity.
- Preferably, use ONTAP 9.8 or later to gain the benefits of proactive resizing. ONTAP 9.8 and later allow elastic sizing and volume autogrow to co-exist.
- Use quota enforcement to limit and monitor the capacity in qtrees or by user.
- Before deploying a FlexGroup volume, use [XCP to scan the file system and analyze the file sizes](#) to understand average file size, largest file size, and so on.
- You should size a FlexGroup volume so that member volumes are less likely to become imbalanced. The largest file size should not exceed 1% to 5% of the member volume's capacity, but keep in mind that a smaller member volume size means that, for 1% to 5% of volume capacity, the largest files must be smaller, relatively speaking. Smaller files finish writing to the storage system faster and do not create enough discrepancy to shift the ingest algorithm very much. When possible, avoid member volume sizes under 1TB when file sizes are 50GB or less (minimum member volume size is 100GB). Since the notion of a "large file" is 1% to 5% of the member volume space, that "large file" size value is much smaller in a 100GB member volume (1GB to 5GB) than it is in a member volume that is 1TB (10GB to 50GB).

Performance expectations: read-heavy workloads

Performance in a FlexGroup volume can greatly exceed that of a FlexVol volume or competitor systems for write-metadata-heavy workloads (high `CREATE` and `SETATTR` calls) that ingest many files. However, other workloads, such as file streams, file appends or read-heavy workloads, don't see the same extreme performance gains over FlexVol volumes that the ingest-heavy workloads see. This is because a FlexGroup volume is designed to overcome the bottleneck of serial processing of write metadata workloads by providing more volume affinities to those workloads. Basic read and writes don't face this serial processing bottleneck.

In some cases (especially with all local traffic), a set of multiple FlexVol volumes might perform slightly better than a FlexGroup volume for random and sequential read/write workloads. However, the complexity involved with creating and managing multiple FlexVol volumes versus a single FlexGroup volume might outweigh the slight performance gains.

For read-heavy workloads, using FlexGroup volumes has some benefits over using single FlexVol volumes, such as the following:

- Scaling across multiple CPUs and nodes to load balance reads to multiple files
- Single namespace for a large-capacity bucket

When deciding whether to use a FlexGroup volume, consider support for specific features. See the earlier section on [what is and is not currently supported with FlexGroup volumes](#).

In addition, when using read-heavy workloads, also consider deploying FlexCache volumes attached to a FlexGroup origin volume to distribute the workload across more volume affinities in the cluster, or even across other SVMs in the same cluster or other clusters across multiple sites or in the cloud.

Data imbalances in FlexGroup volumes

In rare cases, a FlexGroup workload might have an imbalance of capacity in the member volumes. On its own, this does not indicate a problem; this is only a problem if performance is noticeably suffering or capacity imbalances are causing clusters to run out of available space. In most cases, a capacity imbalance does not need to be addressed. ONTAP performs the work to balance out the workload if there are consistent new file creations. This is especially true in ONTAP 9.8 and later due to [proactive resizing](#). In cases where the data is static, the imbalance does not resolve. However, the performance should remain roughly the same as if there was no data imbalance.

Data imbalances can occur in the following scenarios:

- A mix of large and small files are written to a FlexGroup volume.
- A file is written to a FlexGroup and then appended later, thus growing and increasing used capacity.
- Multiple large files are written at once to the same folder; ONTAP does not know how large these files will get, so they are placed to the local member volume for performance considerations.
- Large datasets get deleted and you happen to delete more files on one member volume than other member volumes.
- A user creates a very large file (such as a zip file of many existing files).

Each ONTAP release makes adjustments to the ingest algorithms to try to address wider ranges of workload scenarios, so use the latest ONTAP release available. If issues are present, open a technical support case to isolate and remediate the issue.

Post-placement rebalance

Currently, [ONTAP has no method to rebalance the files that have already been ingested](#). The only way to effectively rebalance is to copy the data from a FlexGroup volume to a new, empty FlexGroup volume all at once. This process is disruptive, because clients and applications must point to the new FlexGroup volume after the data has been moved. Also, this process is performed at a file level, so it could take a considerable amount of time. Rebalancing the files should be considered only if the imbalance of volumes creates an issue that affects production. As mentioned, capacity imbalances are usually imperceptible to client activity. Most customers don't notice an imbalance until they are alerted to a capacity threshold. If data rebalance is necessary, consider using the [XCP Migration Tool](#) to speed up the file copy process.

7. Initial FlexGroup Design Considerations

This chapter covers initial ONTAP FlexGroup volume design considerations. In presenting this information, Fujitsu assumes that no previous FlexGroup volumes have been created on the cluster. Fujitsu also assumes that you have experience with and knowledge about managing ONTAP through the CLI and the GUI and that you have administrator-level access to the storage system.

Cluster considerations

An ONTAP cluster that uses only NAS functionality (CIFS/SMB and NFS) can expand to up to 24 nodes (12 HA pairs). Each HA pair is a homogenous system (that is, two ETERNUS AX series nodes and so on), but the cluster itself can contain mixed system types. For example, a 10-node cluster could have a mixture of four ETERNUS AX series nodes, four ETERNUS HX series systems, and two hybrid nodes for storage tiering functionality.

A FlexGroup volume can potentially span an entire 24-node cluster. However, keep the following considerations in mind.

- **FlexGroup volumes should ideally span only hardware systems that are identical.**
Because hardware systems can vary greatly in terms of CPU, RAM, and overall performance capabilities, the use of only homogenous systems helps promote predictable performance across the FlexGroup volume. Data is balanced anywhere a FlexGroup volume has member volumes deployed; the storage administrator does not control this placement.
- **FlexGroup volumes should span only disk types that are identical.**
Like hardware systems, disk type performance can vary greatly. Since a FlexGroup volume can span multiple nodes in a cluster and the storage administrator has no control over where the data is placed, you should make sure that the aggregates that are used are either all SSD, all spinning, or all hybrid. Mixing disk types can lead to unpredictable performance.
- **Disk sizes are not hugely important.**
While it is important to use similar disk types on aggregates a FlexGroup might span, disk sizes are less important. For instance, if your aggregates have 3TB disks but you bought a set of new 16TB disks, feel free to deploy a FlexGroup across them, provided they are the same media type. The main caveat here is that the member volumes you deploy must be equivalent in size to the others.
- **FlexGroup volumes can span portions of a cluster.**
A FlexGroup volume can be configured to span any combination of nodes in the cluster, from a single node, to an HA pair, to all 24 nodes. The FlexGroup volume does not have to be configured to span the entire cluster. However, doing so can take advantage of all the hardware resources that are available.

ONTAP version considerations

Each release of ONTAP includes new features and improvements for FlexGroup volumes. Although Fujitsu recommends using the latest available patched release of ONTAP, many storage administrators are unable or unwilling to do that.

If you must run an older ONTAP version, familiarize yourself with the feature gaps in that release in ["FlexGroup Feature Support and Maximums"](#), and if possible, test the workload on a FlexGroup volume before deploying in production.

Failure domains

A failure domain is an entity that, if failure occurs, can negatively impact workloads. For example, in an ONTAP cluster, if a both nodes of an HA pair fail (a rare occurrence), the volumes on those nodes become unavailable because there is nowhere for them to fail over to. As a result, the HA pair is considered a failure domain in the cluster. However, a single node in an HA pair can fail in a cluster without disruption because its partner can take it over. In this situation, a single node would not be considered a failure domain.

Errors within a failure domain (such as RAID errors, losing a disk, multipath configuration errors, and metadata inconsistencies) are handled in ONTAP and do not negatively affect the FlexGroup volume.

FlexGroup volumes can span multiple nodes and HA pairs, and thus, multiple failure domains. However, even if a FlexGroup volume spans an entire 10-node cluster, the failure domain is still the HA pair. If you lose access to members in a FlexGroup volume (such as in the rare instance of failure of the HA pair), write access is disabled until all those members are repaired and reintroduced into the FlexGroup volume. The more HA pairs a FlexGroup volume spans, the higher the probability for failure is, because you are now spanning more failure domains. The fewer HA pairs that are used, the lower the probability for failure, but you see less overall performance for the FlexGroup because fewer hardware resources are available for the workload.

Therefore, when planning deployment, consider how many nodes to span in a FlexGroup volume and what SLAs are acceptable, and weigh those considerations against the capacity required and performance needed.

Aggregate layout considerations

An aggregate is a collection of physical disks that are laid out into RAID groups and provide the back-end storage repositories for virtual entities such as FlexVol and FlexGroup volumes. Each aggregate is owned by a specific node and is reassigned during [storage failover](#) events.

Aggregates have dedicated NVRAM partitions for consistency points to avoid scenarios in which slower or degraded aggregates cause issues on the entire node. These consistency points are also known as per-aggregate consistency points and allow mixing of disk shelf types on the same nodes for more flexibility in the design of the storage system.

Best Practice 4: Aggregate Usage with FlexGroup

For consistent performance when using FlexGroup volumes or multiple FlexVol volumes, make sure that the design of the FlexGroup volume or FlexVol volumes spans only aggregates with the same disk type and RAID group configurations for active workloads. For tiering of cold data, predictable performance is not as crucial, so mixing disk types or aggregates should not have a noticeable effect.

[Table 8](#) shows the best practices that Fujitsu recommends for aggregate layout when you use FlexGroup volumes. Keep in mind that these practices are **not** hard requirements. The one-aggregate-per-node recommendation for the ETERNUS AX series systems originates from disk cost associated with RAID Triple Erasure Coding (RAID-TEC), because you might not want to use up expensive SSD space for the additional parity drives required for more than one RAID group. However, with ADP, partitions are spread across data disks, so in those cases, two aggregates per node on the ETERNUS AX series systems are better because there are more available [volume affinities](#) per node with two aggregates present.

Table 8 Best practices for aggregate layout with FlexGroup volumes

| Spinning Disk or Hybrid Aggregates | ETERNUS AX series |
|------------------------------------|--|
| Two aggregates per node | One aggregate per node (without ADP) Two aggregates per node (with ADP) |

Note

For consistent performance, aggregates should have the same number of drives and RAID groups across the FlexGroup volume.

For more information about aggregate layouts when dealing with existing FlexVol volumes, see ["Failure domains" \(page 44\)](#).

Deploying a FlexGroup volume on aggregates with existing FlexVol volumes

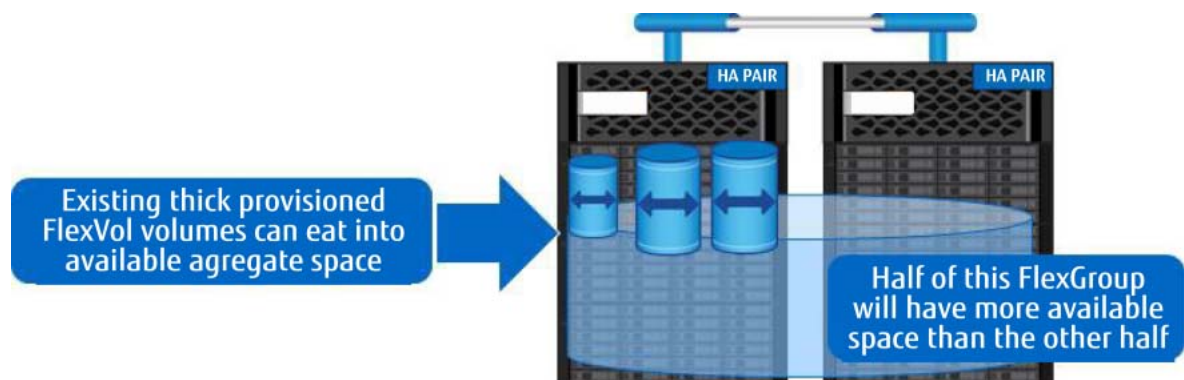
Because a FlexGroup volume can span multiple aggregates in a cluster and can coexist in the same SVM as normal FlexVol volumes, it is possible that a FlexGroup volume might have to share an aggregate with preexisting FlexVol volumes. Therefore, it is essential to consider the factors described in this section when you're deploying a FlexGroup volume.

Consider the capacity footprint of the existing FlexVol volumes

A FlexGroup volume can span multiple aggregates and each of those aggregates might not have the same number of FlexVol volumes on them. Therefore, the aggregates might have disparate free space that can affect the ingest distribution of a FlexGroup volume that has space guarantees disabled, because the existing FlexVol volume capacity might eat into the FlexGroup volume's capacity.

For example, if aggr1 on node1 has four FlexVol volumes at 1TB each and aggr2 on node2 has two FlexVol volumes at 1TB each, then node1's aggregate would have 2TB less space than node2. If you deploy a FlexGroup volume that spans both nodes and is overprovisioned to fill both aggregates, then node1's member volumes already have "space used" in their capacity reports, which would cause node2's members to absorb most of the ingest of data until the capacity used is even across all member volumes.

Figure 18 How FlexVol capacity can affect FlexGroup load distribution



Note

This is an issue only if the FlexGroup volume is thin provisioned. Space-guaranteed FlexGroup volumes would not have other volumes eating into the space footprint. However, space-guaranteed FlexGroup volumes might not be created as large as desired if other volumes in the system prevent the space from being allocated.

Consider the performance effect of the existing FlexVol volumes

When you deploy a FlexGroup volume, it is also important to consider the amount of work the existing FlexVol volumes are doing. If a set of FlexVol volumes on one node is being hit heavily at given times, that can negatively affect the performance of a FlexGroup volume that spans the same nodes and aggregates as the existing FlexVol volumes. This is similar to the effect that can be seen with FlexVol volumes, but because a FlexGroup volume can span multiple nodes, the performance effect might appear to be intermittent from the client perspective, depending on which node the data I/O is occurring.

One way to mitigate this effect is to make use of storage QoS policies to help limit IOPS and throughput to those volumes or guarantee performance with QoS minimums on the FlexGroup volume. Alternately, you can use non-disruptive volume move to redistribute the volumes across nodes to balance the performance effect.

Consider the volume count limits

ONTAP has volume count limits per node that depend on the type of node in use (either ETERNUS AX series or ETERNUS HX series) and the personality of the node. A system with the data protection personality allows more total volumes than a system without.

Additionally, there is a cluster-wide volume limit of 12,000 regardless of the node type in use that can affect how many FlexVol volumes can be provisioned per ONTAP cluster.

Because FlexGroup volumes generally contain multiple FlexVol member volumes, these member volumes count against this total limit. In addition, many ONTAP features also leverage FlexGroup volumes for their architectures. For instance, a single FlexGroup volume might use 16 member FlexVol volumes. If you use FlexClone for that volume, then you've used 32 FlexVol volumes. If you create a FlexCache volume in that cluster (which is also a FlexGroup), then you use whatever number ONTAP selects for that cache volume.

In most cases, you should not need to create multiple small FlexGroup volumes. Instead, provision a larger FlexGroup volume and use [qtrees](#) to separate workloads.

Best Practice 5: Deploying FlexGroup Volumes with Existing FlexVol Volumes in Place

Before deploying a FlexGroup volume:

- If you have existing FlexVol volumes, be sure to verify that adding multiple FlexGroup volumes and their corresponding features to the mix do not exceed the volume count limits.
 - Be sure to use the performance headroom features in Active IQ Unified Manager and ONTAP System Manager to review which nodes are being more heavily utilized.
 - If there is an imbalance, use nondisruptive volume moves to migrate "hot" volumes to other less-utilized nodes to achieve as balanced a workload across nodes as possible.
 - Be sure to evaluate the free space on the aggregates to be used with the FlexGroup volume and make sure that the available space is roughly equivalent.
 - If the effect of volume count limit is a potential factor, create the FlexGroup volumes across nodes that have room to add more new volumes, or use nondisruptive volume moves to relocate volumes and balance out volume counts.
 - Alternately, create FlexGroup volumes with fewer member volumes if volume count limits are a concern.
-

Flash Cache and Flash Pool

Flash Cache cards and Flash Pool aggregates are supported with FlexGroup volumes, but, if you choose to use them, be sure to have one on any node that participates in a FlexGroup volume for consistent performance results. Flash Cache cards are expected to provide the same performance benefits for FlexGroup volumes that they provide for FlexVol volumes.

Advanced disk partitioning

FlexGroup volumes have no bearing on the use of ADP. No special considerations need to be made.

SyncMirror (mirrored aggregates)

FlexGroup volumes can reside on aggregates that participate in a SyncMirror configuration, which is a way to replicate aggregates internally for extra data protection functionality. The FlexGroup should reside entirely on SyncMirror aggregates (for example, all member volumes are on SyncMirror aggregates or none are). Otherwise, the SyncMirror is not useful.

For more information about SyncMirror, contact Fujitsu Support.

Note

SyncMirror does not provide the same functionality as StrictSync (SnapMirror Synchronous). FlexGroup volumes currently do not support StrictSync or SnapMirror Synchronous. SyncMirror is more akin to MetroCluster. For the latest data protection information regarding FlexGroup volumes, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays Data Protection and Backup for ONTAP FlexGroup Volumes](#).

MetroCluster

ONTAP introduced support for FlexGroup volumes on MetroCluster deployments (IP).

MetroCluster software is a solution that combines array-based clustering with synchronous replication to deliver continuous availability and zero data loss at the lowest cost. There are no stated limitations or caveats for FlexGroup volumes with MetroCluster.

Note

When using Volume Encryption (VE) or Aggregate Encryption (AE) with MetroCluster, be sure to complete the MetroCluster configuration before enabling VE/AE.

Cloud Volumes ONTAP

ONTAP introduced official support for Cloud Volumes ONTAP—an ONTAP solution running in the cloud. You can now deploy a FlexGroup volume using Cloud Volumes ONTAP (CVO).

FlexGroup volumes running in Cloud Volumes ONTAP can use the same feature sets available in the ONTAP version deployed to the Cloud Volumes ONTAP instance. Some common use cases seen for Cloud Volumes ONTAP and FlexGroup include the following:

- Data lake for analytics
- EDA repositories for use with Amazon Elastic Compute Cloud (Amazon EC2) instances
- Data backup and archive for use with on-premises SnapMirror

Although FlexGroup volumes can support multiple petabytes in a single namespace for on-premises deployments, Cloud Volumes ONTAP instances max out at 368TB per instance and FlexGroup volumes cannot span more than one cluster instance. FlexGroup volumes in CVO can only be created by using the CLI or ONTAP System Manager. Currently, you cannot use Cloud Manager to create FlexGroup volumes.

Capacity considerations

Although FlexGroup allows massive capacity and file count possibilities, the FlexGroup volume itself is still limited to the physical maximums of the underlying hardware. The [current stated maximums](#) (20PB, 400 billion files) are only tested maximums; the [theoretical maximums](#) could go a bit higher, but the official supported member volume count in a FlexGroup volume currently stands at 200. If you require more than 200 member volumes in a FlexGroup volume, contact your sales representative or Fujitsu Support.

Also, there are node-specific aggregate size limitations that allow only a set number of 100TB FlexVol volumes. Be sure to review your hardware's physical capacity limitations for more information.

For example, the ETERNUS HX6100 allows 400TB aggregates, which means that we would see a maximum of four 100TB volumes allowed per aggregate.

However, Fujitsu recommends not reaching the 100TB limit for member volumes, because doing so would make it impossible to expand member volumes further in the future in the event a member volume runs out of space (you would have to add new 100TB member volumes to increase capacity in that case). Instead, aim to leave a cushion of no less than 10% to 20% of the total maximum FlexVol member space to provide for emergency space allocation features such as autogrow, elastic sizing, and proactive resizing to take effect.

These numbers are raw capacities before features such as Snapshot reserve, WAFL reserve, storage efficiencies and FabricPool cloud tiering are factored in. To correctly size your FlexGroup solution, contact Fujitsu Support to get assistance.

Maximums and minimums

This section covers the maximums and minimums that are specific to ONTAP FlexGroup volumes. [Table 9](#) lists the maximum values and shows whether the maximum is hard-coded/enforced or a recommended/tested value.

Table 9 FlexGroup maximums

| | Value | Value Type |
|---|----------------------|---------------------|
| FlexGroup volume size | 20PB | Tested/recommended* |
| File count | 400 billion | Tested/recommended* |
| Cluster node count | 24 (12 HA pairs) | Hard-coded/enforced |
| FlexVol member volume size | 100TB | Hard-coded/enforced |
| FlexVol member volume file count | 2 billion | Hard-coded/enforced |
| SnapMirror volume count (member per FlexGroup) | 200 | Hard-coded/enforced |
| SnapMirror volume count (FlexGroup total per cluster) | 6,000 | Hard-coded/enforced |
| File size | 16TB | Hard-coded/enforced |
| FlexVol member constituent count | 200 | Tested/recommended* |
| Aggregate size/count | Same as ONTAP limits | Hard-coded/enforced |

Table 10 FlexGroup minimums

| | Value | Value Type |
|------------------------|------------|---------------------|
| FlexVol member size | 100GB | Tested/recommended* |
| Data aggregate count | 1 | Hard-coded/enforced |
| SnapMirror schedule | 30 minutes | Tested/recommended* |
| Snapshot copy schedule | 30 minutes | Tested/recommended* |

Note

*Limits described as tested/recommended are tested limits based on a 10-node cluster. If allowed by the platform, actual limits are not hard-coded and can extend beyond these limits up to 24 nodes. For more information, refer to ["Theoretical or absolute maximums"](#). However, official support for the number of member volumes is 200. If you need to exceed this limit, contact your sales representative to start the qualification process for more member volumes.

Maximum number of FlexGroup volumes in a cluster

A FlexGroup volume can consist of a single FlexVol member volume or hundreds of FlexVol member volumes. The maximum number of FlexVol member volumes is physically constrained only by the total volume count in a cluster. As a result, a FlexGroup volume could [theoretically](#) have up to 24,000 member volumes in a 24-node cluster.

The total number of FlexGroup volumes is similarly constrained by the total volume count in a cluster. Each FlexGroup volume's member volumes are part of the volume count, so the number of FlexGroup volumes allowed in a cluster depends on the number of member volumes.

For example, a two-node cluster may have 2,000 volumes to work with. As a result, you could have one of the following configurations (although others are possible), all of which add up to 2,000 volumes:

- 10 FlexGroup volumes with 200 member volumes
- 20 FlexGroup volumes with 100 member volumes
- 40 FlexGroup volumes with 50 member volumes
- 200 FlexGroup volumes with 10 member volumes

Keep in mind that the existence of other FlexVol volumes in the cluster (including SVM root volumes) affects the total number of available member volumes. FlexVol member volume limits also can be constrained by the FlexGroup volumes participating in SnapMirror relationships. For the latest details on those limitations, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays Data Protection and Backup for ONTAP FlexGroup Volumes](#).

ONTAP, in most cases, creates multiple member volumes by default. For details on FlexGroup creation methods, refer to ["FlexVol member volume layout considerations"](#). If you create multiple FlexGroup volumes, you might unknowingly begin to use up the volume count in your cluster. In general, use the automated `volume create -auto-provision-as` CLI command to create new FlexGroup volumes rather than getting bogged down in the details of member volume counts. Overall, it is better to create fewer, larger FlexGroup volumes and divide workloads using qtrees. Qtrees in ONTAP 9.8 offer quota enforcement, granular statistics, and QoS policies (currently, qtree QoS is only supported for NFS).

Theoretical or absolute maximums

The stated supported limits for a FlexGroup volume are 200 constituent volumes, 20PB, and 400 billion files. However, these are simply the tested limits in a 10-node cluster. When you factor in the maximum volumes that are allowed per node in a cluster, the limits can potentially expand dramatically.

Ultimately, the architectural limitation for a FlexGroup volume is the underlying hardware capacities and the total number of allowed volumes in a single cluster.

Table 11 Theoretical maximums for FlexGroup based on allowed volume count in ONTAP

| Maximum Cluster Size | Current Architectural Maximum Member Volumes per Cluster (ONTAP 9.8) | Theoretical Maximum Capacity per FlexGroup Volume | Theoretical Maximum Inodes per FlexGroup Volume |
|----------------------|--|---|--|
| 24 nodes | 12,000 | ~1200PB (based on 100TB per member volume × 11,999 FlexGroup member volumes) | ~24 trillion inodes (based on 2 billion inodes × 11,999 FlexGroup member volumes) |

Note

The main limiting factor in the number of 100TB member volumes allowed in a cluster is the underlying physical hardware limitations, which vary depending on platform.

If you want to exceed the stated 20PB, 400 billion file, and 200-member volume limits, contact your sales representative to begin a qualification process.

FlexVol member volume layout considerations

FlexVol volumes are the building blocks of a FlexGroup volume. Each FlexGroup volume contains several member FlexVol volumes to provide concurrent performance and to expand the capacity of the volume past the usual 100TB limits of single FlexVol volumes.

Standard FlexVol volumes are provisioned from the available storage in an aggregate. FlexVol volumes are flexible and can be increased or decreased dynamically without affecting or disrupting the environment. A FlexVol volume is not tied to any specific set of disks in an aggregate and exists across all the disks in the aggregate. However, individual files themselves are not striped; they are allocated to individual FlexVol member volumes.

Figure 19 FlexVol and FlexGroup architecture comparison



Because of this architecture and the potential for [large files](#) to affect FlexGroup operations, there are some considerations you should keep in mind when provisioning a FlexGroup volume.

When designing a FlexGroup volume, consider the following for the underlying FlexVol member volumes:

- When you use automated FlexGroup creation methods such as `volume create -auto-provision-as flexgroup` or ONTAP System Manager, the default number of member FlexVol volumes in a FlexGroup volume depends on several factors covered in this section.

Note

For nearly all use cases, Fujitsu recommends that you let ONTAP determine the member volume count per node, provided you are creating larger FlexGroup volumes (10TB or greater). For smaller FlexGroup volumes, closer attention should be paid to the file sizes in the workload and the percentage of capacity per member volume they would potentially use.

FlexVol member volumes are deployed in even capacities, regardless of how the FlexGroup volume was created. For example, if an eight-member, 800TB FlexGroup volume was created, each member is deployed with 100TB. If a larger or smaller quantity of FlexVol member volumes is required at the time of deployment, use the `volume create` command with the `-aggr-list` and `-aggr-list-multiplier` options to customize the number of member volumes deployed per aggregate. Refer to ["When would I need to manually create a FlexGroup volume?"](#).

Deployment method #1: Command line

Using the ONTAP command line is the currently recommended way to deploy a FlexGroup volume. However, there are two different ways to do this from the CLI: manually and automatically. Both commands use the `volume create` command set.

■ Automated FlexGroup creation (auto-provision-as) – CLI

This is the preferred method for FlexGroup creation, as it combines ease of use with predictable deployment logic and warnings during creation to help prevent misconfigured FlexGroup volumes. To use the automated CLI method, run the `volume create -auto-provision-as flexgroup` command. By default, this command provisions a FlexGroup volume with the following parameters:

- N number of member volumes (100GB each; four per aggregate in the cluster, up to two aggregates per eight member volumes)
- Total size = 100GB per member * number of member volumes (16 member volumes = 1.6TB)
- All nodes and data aggregates in the cluster used (regardless of node or hardware type)

When you run the command, you see a warning that informs you of the configuration. Be sure to review that before typing `Y`. Are the member volumes the size you want? Are the listed aggregates correct and using the same media types?

In most cases, the default values specified with no options are not adequate. You likely want to specify aggregates or nodes to use in the deployment.

As such, there are some additional configuration option flags for FlexGroup auto provisioning.

```
-support-tiering -use-mirrored-aggregates
-encryption-type -nodes
-size -state
-policy -user
-group -security-style
-unix-permissions -junction-path
-comment -max-autosize
-min-autosize -autosize-grow-threshold-percent
-autosize-shrink-threshold-percent -autosize-mode
-space-guarantee -type
-percent-snapshot-space -snapshot-policy
-language -foreground
-nvfail -qos-policy-group
-qos-adaptive-policy-group -caching-policy
-encrypt -is-space-reporting-logical
-is-space-enforcement-logical -tiering-policy
-tiering-object-tags -analytics-state
```

Use those flags to customize your FlexGroup's size, tiering policies, space guarantees, nodes , aggregates and much more.

The following is the general behavior of the automated CLI commands:

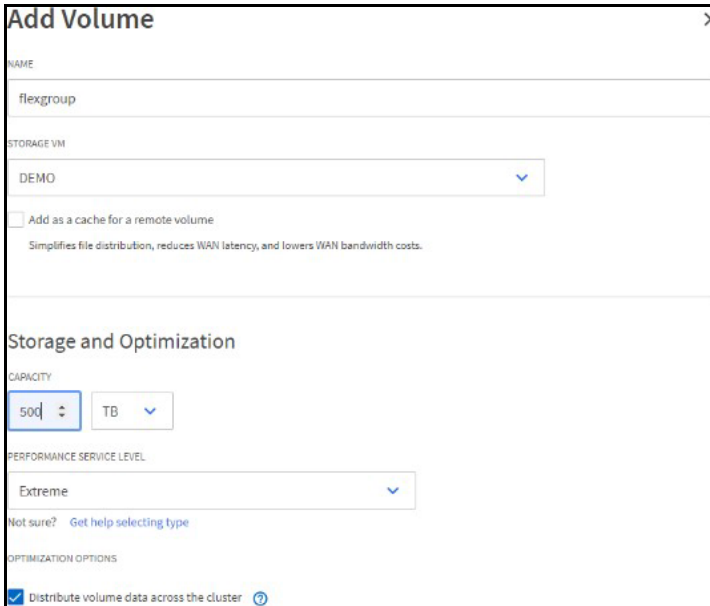
- Uses two aggregates per node, if possible. If not, use one aggregate per node.
- Uses the same number of aggregates on each node.
- The automated commands choose the aggregates that have the most free space.
- The automated commands create eight constituents per node if there are eight or fewer nodes.
- In clusters with more than eight nodes, scale back to four member volumes per node.
- Uses the fastest aggregates that you can. First try SSD, then Hybrid, and then spinning disk.
- CPU utilization, node performance, aggregate capacity, etc. are not currently considered.

■ Manual FlexGroup creation—CLI

The CLI also provides a more manual approach for creating FlexGroup volumes. In most cases, use the automated command, because it covers most use cases. However, if you need to customize the number of member volumes per aggregate, specify the aggregates to be used or have other reasons covered in "[When would I need to manually create a FlexGroup volume?](#)", you still use the `volume create` command. However, instead of the `-auto-provision-as` option, you must specify `-aggr-list` along with it. Specifying `-aggregate` creates a normal FlexVol volume and does not allow you to specify `-aggr-list`. To control the number of member volumes per aggregate, use `-aggr-list-multiplier`. Your member volume count is the number of aggregates you specified multiplied by `-aggr-list-multiplier`.

Deployment method #2: ONTAP System Manager

Figure 20 ONTAP System Manager FlexGroup volume creation



The screenshot shows the 'Add Volume' dialog box in ONTAP System Manager. The 'NAME' field is set to 'flexgroup'. The 'STORAGE VM' dropdown is set to 'DEMO'. There is an unchecked checkbox for 'Add as a cache for a remote volume'. The 'Storage and Optimization' section includes a 'CAPACITY' field set to '500' with a unit dropdown set to 'TB'. The 'PERFORMANCE SERVICE LEVEL' dropdown is set to 'Extreme'. Under 'OPTIMIZATION OPTIONS', the checkbox 'Distribute volume data across the cluster' is checked.

ONTAP System Manager has an easy-to-use GUI for volume creation. However, there are some caveats to consider when deploying with the System Manager GUI that make using the CLI a better choice when provisioning FlexGroup volumes.

To create a FlexGroup volume in ONTAP System Manager, the only thing you need to do to ensure the volume is a FlexGroup and not a FlexVol is to click More Options and select the Distribute Volume Data Across the Cluster box.

This tells System Manager to create a FlexGroup volume that spans multiple nodes; no aggregate or node specification is required.

System Manager deploys a FlexGroup volume according to the following rules.

- Member volumes are never smaller than 100GB
- The smallest allowed FlexGroup is 100GB (one 100GB member volume)
- Smaller FlexGroup volumes deploy fewer member volumes when necessary to adhere to the 100GB rule; for example, a 200GB FlexGroup deploys two 100GB member volumes.
- Aggregate and node selection for the FlexGroup is performed automatically. To specify nodes or aggregates, use the CLI or REST API.
- When a FlexGroup volume is large enough to accommodate, then the member volume count is capped to the number of volume affinities available per node.
- System Manager limits your initial FlexGroup size to the total space available as if space guarantees were enabled. You can go back into System Manager and grow the volume larger with the Edit functionality.
- System Manager only uses similar aggregates for the FlexGroup. In other words, it does not mix SSD and HDD aggregates.

Deployment method #3: REST APIs

With REST APIs, you can create, monitor, and manage FlexGroup volumes. To use REST APIs to provision a FlexGroup volume, use the same guidance as described in ["When would I need to manually create a FlexGroup volume?"](#).

Deployment method #1: Command line. For example, decide whether to let ONTAP choose the configuration or whether you should manually specify options.

You can find REST API documentation at `https://[your_cluster_IP_or_name]/docs/api`. This site provides examples and an interactive Try It Out feature that enables you to generate your own REST APIs.

For example, to create a FlexGroup volume, you can use the `POST` REST API under `/storage/volumes`. What makes a FlexGroup a FlexGroup (and not a FlexVol) in this call are one or a combination of the following values:

- **Aggregates**
If you specify more than one, then the REST API creates a FlexGroup volume. This is the same behavior as `-aggr-list` in the CLI.
- **constituents_per_aggregate**
Specifies the number of times to iterate over the aggregates listed with `aggregates.name` or `aggregates.uuid` when a FlexGroup volume is created or expanded. If a volume is being created on a single aggregate, the system creates a flexible volume if the `constituents_per_aggregate` field is not specified. If this field is specified, it creates a FlexGroup volume. If a volume is being created on multiple aggregates, the system always creates a FlexGroup volume. This is the same behavior as `-aggr-list-multiplier` in the CLI.
- **Style**
If you specify `style` as `flexgroup` and don't set the `constituents_per_aggregate` value or more than one aggregate, ONTAP automatically provisions a FlexGroup volume of four members per aggregate. This is the same behavior as `-auto-provision-as` in the CLI.

In the REST API documentation, the Try It Out functionality helps guide you as you try to create the correct REST API strings. When you make a mistake, the interface delivers error messages and a list of error codes. Also, a job string URL is given if the REST API command is correct but the job fails for another reason (such as creating a FlexGroup volume that has members that are too small). You can access the job string through the browser with the following URL:

```
https://[your_cluster_IP_or_name]/api/cluster/jobs/job_uuid]
```

This is what a failure message looks like:

```
{
  "uuid": "b5b04f0b-82ea-11e9-b3aa-00a098696eda",
  "description": "POST /api/storage/volumes/b5b02a66-82ea-11e9-b3aa-00a098696eda",
  "state": "failure",
  "message": "Unable to set parameter \"-min-autosize\" to specified value because it is too small. It must be at least 160MB (167772160B).",
  "code": 13107359,
  "start_time": "2019-05-30T10:53:39-04:00",
  "end_time": "2019-05-30T10:53:39-04:00",
  "_links": {
    "self": {
      "href": "/api/cluster/jobs/b5b04f0b-82ea-11e9-b3aa-00a098696eda"
    }
  }
}
```

This is what a successful job looks like:

```
{
  "uuid": "ac2155d1-82ec-11e9-b3aa-00a098696eda",
  "description": "POST /api/storage/volumes/ac2131c5-82ec-11e9-b3aa-00a098696eda",
  "state": "success",
  "message": "success",
  "code": 0,
  "start_time": "2019-05-30T11:07:42-04:00",
  "end_time": "2019-05-30T11:07:46-04:00",
  "_links": {
    "self": {
      "href": "/api/cluster/jobs/ac2155d1-82ec-11e9-b3aa-00a098696eda"
    }
  }
}
```

For a sample REST API string that creates a FlexGroup volume, refer to ["Command Examples"](#).

When would clients experience out of space errors?

Generally speaking, when a NAS client sees an `out of space` error, that means the volume is actually out of space and resulting `df` and `volume show` commands would confirm that.

However, getting an `out of space` error from a NAS storage system is not always straightforward, because it is a generic error from the server telling the client that there are no more available resources. There is no concept of an `out of inodes` error or `reached maximum directory size` in NFS or SMB, so ONTAP resorts to using the standard `out of space` error to let clients know that they cannot write more data.

With a FlexGroup volume, clients can also receive `out of space` errors when a member volume fills to 100%. However, in ONTAP 9.7 and later, this scenario is virtually nonexistent.

The following table shows situations in which `out of space` errors are seen, their causes, and how to address them.

Table 12 Situations in which you see out of space errors

| Situation | How to Identify and Resolve |
|---|---|
| Volume or aggregate has no available space to honor writes. | <ul style="list-style-type: none"> • <code>df</code> or <code>volume show-space</code> output from cluster CLI • View capacity from ONTAP System Manager • Active IQ Unified Manager alerts • EMS messages <p>Resolution: Add more capacity to volume. Add more disks to aggregate. Use FlexGroup volumes to scale across nodes.</p> |
| Quota limit reached | <ul style="list-style-type: none"> • <code>df</code> or <code>volume show-space</code> output from cluster CLI • View capacity from ONTAP System Manager • <code>quota report</code> output from cluster CLI • Quota report from ONTAP System Manager • Active IQ Unified Manager alerts • EMS messages <p>Resolution: Increase the quota limit or notify the client that they need to delete data to stay under their quota.</p> |
| Out of inodes | <ul style="list-style-type: none"> • <code>df</code> or <code>volume show-space</code> output from cluster CLI • View capacity from ONTAP System Manager • <code>df -i</code> or <code>volume show -fields files,files-used</code> command from cluster CLI • Active IQ Unified Manager alerts • EMS messages <p>Resolution: Increase the total files value in the volume. Refer to "High file count considerations" for more information.</p> |

| Situation | How to Identify and Resolve |
|-----------------------|---|
| Maxdirsize exceeded | <ul style="list-style-type: none"> • <code>df</code> or <code>volume show-space</code> output from cluster CLI • View capacity from ONTAP System Manager • <code>df -i</code> or <code>volume show -fields files,files-used</code> command from cluster CLI • Active IQ Unified Manager alerts • EMS messages • Client-side commands to view directory sizes (as shown in "Querying for used maxdirsize values") <p>Resolution: Use the <code>volume file show-inode</code> command from the cluster CLI to find the affected file path. Reduce the file count in the offending directory or contact support to verify if the <code>maxdirsize</code> value is safe to increase. For more information on <code>maxdirsize</code>, refer to "Directory size considerations: maxdirsize".</p> |
| Member volume at 100% | <ul style="list-style-type: none"> • <code>df</code> or <code>volume show-space</code> output from cluster CLI • Active IQ Unified Manager alerts • EMS messages • ONTAP version information <p>Resolution: ONTAP introduces Elastic sizing as a safeguard against file write failures when a member volume fills. ONTAP 9.8 introduces Proactive resizing to proactively resize member volumes to maintain an even balance of free space in member volumes. For best capacity usage results, upgrade to ONTAP 9.8 or later.</p> |

When would I need to manually create a FlexGroup volume?

In most cases, letting ONTAP choose the member volumes is the best option when creating a FlexGroup volume. In other words, don't worry about CPU count and volume affinity best practices for FlexGroup creation—let ONTAP do that for you. Trying to manipulate volume counts can lead to confusion and issues that may affect your FlexGroup later.

However, in some use cases, manual creation might be needed. The following sections describe scenarios in which you might need to manually create FlexGroup volumes.

■ Concern regarding overprovisioning volume counts

In ONTAP, each node and cluster has a finite number of FlexVol volumes allowed. The limits are dependent on platform and ONTAP version, but, because a FlexGroup volume is composed of multiple FlexVol volumes, those limits also apply to FlexGroup volumes, as each FlexVol member volume counts against the total volume count limit. If you have FlexGroup volumes with many member volumes or you want to create many FlexGroup volumes in a cluster, then you would need to consider the overall volume limits per node. You might also need to manually create the FlexGroup volumes to modify the default volume counts or aggregate placement to keep total FlexVol volume numbers below the node limits.

■ Large files with limited capacity

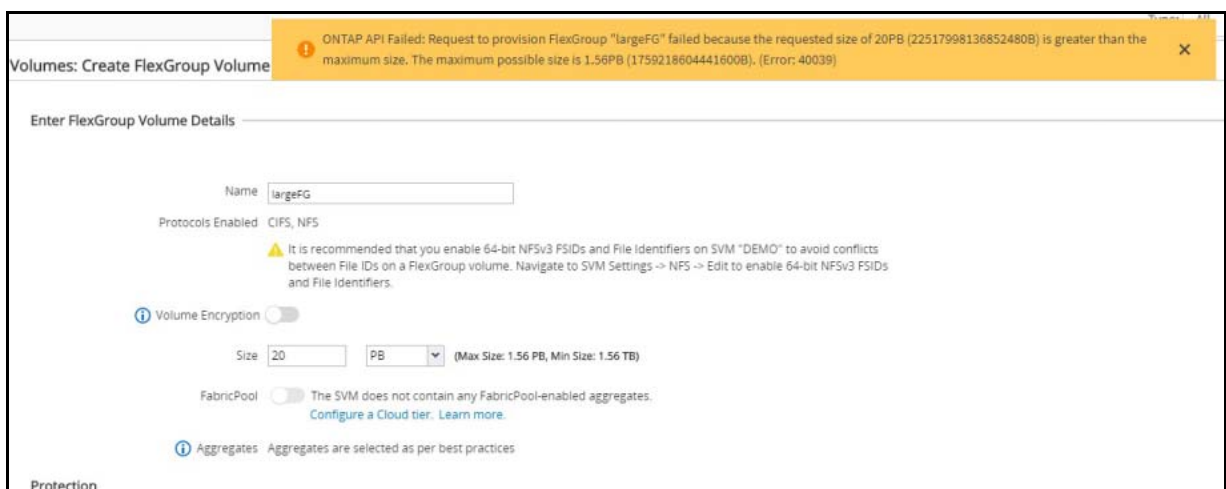
If you have a workload with larger files but cannot provision volumes that are tens or hundreds of TBs and you want to comply with [best practices for large files](#), you might need to adjust the member volume count to create fewer members at larger individual capacities.

For example, if you want to create a 16TB FlexGroup volume across four nodes, using the ONTAP default methods would create a minimum of 32 member volumes that are 500GB in size each. If your average file size is 250GB, then 500GB member volumes are not large enough to distribute the data effectively. Manually creating a FlexGroup volume with fewer, larger member volumes work s better for those use cases.

■ A need for a large amount of capacity or high file counts

FlexVol volumes are limited to 100TB in size and can contain up to two billion files. If you have a two- node cluster and you let ONTAP create a FlexGroup volume, you get at most 16 member volumes in a single FlexGroup volume in some cases, because it is code-limited to the best practice of eight per node. In the following example, the two-node cluster can only create a FlexGroup volume with a maximum of 1.56PB of capacity (eight members per node; 16 members total; 100TB per member volume).

Figure 21 Error when creating a FlexGroup volume beyond the allowed maximum in System Manager



The auto-provision-as option gives the same error:

```
cluster::*> vol create -vserver DEMO -volume largeFG -auto-provision-as flexgroup -size 2PB

Error: command failed: Request to provision FlexGroup volume "largeFGT" failed because the
requested size of 2PB (2251799813685248B) is greater than the maximum
possible size is 1.56PB (1759218604441600B).
```

If you desire a larger FlexGroup volume than what the automated tools allow, you need to create the FlexGroup manually to allow a higher number of member volumes by using the `-aggr-list-multiplier` option. For a 20PB FlexGroup volume, you would need at least 200 member volumes. Ideally, you would size the Flexgroup member volumes to a value less than 100TB, in case you later need room for those volumes to grow. 80-90TB should be the target maximum member volume value.

Similar considerations should be made if the file count needs to exceed the maximum files allowed. In the 16-member FlexGroup example, a maximum of 32 billion files is allowed. If more files are needed, increase the `maxfiles` value first. If that is not possible (for example, the `maxfiles` value is at the limit), then add more member volumes.

For an example of how to create a FlexGroup volume from the CLI and specify the number of members, refer to ["Command Examples"](#) later in this document.

■ Avoiding the cluster network

A less common scenario is the desire to avoid the cluster network by [creating a FlexGroup volume across a single node](#) or to reduce exposure to [failure domains](#). In this use case, use the ONTAP CLI to manage which aggregates are specified using the `-aggr-list` and `-aggr-list-multiplier` options with the `volume create` command.

■ Do I need a large number of member volumes?

Usually, you do not need to exceed the best practice volume count for a FlexGroup volume. However, if you need more capacity or higher file counts, you can increase the number of member volumes at initial deployment, or you can do so later by using the volume expand command. In general, it's better to increase member volume counts sooner than later, before data starts to fill the existing member volumes. Adding member volumes later creates an imbalance, which ONTAP must adjust for and it might affect workloads. For more information about when you might need to stray from ONTAP best practices for member volume counts, see the section above.

■ Member count considerations for large and small files

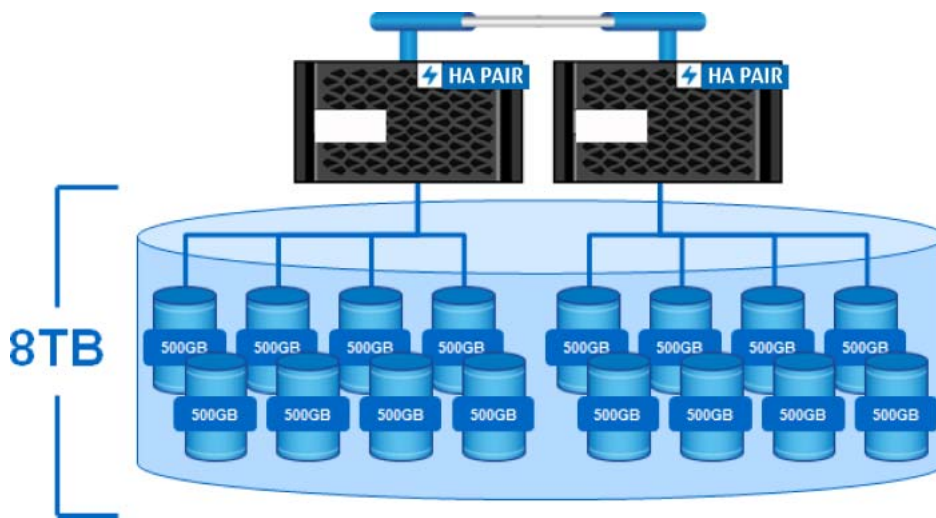
FlexGroup volumes work best in a high-file-count environment of many small files. However, they also work well with larger files. As mentioned in "[What are large files?](#)", large files should be considered in terms of percentage of the total space allocated to a member volume.

When larger files are present in a workload, the initial deployment size of a FlexGroup should be kept in mind. By default, a FlexGroup deploys eight-member volumes per node, so any capacity footprint that is defined at the FlexGroup level effectively gets divided into [total space/n number of member volumes].

For example, if an 8TB FlexGroup is deployed across two nodes in a cluster and the member count is 16, then each member volume is about 500GB in size.

In many workloads, the distribution shown in [Figure 22](#) would work well. However, if larger files in a workload would potentially fill in member volumes' large chunks of capacity used, then performance or even accessibility could be affected. In ONTAP 9.8, "Proactive resizing" helps make these "member volume full" scenarios less frequent.

Figure 22 How capacity is divided among member volumes



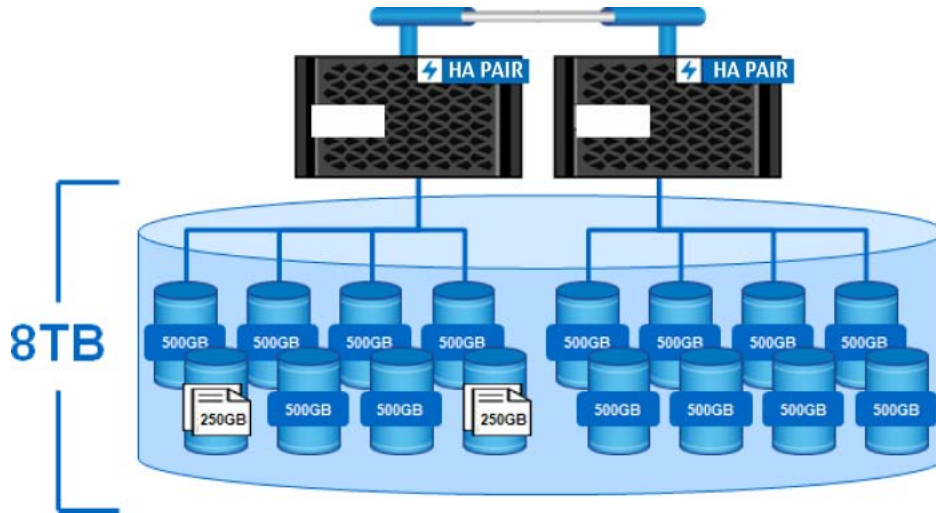
Best Practice 6: Best ONTAP Version to Run for Large File Workloads

In general, the latest ONTAP release available is the best version to run when using FlexGroup volumes, but we understand that not everyone can upgrade to the latest ONTAP release. Fujitsu recommends using the latest patch release of ONTAP 9.8. If that version is not possible, then use the latest patch release of ONTAP 9.7.

For example, if some files in a workload are 250GB, then each time a file is written to a FlexGroup volume with 500GB members, 50% of the total capacity of a member volume is filled.

If a second 250GB file attempts to write to that 500GB member volume, then the member volume runs out of available capacity before the file completes its write.

Figure 23 Effect of larger files in a FlexGroup member volume



Remember, files in a FlexGroup volume do not stripe; they always write to a single FlexVol member volume. Therefore, there must be enough space in a single member volume to honor the write.

[Elastic sizing](#) provides some relief by pausing before sending the client an out of space error and instead borrowing free space from other member volumes in the same FlexGroup if available. However, if volume auto-size is enabled, elastic resizing is disabled for that volume. Elastic resizing is not intended to be a way to avoid capacity management, but instead is a reactive way to reduce the effect of capacity issues. It's still imperative that the member volume capacities remain below 80-90% for best results.

Proactive resizing in ONTAP 9.8 combines the benefits of elastic sizing and volume autosize. Rather than waiting for a file creation to run out of space, ONTAP increases member volume sizes at a free space threshold proactively to reduce the effect of capacity imbalances and reduce the need to manage capacity from individual member volumes. In addition, volume autogrow can be used in conjunction with proactive resizing, so, if the total FlexGroup capacity is at a threshold, ONTAP automatically increases the size to a specified value.

Eventually, more storage must be added when physical space is exhausted. Also, adding nodes or disks to a FlexGroup volume is nondisruptive, easy, and fast.

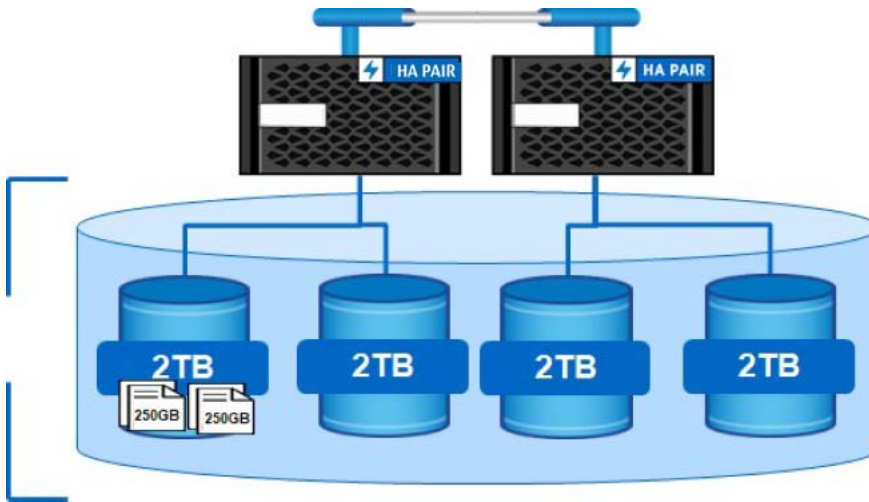
A better approach to sizing a FlexGroup volume is to analyze your workload and average file sizes before deploying a new FlexGroup volume or before allowing new workloads to access existing FlexGroup volumes. Fujitsu offers the XCP Migration Tool, which can quickly analyze files and report on sizes. For more information about XCP, refer to ["Migrating to ONTAP FlexGroup"](#).

After you have a good idea of what size files are going to land in a FlexGroup volume, you can make design decisions about how the volume should be sized at initial deployment.

Options include, but are not limited to the following:

- **Leave the member volume count at the defaults and grow the FlexGroup volume**
Size the total FlexGroup volume to a value large enough to accommodate member volume sizes that can handle the workload. In our example, the FlexGroup volume is 80TB, which provides 16-member volumes at 5TB per volume. However, this approach requires more physical capacity (unless you utilize thin provisioning).

Figure 24 Fewer, larger member volumes



- Manually reduce the member volume count and leave the FlexGroup capacity as is**
 Rather than accept the default values from the automated commands, you can use the CLI to create a FlexGroup volume that is identical in total capacity but contains fewer (but larger) member volumes. In our example, reducing the member volume count to two per node in an 8TB FlexGroup would provide member volume sizes of 2TB each. This would reduce the number of volume affinities available (and could reduce the overall performance of the FlexGroup volume for file ingest), but it would allow larger files to be placed.

After the large files are placed in member volumes, performance should be similar to what you would see from a FlexVol volume or a FlexGroup volume with more member volumes.

■ Capacity management features

The following table shows which capacity management features are available depending on your ONTAP release.

Table 13 Capacity Management Decision Matrix

| ONTAP Version | Capacity Management Features |
|---------------------|--|
| ONTAP 9.7 | <ul style="list-style-type: none"> Thin provisioning Capacity Alerting Storage efficiencies Volume autosize (autogrow/autoshrink) Qtrees and monitoring quotas Quota enforcement FabricPool autotiering Elastic sizing |
| ONTAP 9.8 and later | <ul style="list-style-type: none"> Thin provisioning Capacity Alerting Storage efficiencies Volume autosize (autogrow/autoshrink) Qtrees and monitoring quotas Quota enforcement FabricPool autotiering Elastic sizing Proactive resizing |

Aggregate free space considerations

When you create a FlexGroup volume, it is ideal for the aggregate (or aggregates) that the FlexGroup is deployed on to have the following characteristics:

- A roughly even amount of free space across multiple aggregates (especially important when using thin provisioning)
- At least 10GB or 0.6% free space (whichever is less)

ONTAP 9.7 and later versions no longer check for deduplication metadata.

■ Why Is member volume capacity important?

The goal of a FlexGroup volume is to manage it from the FlexGroup level, while not having to pay much attention to the underlying member volumes. In most cases, this is how FlexGroup volumes operate. However, in releases prior to ONTAP 9.8, member volume capacity had to be considered more often when creating and managing FlexGroup volumes.

Available free space in a member volume affects how often new files are ingested locally or remotely in a FlexGroup volume, which in turn can affect performance and capacity distribution in the FlexGroup volume for new file creation.

Average and largest file sizes in a workload are important to consider when you are designing an initial FlexGroup volume, because [large files](#) can fill up individual member volumes faster, causing more remote allocation of new file creations, or even causing member volumes to run out of space before the other member volumes do. Files that are already placed in the member volume remain in place, which means that, as they grow, the used capacity in that volume increases. However, already placed files generally do not see the same performance effect in unbalanced FlexGroups as new file creations do.

Capacity imbalances are not a problem in and of themselves, but they can be the root cause for performance issues or capacity utilization issues. If you have a performance issue and suspect capacity imbalance, open a Fujitsu support case for assistance in analyzing the performance data. In some cases, capacity imbalances might appear to be an issue, but the actual problem has a different root cause.

Best Practice 7: Member Volume Size Recommendations

Fujitsu recommends sizing a member volume such that the largest file does not exceed 1% to 5% of the member volume's capacity. Avoid creating FlexGroup volumes with fewer than two members per node.

Initial volume size considerations

A common deployment issue is undersizing a FlexGroup volume's member volume capacity. This is often done unbeknownst to the storage administrator because all they care about is the total capacity; they don't usually stop to think about underlying member volumes. To them, 80TB should be 80TB. But in a FlexGroup, 80TB is actually 80TB divided by the total number of member volumes.

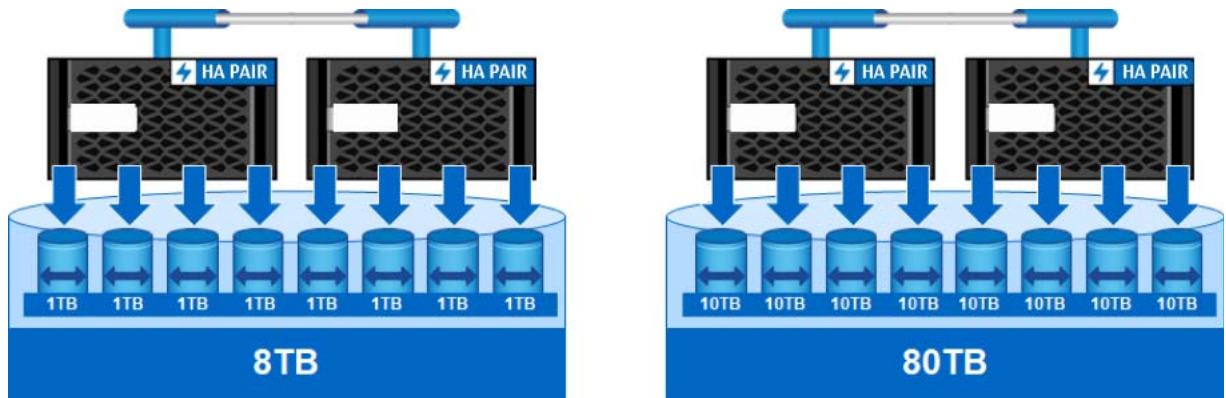
FlexGroup volumes can be created at almost any capacity, but it is important to remember that several FlexVol member volumes make up the total size of the FlexGroup volume. By default, automated FlexGroup commands create a default number of member volumes, depending on the deployment method used (refer to ["FlexVol member volume layout considerations"](#) for details).

Best Practice 8: Simplifying FlexGroup Deployment

If you want to stop worrying so much about member volumes, upgrade to ONTAP 9.8 to get the benefits of proactive resizing, which makes member volume sizes less of a concern.

For example, in an 80TB FlexGroup volume with eight member volumes, each member volume is 10TB in size. These member volume sizes are intended to be inconsequential in most workload cases since 10TB is a pretty large size to work with, but it's important to know what the file sizes of the workload are to help plan the capacity accordingly. For example, if you know your workload has 500GB files, then 10TB member volumes are fine, whereas 1TB member volumes would be problematic.

Figure 25 FlexGroup volumes—member sizes versus FlexGroup volume capacity



■ Initial space consumption on new FlexGroup volumes

Each FlexGroup member volume sets aside a small amount of space (around 50MB) for internal use. When a member volume is sized to the minimum of 100GB, the used space is around 0.05%, which is negligible to ONTAP. However, used space still shows up in the output of empty FlexGroup volumes, so this is something to keep in mind as a nonissue when deploying a FlexGroup volume.

For example:

```
cluster::*> vol show -vserver DEMO -volume fgautogrow* -fields used
vserver volume          used
-----
DEMO    fgautogrow__0006  57.48MB
DEMO    fgautogrow__0008  57.48MB
DEMO    fgautogrow__0001  57.50MB
DEMO    fgautogrow__0004  57.50MB
DEMO    fgautogrow__0005  57.52MB
DEMO    fgautogrow__0007  57.52MB
DEMO    fgautogrow__0002  57.57MB
DEMO    fgautogrow__0003  57.57MB
DEMO    fgautogrow         460MB
```

■ Shrinking a FlexGroup volume

Even with volume shrink support, avoid oversizing the volumes at the initial creation. If you make them too large, your administration options might be limited when you need to grow capacity later and you have to add new member volumes, because they would need to be added in identical member volume sizes.

Snapshot copies and snapshot reserve

ONTAP Snapshot copies are designed to create a point-in-time copy of a filesystem without using any space until data is overwritten. When data is changed or deleted, the space is marked as removed from the active file system (AFS) and ONTAP redirects pointers to the Snapshot copy. After this is done, the Snapshot copy shows the space used.

Snapshot Reserve

By default, volumes assign a 5% Snapshot copy reservation. This means that if you provision a volume that is 100TB, then 5% of that volume (5TB) is allocated for Snapshot copies. As a result, the volume size output (such as `volume show` and `df`) on the storage system show 95TB of usable space in that scenario, which is also what clients see as available space. In a FlexGroup volume, the snapshot reserve is set at the FlexGroup level, but it is applied to each member volume. Although the snap reserve on a 100TB FlexGroup volume might be 5TB, the individual member volumes would share that 5TB evenly. If there are eight member volumes, then each has 640GB of capacity reserved for Snapshot copies.

Snapshot Spill

If Snapshot copy use grows beyond the size of the snapshot reserve, then that space starts to use capacity from the AFS. ONTAP reports the space used by Snapshot copies as being greater than 100%, and the total used space in the volume increases, even if no physical data exists in the volume.

For an example, refer to ["Snapshot spill example"](#).

■ Snapshot spill and snapshot scanners

ONTAP performs periodic scans of volumes to check for used and changed blocks in WAFL. This is used to properly calculate the used space in the volume.

These scanners are low priority jobs that defer to production workloads, so the speed which they complete can depend on the load on the system. You can view these scanners as follows:

```
Mon Dec 21 19:24:30 -0600 [CLUSTER: scan_ownblocks_calc_wkr: wafl.scan.ownblocks.done:info]:  
Completed block ownership calculation on volume vol__0003@vserver:ebf4370e-208a-11eb-921e-  
d039ea2020c8. The scanner took 202 ms.
```

If volumes are small (for example, 100GB), they can fill faster. In these cases, it is possible for incoming data writes to perform faster than the scanners, especially on faster the ETERNUS AX series systems. As a result, the capacity reporting doesn't react in time for storage administrators to address capacity needs by adding space or deleting snapshots. In these cases, the volume might report `out of space` if volume autogrow is not enabled, because there is no space left to borrow from other member volumes.

■ Snapshot spill remediation tips

Snapshot spill is a normal function of how Snapshot copies work in ONTAP when the snapshot reserve is overrun. The effect of snapshot spill and how quickly it can grow depends on the total size of the volume. Smaller volumes have lower total snapshot reserve space and are more susceptible to snapshot spill. To minimize the effect of snapshot spill, you can do one or more of the following:

- Increase the total FlexGroup volume size, which also increases the total available snapshot reserve and makes snapshot spill less common.
- Avoid creating small FlexGroup volumes if you are using Snapshot copies.
- Use larger snapshot reservation percentages if you have more data churn; clients only see available space in the Active File System and do not see reserved snapshot space.
- Delete larger snapshots when possible. FlexGroup volumes do not currently support snapshot autodelete, so you need to delete Snapshot copies manually or via script.
- Set snap reserve to 0; this causes the snapshot used space to be reflected in the AFS immediately.

Volume autosize (autogrow and autoshrink)

Volume autogrow for FlexGroup volumes is supported. This support enables a storage administrator to set an autogrow policy for the FlexGroup volume that allows ONTAP to increase the FlexVol size to a predefined threshold when a volume approaches capacity. Applying volume autogrow to a FlexGroup volume is done in the same way as with a FlexVol volume; you specify thresholds and configure different options. Details can be found in the Fujitsu manual site.

The configuration options are the same as a FlexVol and include the following:

```
[-max-autosize {<integer>[KB|MB|GB|TB|PB]] - Maximum Autosize
This parameter allows the user to specify the maximum size to which a volume can
grow. The default for volumes is 120% of the volume size. If the value of this
parameter is invalidated by manually resizing the volume, the maximum size is reset
to 120% of the volume size. The value for -max-autosize cannot be set larger than
the platform-dependent maximum volume size. If you specify a larger value, the
value of -max-autosize is automatically reset to the supported maximum without
returning an error.

[-min-autosize {<integer>[KB|MB|GB|TB|PB]] - Minimum Autosize
This parameter specifies the minimum size to which the volume can automatically
shrink. If the volume was created with the grow_shrink autosize mode enabled, then
the default minimum size is equal to the initial volume size. If the value of the -
min-autosize parameter is invalidated by a manual volume resize, the minimum size
is reset to the volume size.

[-autosize-grow-threshold-percent <percent>] - Autosize Grow Threshold Percentage
This parameter specifies the used space threshold for the automatic growth of the
volume. When the volume's used space becomes greater than this threshold, the
volume will automatically grow unless it has reached the maximum autosize.

[-autosize-shrink-threshold-percent <percent>] - Autosize Shrink Threshold
Percentage
This parameter specifies the used space threshold for the automatic shrinking of
the volume. When the amount of used space in the volume drops below this threshold,
the volume will shrink unless it has reached the specified minimum size.

[-autosize-mode {off|grow|grow_shrink}] - Autosize Mode
This parameter specifies the autosize mode for the volume. The supported autosize
modes are:

o off - The volume will not grow or shrink in size in response to the amount of used
space.
o grow - The volume will automatically grow when used space in the volume is above
the grow threshold.
o grow_shrink - The volume will grow or shrink in size in response to the amount of
used space.

By default, -autosize-mode is off for new volumes, except for DP mirrors, for which
the default value is grow_shrink. The grow and grow_shrink modes work together with
Snapshot autodelete to automatically reclaim space when a volume is about to become
full. The volume parameter -space-mgmt-try-first controls the order in which these
two space reclamation policies are attempted.

[-autosize-reset [true]] - Autosize Reset
This allows the user to reset the values of autosize, max-autosize, min-autosize,
autosize-grow-threshold-percent, autosize-shrink-threshold-percent and autosize-
mode to their default values. For example, the max-autosize value will be set to
120% of the current size of the volume.
```


How volume autosize works in a FlexGroup volume

ONTAP pauses the operation briefly while it searches other member volumes for available free space. If there is available free space, then the member volume grows, while shrinking another member volume by the same amount; this maintains the same total FlexGroup volume size. This is known as Elastic sizing and is covered in more detail later.

If volume autosize is enabled, rather than pausing and borrowing space from another member volume and keeping the same total capacity (which adds some latency to the workload), volume autogrow instead grows member volumes by a configured capacity when a capacity threshold has been reached. This increases the total FlexGroup volume size by the amount the member volume grew.

For example, if you had a 10TB FlexGroup volume and a member volume automatically grew by 1TB, then you now have an 11TB FlexGroup volume with volume autogrow.

■ Volume autoshrink

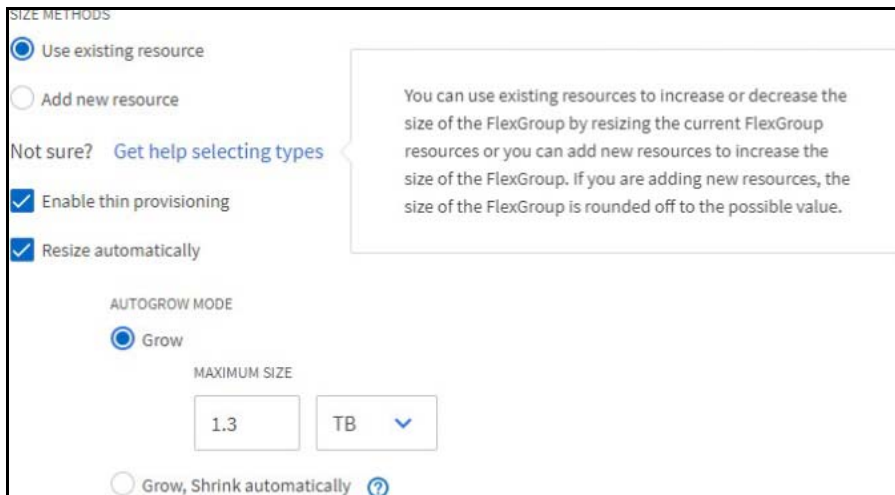
In addition to autogrow, the volume autosize feature also has an autoshrink functionality. This can be enabled or disabled via the `-autosize-mode` option. Autoshrink allows ONTAP to shrink a member volume back to a normal size if the capacity used reaches the configured `-autosize-shrink-threshold-percent` value. If that 11TB FlexGroup volume has 1TB free up in the member volume that grew, it would shrink by whatever you have configured, but no smaller than the original volume size by default.

■ How to enable volume autosize

Enabling volume autosize can be done in several different ways.

Procedure ▶▶▶ —————

- 1 You can use ONTAP System Manager during volume creation or use Edit:



The screenshot shows the 'SIZE METHODS' configuration window in ONTAP System Manager. It includes the following elements:

- SIZE METHODS:**
 - Use existing resource
 - Add new resource
- Not sure? [Get help selecting types](#)**
- Enable thin provisioning
- Resize automatically
- AUTOGROW MODE:**
 - Grow
 - Grow, Shrink automatically [?](#)
- MAXIMUM SIZE:**
 - Input field: 1.3
 - Unit dropdown: TB

A tooltip on the right side of the window reads: "You can use existing resources to increase or decrease the size of the FlexGroup by resizing the current FlexGroup resources or you can add new resources to increase the size of the FlexGroup. If you are adding new resources, the size of the FlexGroup is rounded off to the possible value."

- 2 You can use the `volume autosize` command via the CLI:

```
cluster::> volume autosize -server DEMO -volume fgautogrow -maximum-size 100g  
-grow-threshold-percent 80 -autosize-mode grow
```

- You can check that it's enabled with the following commands:

```
cluster::> vol autosize -vserver DEMO -volume fgautogrow
Volume autosize is currently ON for volume "DEMO:fgautogrow".
The volume is set to grow to a maximum of 100g when the volume-used space is
above 80%. Volume autosize for volume 'DEMO:fgautogrow' is currently in mode
grow.
```

3 You can use the `volume modify` command via the CLI:

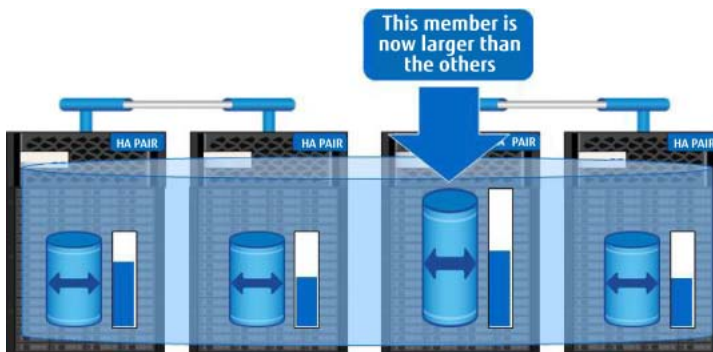
```
cluster::> volume modify -vserver DEMO -volume fgautogrow -autosize-mode grow_shrink
-autosize-grow-threshold-percent 95% -autosize-shrink-threshold-percent 50% -max-autosize
1.20PB -min-autosize 1PB
```

After a member volume has been grown through autogrow, there is an imbalance of member volume available size/allocation. This is by design.

In addition, the total FlexGroup size is now larger due to the larger member volume size. If you do not want the total FlexGroup volume size to grow, you can leave volume autogrow disabled and instead use the other capacity management feature in ONTAP, such as elastic sizing and proactive resizing (ONTAP 9.8 and later).



Figure 26 Member volume size allocation after a volume autosize operation



■ Volume autosize interaction with elastic sizing

[Elastic sizing](#) provides a way for file writes to complete in nearly filled member volumes by borrowing space from other member volumes. This takes place without growing the total size of the FlexGroup volume. As space is freed up in the filled member volume, elastic sizing begins to normalize the member volume sizes back to their original capacities.

Volume autosize on the other hand adds space to the total size of the FlexGroup volume by automatically growing a member volume when it reaches a space threshold.

Elastic sizing is enabled for FlexGroup volumes by default. If you enable volume autosize in ONTAP 9.7, elastic sizing no longer takes effect for that volume. ONTAP 9.8 and later enables the use of volume autosize with elastic sizing.

■ Volume autosize interaction with proactive resizing

Proactive resizing is available in ONTAP 9.8 and later and is covered in ["Proactive resizing"](#).

Volume autosize works in conjunction with proactive resizing. Proactive resizing adjusts member volume capacities, and, if a capacity threshold for autosize is reached, ONTAP applies volume autosize. If volume autosize is disabled, then proactive resizing works on its own. For more detail on how proactive resizing works with autosizing enabled, refer to ["Proactive resizing behavior - volume autosize enabled"](#).

Elastic sizing

Files written to a FlexGroup volume live in individual member volumes. They do not stripe across member volumes.

There are a few reasons why a member volume might fill up:

- You try to write a single file that exceeds the available space of a member volume. For example, a 10GB file is written to a member volume with 9GB available.
- If a file is appended over time, it eventually fills up a member volume—for example, if a database resides in a member volume.
- Snapshot copies eat into the active file system space available.

FlexGroup volumes do a good job of allocating space across member volumes, but, if a workload anomaly occurs, it can have a negative effect. For example, your volume is composed of 4,000 files but then a user zips some up and create a giant single tarball file.

One solution is to grow volumes, either manually or by using volume autogrow. Another solution is to delete data. However, administrators often don't see capacity issues until it's too late.

For example, a FlexGroup volume can be hundreds of terabytes in size, but the underlying member volumes and their free capacities are what determine the space available for individual files. If a 200TB FlexGroup volume has 20TB remaining (10% of the volume), the amount of space available for a single file to write is not 20TB; instead, it is closer to $20\text{TB}/[\text{number of member volumes in a FlexGroup}]$, provided all member volumes in the FlexGroup volume have evenly distributed capacities.

In a two-node cluster, a FlexGroup volume that spans both nodes is likely to have 16 member volumes. That means if 20TB are available in a FlexGroup volume, the member volumes would have 1.25TB available.

The elastic sizing feature helps avoid "out of space" errors in this scenario. This feature is enabled by default and does not require administrator configuration or intervention.

Elastic sizing is not a panacea; it is intended to be reactive to prevent file writes from failing. Capacity management to keep adequate member volume space available is still necessary, even with elastic sizing enabled.

■ Elastic sizing: an airbag for your data

One of our FlexGroup developers refers to elastic sizing as an airbag: it is not designed to stop you from getting into an accident, but it does help soften the landing when it happens. In other words, it's not going to prevent you from writing large files or running out of space, but it is going to provide a way for those writes to complete. In fact, in some cases, the peace of mind you get from elastic sizing can cause you to ignore capacity issues until the entire FlexGroup is out of space or until a performance issue occurs.

Here's how it works at a high level:

Procedure ▶▶▶ —————

- 1** When a file is written to ONTAP, the system has no idea how large that file will become. The client doesn't know. The application usually doesn't know. All that's known is "hey, I want to write a file."
- 2** When a FlexGroup volume receives a write request, it is placed in the best available member based on various factors, such as free capacity, inode count, time since last file creation, and so on.
- 3** When a file is placed, since ONTAP doesn't know how large a file will become, it also doesn't know if the file is going to grow to a size that's larger than the available space. So, the write is allowed as long as we have space to allow it.

- 4 If/when the member volume runs out of space, right before ONTAP sends an `out of space` error to the client, it queries the other member volumes in the FlexGroup volume to see if there's any available space to borrow. If there is, ONTAP adds 1% of the volume's total capacity in increments (in a range of 10MB to 10GB) to the volume that is full (while taking the same amount from another member volume in the same FlexGroup volume), and then the file write continues.
- 5 During the time ONTAP is looking for space to borrow, that file write is paused. This appears to the client as a performance issue, usually as latency. But the overall goal here isn't to finish the write fast—it's to allow the write to finish at all. Usually, a member volume is large enough to provide the 10GB increment (1% of 1TB is 10GB), which is often more than enough to allow a file creation to complete. In smaller member volumes, the effect on performance could be greater, because the system needs to query to borrow space more often due to smaller increments, and files don't have to be as large to fill the volume.
- 6 The capacity borrowing maintains the overall size of the FlexGroup volume. For example, if your FlexGroup volume is 40TB in size, it remains 40TB.
- 7 After files are deleted or volumes are grown and space is available in that member volume again, ONTAP re-adjusts the member volumes back to their original sizes to maintain an evenness in space, but only when a member volume's capacity is within 75% of the average free space of the other member volumes in the FlexGroup.



Figure 27 File write behavior before elastic sizing

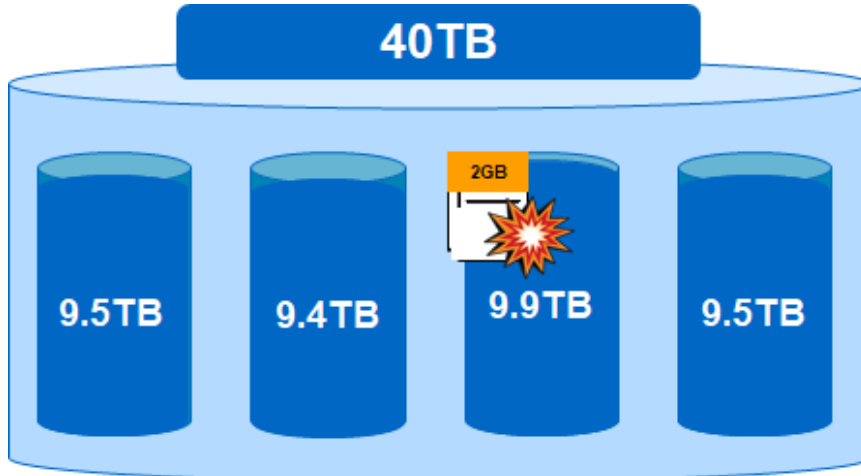
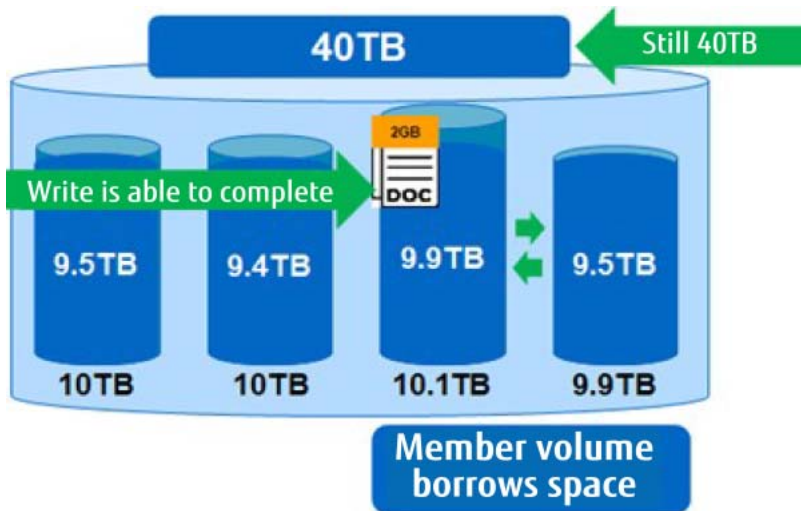


Figure 28 File write behavior after elastic sizing



Ultimately, elastic sizing helps mitigate file write failures in full member volumes and removes some the administrative overhead of managing capacity, because a full member is no longer an urgent event. ONTAP can still write files as long as there is available free space in other member volumes.

However, due to its reactive nature, it's best to upgrade to ONTAP 9.8 to make use of [proactive resizing](#).

■ When to use volume autogrow versus elastic sizing

When volume autogrow is enabled on a FlexGroup volume, elastic sizing is disabled for that volume in ONTAP 9.7, because the two features are essentially redundant.

However, there are some differences in how they work and when you'd want to use one over the other.

- Volume autogrow should be used when the total capacity of the FlexGroup volume can be grown to accommodate new data being written to it.
- Elastic sizing is enabled by default and should be used when the total size of the FlexGroup volume should not be allowed to grow past the specified capacity.

■ Performance effect of elastic sizing

Each time a file write must pause for ONTAP to find more space in the FlexGroup volume, client latency occurs. The amount of latency seen for a write operation to a file depends on the number of times the write must pause to find more space. For example, if a member volume has just 10GB available, but a 100GB file is being written, then elastic sizing causes the write to pause a number of times to allow the write to complete. That number is determined by the member volume total size, which can be anywhere from 10MB to 10GB.

The following example shows a test in which a file was copied to a FlexGroup volume. In the first test, the FlexGroup constituent was not large enough to hold the file, so elastic sizing was used. The 6.7GB file took around 2 minutes to copy:

```
[root@centos7 /]# time cp Windows.iso /elastic/  
real    1m52.950s  
user    0m0.028s  
sys     1m8.652s
```

7. Initial FlexGroup Design Considerations

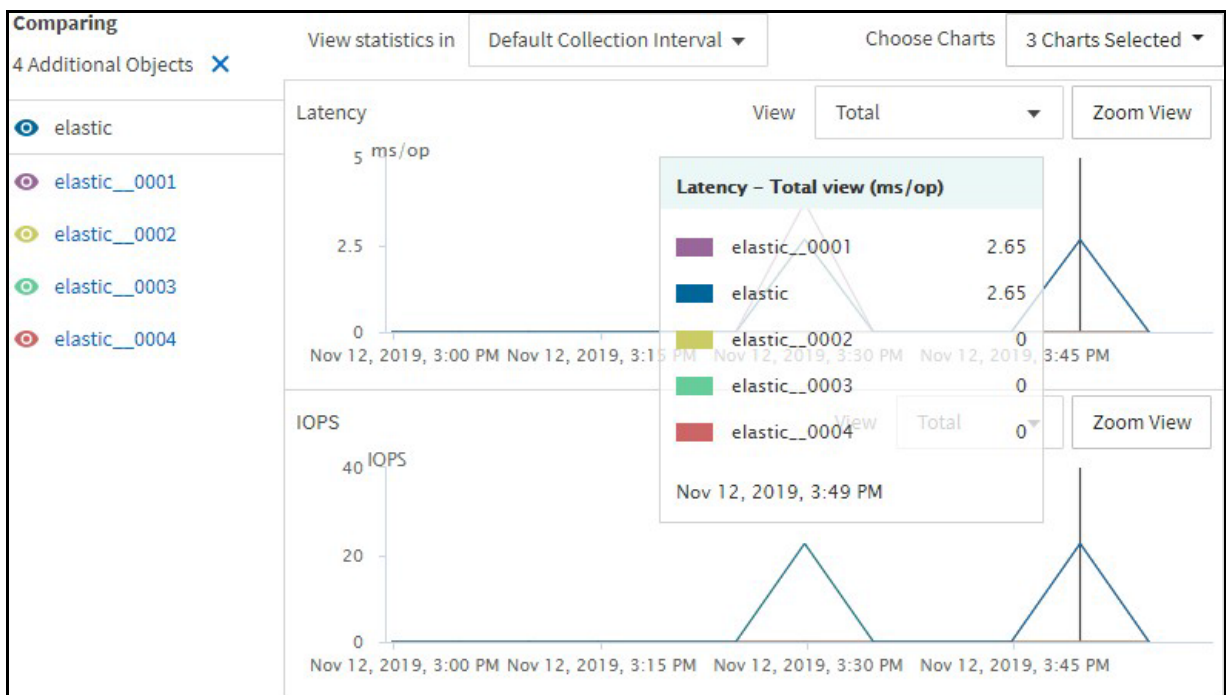
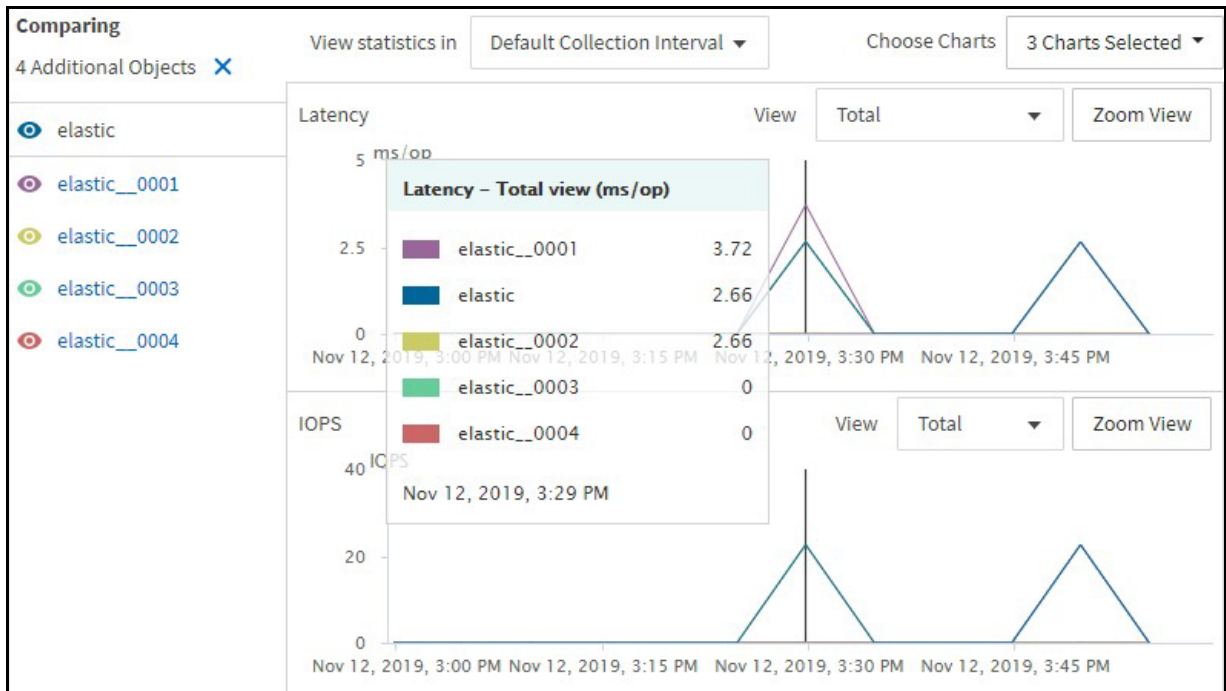
Capacity considerations

When the FlexGroup constituent volume was large enough to avoid elastic sizing, the same copy took 15 seconds less:

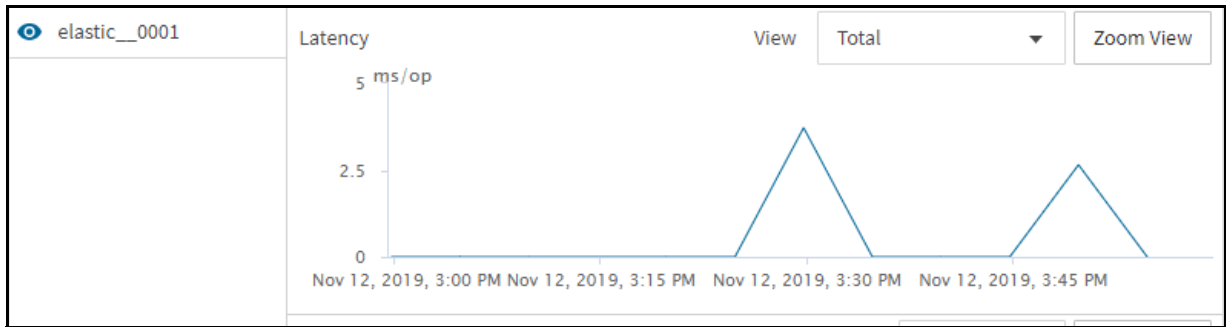
```
[root@centos7 /]# time cp Windows.iso /elastic/  
real    1m37.233s  
user    0m0.052s  
sys     0m54.443s
```

That shows there can be a real latency effect with elastic sizing.

The following graphs illustrate the latency hit on the constituent volume:



The constituent volume 0001 has about 0.5ms more latency when elastic sizing is in effect:



If you suspect that elastic sizing might be causing performance issues, you can do one of the following:

- Open a support case to confirm symptoms and logging.
- Grow the FlexGroup volume to make sure that there is enough space to remove elastic sizing from the equation.

ONTAP 9.8 and later introduces a new EMS event (`fg.member.elastic.sizing`) that lets you know that elastic resizing has occurred on a FlexGroup member volume.

However, with proactive resizing, every member volume resize event is considered elastic sizing and does not indicate a performance issue, but instead that the FlexGroup has some capacity issues that might need to be addressed by adding more space to the FlexGroup.

See an example of an elastic sizing EMS in "[Event management system examples](#)".

Keeping more than 20% available free space in a FlexGroup member volume is the ideal way to avoid the need for elastic sizing. This would require close management of capacity in ONTAP 9.7.

However, ONTAP maintains free space for you in ONTAP 9.8 with proactive resizing.

Proactive resizing

ONTAP 9.8 introduces a new feature for capacity management, with the goal of removing capacity management tasks from the storage administrator and instead letting ONTAP manage FlexGroup member volume capacity.

The following issues should be considered regarding proactive resizing:

- Member volumes remain the same size if the member volume capacity is less than 60%, even if there is a large capacity disparity.
- Proactive resizing adjusts member volume sizes at between 60% to 80% used capacity in small increments so as to maintain a relatively even balance of available space.
- After 80% used capacity, the goal is to maintain even capacity usage by adjusting the total member volume sizes up or down.
- When a resize occurs, it is not large; the range is between 10M and 10GB. But it also does not affect performance the way elastic sizing does, because there is no pause needed to check for free space. Resizing occurs before any capacity issues appear.
- Volume autosize is implemented if a member volume reaches the autogrow threshold, provided you have enabled volume autosize.

This free space buffer helps maintain even file ingest across the volume, reduces capacity imbalance, and improves capacity management for FlexGroup volumes in ONTAP.

■ Proactive resizing behavior – volume Autosize disabled

In the example below, a 400GB FlexGroup volume with four 100GB member volumes is used. A client creates 32 10GB files in the FlexGroup volume across four folders. The FlexGroup volume has volume autosize disabled. This means that the FlexGroup volume size we've specified remains that size, even if we reach 100% capacity.

At the start of the job after the first files are written, this is how the capacity balance appears:

Figure 29 Initial FlexGroup data balance – proactive resize, autosize disabled

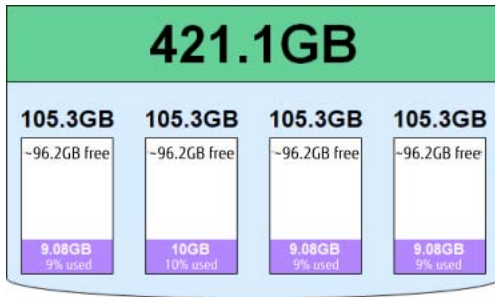
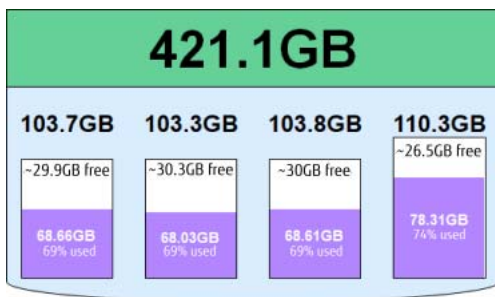
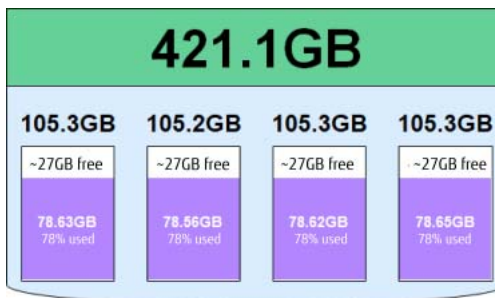


Figure 30 FlexGroup data balance, ~68% used – proactive resize, autosize disabled



At around 70% capacity usage, we can start to see the member volumes resize a bit to maintain a balanced free space, but the total capacity remains the same.

Figure 31 FlexGroup data balance, job complete – proactive resize, autosize disabled



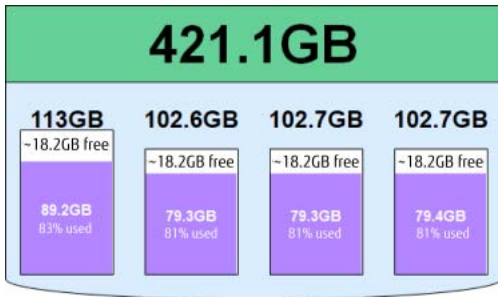
After the job finishes, ONTAP sees that the used space is even across all member volumes and proactive resizing shrinks the member volumes back down to their original sizes and makes them all the same because there is sufficient free space. The total FlexGroup size has not changed.

So, what happens when a new 10GB file is written after this?

When a file is written, it ends up in one of the member volumes. That creates a data imbalance, but ONTAP reacts accordingly by resizing the other member volumes to maintain an even amount of free space.

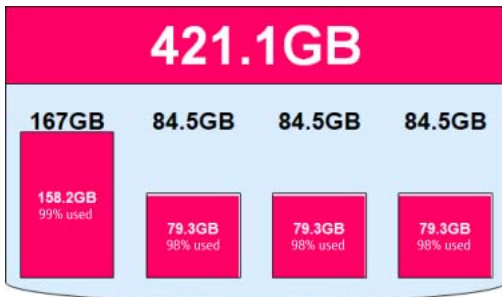
Here is the data balance after the new 10GB file is written:

Figure 32 FlexGroup data balance, new large file – proactive resize, autosize disabled



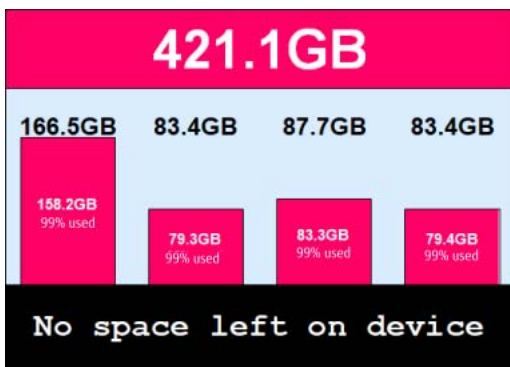
As you can see, the new file ended up in the first member volume. Elastic sizing grew that member volume to 113GB and shrunk the other member volumes while maintaining roughly the same amount of free space available and the total FlexGroup size.

Figure 33 FlexGroup data balance, 80GB file – proactive resize, autosize disabled



Then we write a new file to the FlexGroup again. This time, the file is too big to fit into a single member volume (80GB) and is almost too large to fit into the FlexGroup itself. When that happens, ONTAP uses proactive resizing, but we're not in a situation where every member volume has just 1GB of free space remaining. That means the next 10GB file will fail, because the entire FlexGroup is out of space and autosize is disabled.

Figure 34 FlexGroup data balance, out of space – proactive resize, autosize disabled



As a result, the next file creation fails, but proactive resizing becomes much more aggressive in adding free space to the member volume to avoid an out of space error. But when a FlexGroup volume itself is out of space, then the only remediation is growing the FlexGroup volume manually – or enabling volume autosize.

■ Proactive resizing behavior – volume autosize enabled

In the example below, a 400GB FlexGroup volume with four 100GB member volumes is used. A client creates 32 10GB files in the FlexGroup volume across four folders. The FlexGroup volume has volume autosize enabled with the default settings, which means the following:

- The FlexGroup volume maintains the same capacity, even if proactive resizing occurs, until the 92% used-space threshold is reached.
- After the used-space threshold is reached, the volume increases no more than 20%, as per the default settings. In this case, 566.7GB is the maximum size the volume grows, which is greater than 20% because this volume's size was increased and then later decreased.
- If the used capacity falls below 50%, then the volume shrinks back to the original size of 421.1GB.

These are the autosize settings for the FlexGroup:

```
cluster::> vol autosize -vserver DEMO -volume FG_SM_400G
Volume autosize is currently ON for volume "DEMO:FG_SM_400G".
The volume is set to grow to a maximum of 566.7g when the volume-used space is above 92%.
The volume is set to shrink to a minimum of 421.1g when the volume-used space falls below 50%.
Volume autosize for volume 'DEMO:FG_SM_400G' is currently in mode grow_shrink.
```

When a FlexVol or FlexGroup volume is smaller, the default growth threshold percentage is lower. For example:

- A 100GB volume has a default grow threshold of 90% and a shrink threshold of 50%.
- A 10TB volume has a default grow threshold of 98% and a shrink threshold of 50%.

At the start of the job (after the first files are written), this is how the capacity balance appears:

Figure 35 Initial FlexGroup data balance – proactive resize, autosize enabled

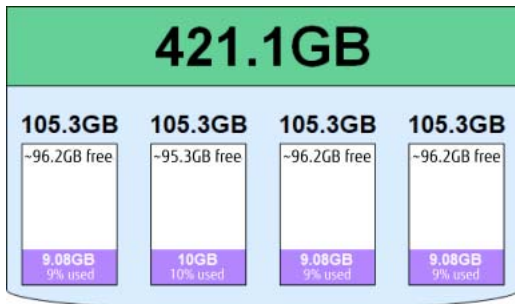
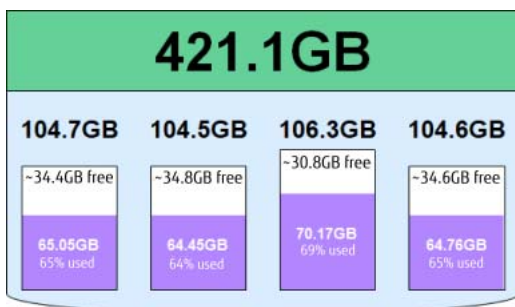
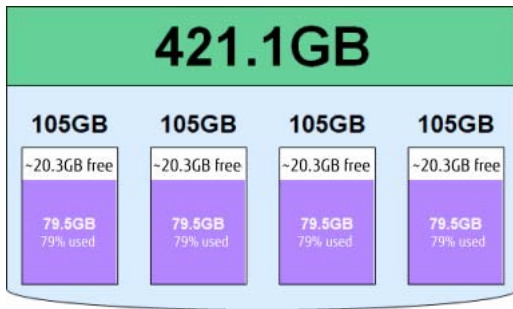


Figure 36 FlexGroup data balance, ~68% used – proactive resize, autosize enabled



At around 66% capacity usage, we can start to see the member volumes resize a bit to maintain balanced free space, but the total capacity remains the same, just like when autosize is disabled.

Figure 37 FlexGroup data balance, job complete – proactive resize, autosize enabled



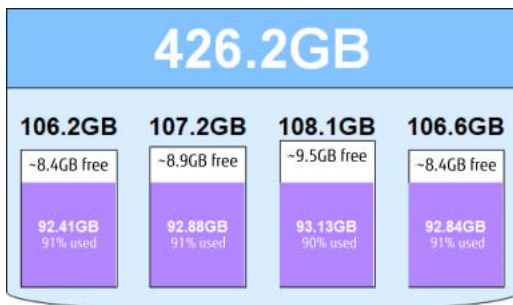
After the job finishes, ONTAP sees that the used space is even across all member volumes and proactive resizing shrinks the member volumes back down to their original sizes and makes them all the same size, because there is sufficient free space available. The total FlexGroup size has not changed.

As you can see, FlexGroup volumes with autosize enabled act just like FlexGroup volumes when autosize is disabled when the free space thresholds are below where autosize would kick in.

In the above graphic, we have roughly ~81GB free space available in the entire FlexGroup volume. If we keep writing 10GB files, we eventually reach the autosize threshold and ONTAP starts to react accordingly – this time with autosize growing the member volumes that need extra space, rather than by borrowing free space from other member volumes.

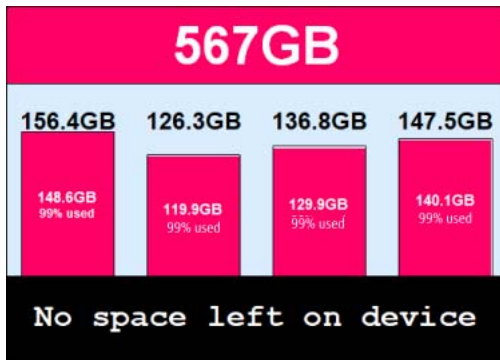
This results in the entire FlexGroup volume's capacity increasing. In the next test run, we created a new folder in the same FlexGroup and re-ran the test that creates 32 10GB files in the FlexGroup volume across four folders.

Figure 38 FlexGroup data balance, second test run – proactive resize, autosize enabled



After one of the FlexGroup member volumes reaches the 92% used-space threshold, autosize grows *only* that member volume. If other member volumes also need to be grown when they hit 92%, then those are also increased. This increases the overall capacity of the FlexGroup volume. Proactive resizing also adjusts the other member volume capacities up or down so that a relatively even amount of free space is available per member volume.

Figure 39 FlexGroup data balance, autosize limit – proactive resize, autosize enabled



Volume autosize defaults to only allow a volume to grow to 120% of the total volume size. Because the second test run needed 320GB of capacity to successfully complete and volume autogrow only allowed the volume to grow to 566.7GB total capacity, the job failed due to lack of space.

■ Autosize considerations: smaller FlexGroup volumes

Since autosize capacity is based on the percentage of total size, smaller FlexGroup volumes (such as a 420GB FlexGroup volume) have less runway for growth by default than a larger FlexGroup volume would. The default autogrowth maximum is capped to 120% of the total volume size. If the volume is ever grown manually and shrunk back down, then the autogrow value reflects the larger volume size.

Table 14 Autosize maximum size examples

| FlexGroup Volume Size | Default Maximum Autosize | Default Size Delta |
|-----------------------|--------------------------|--------------------|
| 420GB | 480GB | +80GB |
| 100TB | 120TB | +20TB |
| 400TB | 480TB | +80TB |

As a result, if you're using volume autosize for FlexGroup volumes, use the following guidance: Use larger FlexGroup volumes and maintain the default autosize values.

- If you use smaller FlexGroup volumes, modify the default `-max-autosize` value to avoid outages.
- If you don't want your end users to get more capacity than what you have provided, you can still use volume autosize if you use `qtrees` and `quotas` to limit the capacity seen and used by your end users.
- If you want to disable volume autosize, be aware that file writes fail when there is no more available space in the FlexGroup volume, even with proactive resizing in ONTAP 9.8.

Best Practice 9: Combining ONTAP Features for Capacity Management

The best way to approach capacity management involves a combination of larger FlexGroup volumes, volume autosize, ONTAP 9.8 or later, `qtrees` and quota enforcement. This story becomes more compelling when automatic tiering to cloud or S3 is performed using FabricPool. Using these features minimizes the capacity management overhead for storage administrators.

Networking considerations

When you use CIFS/SMB or NFS, each mount point is made over a single TCP connection to a single IP address in the cluster. In ONTAP, these IP addresses are attached to data LIFs, which are virtual network interfaces in an SVM.

The IP addresses can live on a single hardware Ethernet port or multiple hardware Ethernet ports that participate in a Link Aggregation Control Protocol (LACP) or another trunked configuration. However, in

ONTAP, these ports always reside on a single node, which means that they are sharing that node's CPU, PCI bus, and so on. To help alleviate potential bottlenecks on a single node, ONTAP allows TCP connections to be made to any node in the cluster, after which ONTAP redirects that request to the appropriate node through the cluster back-end network. This approach helps distribute network connections and load appropriately across hardware systems.

Best Practice 10: Network Design with FlexGroup

FlexGroup networking best practices are similar to FlexVol networking best practices. When you design a NAS solution in ONTAP, consider the following networking best practices regardless of the volume style:

- Create at least one data LIF per node, per SVM to confirm a path to each node.
 - Present multiple IP addresses to clients behind a single fully qualified domain name (FQDN) by using some form of DNS load balancing.
 - When possible, use LACP ports to host data LIFs for throughput and failover considerations.
 - When you manually mount clients, spread the TCP connections across cluster nodes evenly. Otherwise, allow DNS load balancing to handle the client TCP connection distribution.
 - For clients that do frequent mounts and unmounts, consider using on-box DNS to help balance the load. If clients are not mounted and unmounted frequently, on-box DNS does not help much.
 - If the workload is that of a "mount storm" (that is, hundreds or thousands of clients mounting at the same time), use off-box DNS load balancing and/or consider using FlexCache volumes. A mount storm to a single node can result in a denial of service to clients or performance issues.
 - If you're using NFSv4.1, consider leveraging pNFS for data localization and parallel connections to files. pNFS works best with sequential I/O workloads; high metadata workloads might bottleneck over the single metadata server connection.
 - If you have clients that support it, such as the latest SUSE and Ubuntu clients, the Nconnect mount option can provide even greater performance for NFS mounts on single clients.
 - For SMB3 workloads, consider enabling the multichannel and large MTU features on the CIFS server.
 - If you are using jumbo frames on your network, ensure jumbo frames are enabled at each endpoint in the network architecture; mismatched jumbo frame configurations can introduce hard-to-diagnose performance issues for any volume type.
 - NFS clients can get greater performance with multiple mount points from the same client connected to the same volume in ONTAP across multiple network interfaces. However, this configuration can introduce complexity. If your NFS client supports it, use Nconnect.
-

LACP considerations

There are valid reasons for choosing to use an LACP port on client-facing networks. A common and appropriate use case is to offer resilient connections for clients that connect to the file server over the SMB 1.0 protocol. Because the SMB 1.0 protocol is stateful and maintains session information at higher levels of the OSI stack, LACP offers protection when file servers are in an HA configuration. Later implementation of the SMB protocol can deliver resilient network connections without the need to set up LACP ports.

LACP can provide benefits to throughput and resiliency, but you should consider the complexity of maintaining LACP environments when you are deciding. Even if LACP is involved, you should still use multiple data LIFs.

DNS load-balancing considerations

DNS load balancing (both off-box and on-box) provides a method to spread network connections across nodes and ports in a cluster. FlexGroup volumes do not change the overall thinking behind DNS load balancing. Storage administrators should still spread network connections across a cluster evenly, regardless of what the NAS container is. However, because of the design of FlexGroup volumes, remote cluster traffic is a near certainty (pNFS data locality is the exception) when a FlexGroup volume spans multiple cluster nodes. Therefore, network connection and data locality considerations are nullified in those configurations. As a result, some form of DNS load balancing fits in a bit better with a FlexGroup volume, because worrying about data locality is no longer a factor. Ultimately, the decision of which method of DNS load-balancing to use comes down to the storage and network administrators' goals.

Best Practice 11: Use Some Form of DNS Load Balancing

When possible, use some form of DNS load balancing with FlexGroup volumes on nodes that contain FlexGroup member volumes.

■ On-box DNS or off-box DNS?

ONTAP provides a method to service DNS queries through an on-box DNS server. This method factors in a node's CPU and throughput to help determine which available data LIF is the best one to service NAS access requests.

- Off-box DNS is configured by way of the DNS administrator creating multiple "A" name records with the same name on an external DNS server that provides round-robin access to data LIFs.
- For workloads that create mount-storm scenarios, the ONTAP on-box DNS server cannot keep up and balance properly, so it's preferable to use off-box DNS.

Fujitsu recommends as a best practice creating at least one data LIF per node per SVM, especially when using a FlexGroup volume. Because of this, it might be prudent to mask the IP addresses behind a DNS alias through DNS load balancing. Then you should create multiple mount points to multiple IP addresses on each client to allow more potential throughput for the cluster and the FlexGroup volume.

Border Gateway Protocol (BGP)

ONTAP supports BGP to provide a more modern networking stack for your storage system. BGP support provides layer-3 (L3) routing, improved load-balancing intelligence, and virtual IPs (VIPs) for more efficient port utilization.

FlexGroup volumes need no configuration changes to use this new networking element.

Security and access control list style considerations

In ONTAP, you can access the same data through NFS and SMB/CIFS while preserving file ownership and honoring proper file permissions. This is known as multiprotocol NAS access. The same general guidance for multiprotocol NAS that applies to a FlexVol volume applies to a FlexGroup volume; these operate functionally the same for authentication and authorization. That guidance is covered in the product documentation in the "CIFS, NFS, and Multiprotocol Express Guides" and the "CIFS and NFS Reference Guides".

In general, for multiprotocol access, you need the following:

- Valid users (Windows and UNIX)
- Valid name-mapping rules or 1:1 name mappings through local files and/or servers such as LDAP or NIS. ONTAP uses name mappings to coordinate access for clients.
- Volume security style (NTFS, UNIX, or mixed). This can be configured for volumes or qtrees.

- A default UNIX user (pcuser, created by default for Windows to UNIX name mappings). Default Windows users (for UNIX to Windows name mappings) are not configured by default.

When a volume is created, a security style is applied. If you create a volume without specifying a security style, the volume inherits the security style of the SVM root volume. The volume security style determines the style of access control list (ACL) that is used for a NAS volume and affects how users are authenticated and mapped into the SVM. When a FlexGroup volume has a security style selected, all member volumes will have the same security style settings.

You can specify unique security styles in a FlexGroup volume by using qtrees.

Basic volume security style guidance

The following is some general guidance on selecting a security style for volumes:

- With the UNIX security style, Windows users must map to valid UNIX users. UNIX users only need to map to a valid user name if NFSv4.x is being used.
- In the NTFS security style, Windows users must map to valid UNIX users, and UNIX users must map to valid Windows users to authenticate. If a valid UNIX user name exists, NFS clients see proper ownership on files and folders. Authorization (permissions) is handled by the Windows client after the initial authentication. If no valid UNIX user exists, the default UNIX user (pcuser) is used for authentication/UNIX ownership.
- The UNIX security style allows some Windows clients to modify basic mode bit permissions (ownership changes, rwx). However, it does not allow NFSv4.x ACL management over SMB, and it does not understand advanced NTFS permissions.
- A mixed security style allows permissions to be changed from any type of client. However, it has an underlying effective security style of NTFS or UNIX, based on the last client type to change ACLs.
- A mixed security style requires proper name mapping to function properly due to the changing effective security styles.
- If granularity of ACL styles in a FlexGroup volume is desired, consider deploying qtrees. Qtrees allow you to set security styles per logical directory in ONTAP. If you want other home directory features such as FPolicy, anti-virus, native file auditing, and quota enforcement, then use the most recent patched release of ONTAP.
- NFSv4.x and NFSv4 ACL support for FlexGroup volumes was added in ONTAP 9.7.

Best Practice 12: Volume Security Style: Mixed Security Style Guidance

Fujitsu recommends a mixed security style only if clients need to be able to change permissions from both styles of clients. Otherwise, it's best to select either NTFS or UNIX as the security style, even in multiprotocol NAS environments.

More information about user mapping, name service best practices, and so on, can be found in the product documentation. You can also find more information in [FUJITSU Storage ETERNUS AX series All-Flash Arrays](#), [ETERNUS HX series Hybrid Arrays](#) [How to Configure LDAP in ONTAP Multiprotocol NAS Identity Management](#).

■ Changing the security style of a FlexGroup volume

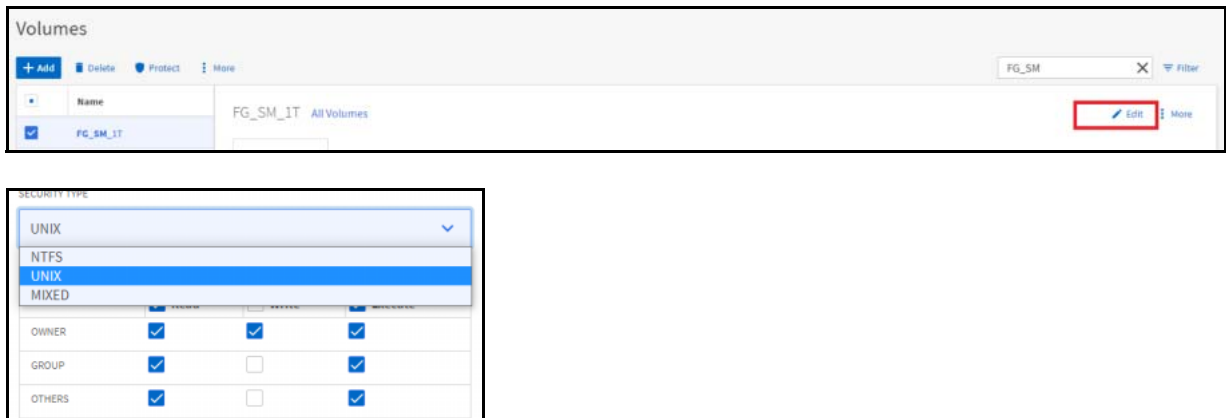
FlexGroup volumes are intended to be managed similarly to FlexVol volumes; changing the security style of volumes is included in that philosophy. Volume security styles can be changed live, with no need for clients to remount. However, the subsequent change in ACL styles means that access permissions might become unpredictable. For the best possible results, Fujitsu recommends changing security styles in a maintenance window on production datasets.

To change the security style of a FlexGroup volume, do one of the following:

- Use `volume modify` from the command line.
- Use the Edit button or Advanced Features when initially creating the FlexGroup volume in ONTAP System Manager.

7. Initial FlexGroup Design Considerations
Security and access control list style considerations

Figure 40 Modifying FlexGroup volume security styles in ONTAP System Manager



■ Using NFSv4.x ACLs with NFSv3 clients

In ONTAP, it is possible to leverage the benefits of NFSv4.x's granular ACL support even if your clients are only using NFSv3, provided you are using an ONTAP release that supports NFSv4 ACLs with FlexGroup volumes by using an administrative client that has mounted the export via NFSv4. By setting the NFSv4 ACLs from that client, clients accessing from NFSv3 honor those NFSv4 ACLs without needing to use NFSv4.x mounts.

8. FlexGroup Administration Considerations

This chapter covers general FlexGroup volume administration considerations, including tasks such as viewing FlexGroup volumes, volume moves, resizing FlexGroup volumes, renames, and so on.

Note

The ONTAP System Manager examples in this document use the new version of System Manager available in ONTAP 9.7 and later.

Viewing FlexGroup volumes

FlexGroup volumes can be created through the ONTAP GUI or through the command line and are designed to be managed, from a storage administrator's perspective, like a regular FlexVol volume. Things such as Snapshot copies, resize, and storage efficiency policies are all managed from the FlexGroup volume level.

However, the FlexGroup volume is not just a FlexVol volume; instead, it is made up of a series of FlexVol member volumes that act in concordance across the FlexGroup volume. ONTAP uses these member volumes for ingest of data to provide automated load balance and parallel operations across the file system, which provides capacity and performance gains.

In most cases, a FlexGroup volume can be managed at the FlexGroup level. For instance, when growing a FlexGroup volume, you can use the GUI or run the `volume size` command at the FlexGroup level to increase the total volume size. ONTAP makes sure that all underlying member volumes are given equivalent capacities, so the storage administrator doesn't have to think about how to distribute capacity. In ONTAP 9.8 and later, capacity management at the FlexGroup level is further simplified with the new [proactive resizing](#) functionality.

In rare cases, you may want to view individual FlexVol member volumes for capacity and performance concerns. These tasks are more commonplace in earlier releases of ONTAP.

Two scenarios where viewing member volumes might be useful:

- To view the member capacity usage (are we getting close to full in a single member volume?)
- To view individual member performance (do I need to use `volume move`?)

The following sections offer guidance on viewing FlexGroup volumes.

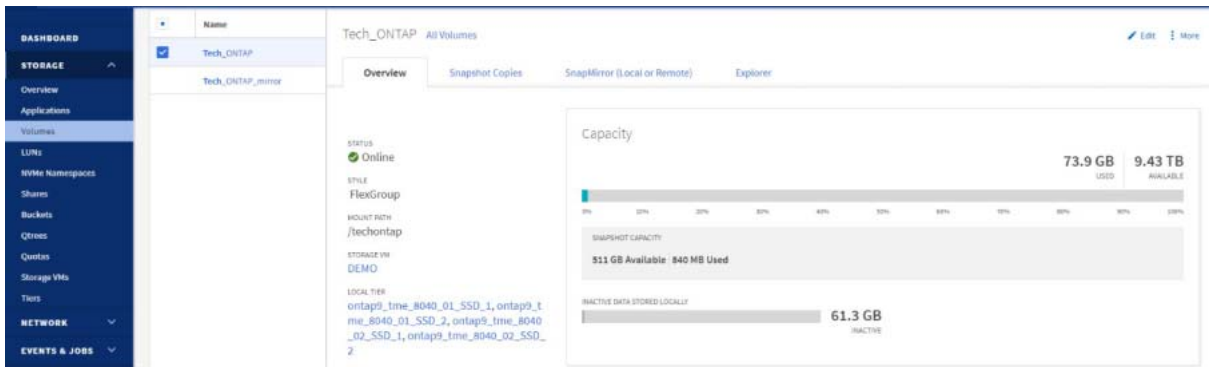
ONTAP System Manager

With ONTAP System Manager, you can view and manage a FlexGroup volume at the FlexGroup level through the FlexGroup tab; however, there are no views for member volumes. This is by design—a FlexGroup volume should be simple to manage. ONTAP System Manager provides useful information about the FlexGroup volume in these views, such as data protection information, real-time performance, and capacity information.

Note

Keep in mind that ONTAP System Manager cannot provide space allocation information for FlexGroup volumes that are thin-provisioned.

Figure 41 ONTAP System Manager FlexGroup volume view



Like a FlexVol volume, you can manage basic tasks from the System Manager GUI, such as snapshots, resizing, and data protection.

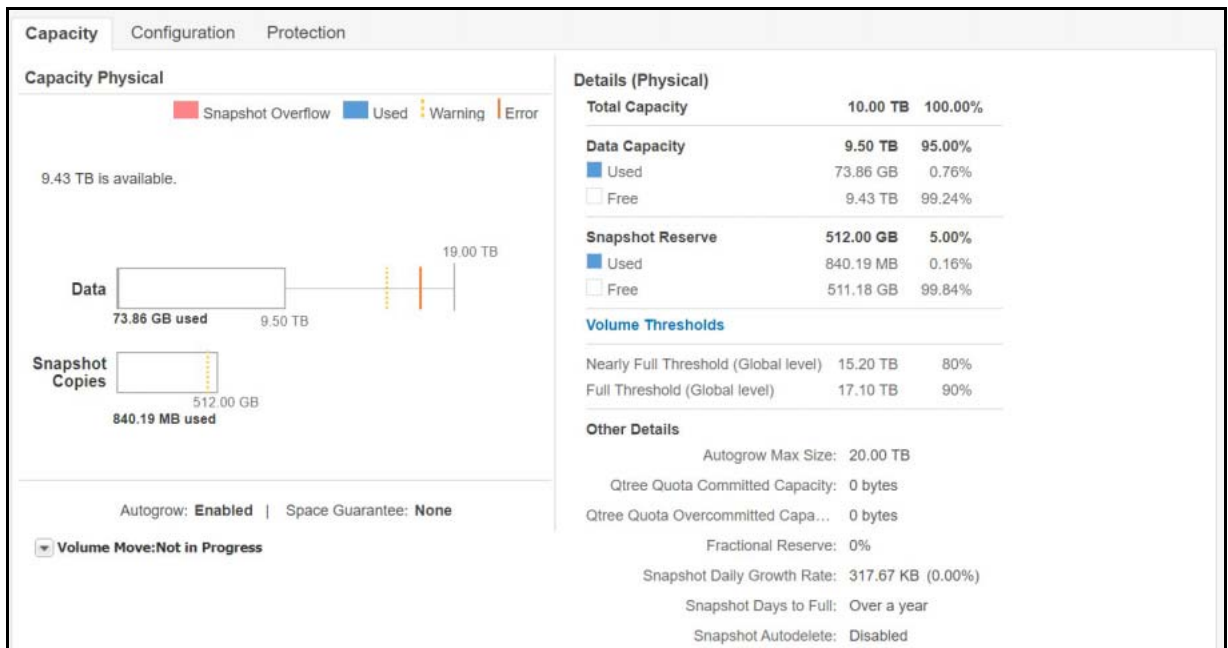
Active IQ Unified Manager

With Active IQ Unified Manager, storage administrators can use a single dashboard to review the health and performance of a ONTAP cluster.

With Active IQ Unified Manager, you can review FlexGroup volume capacity, configurations, and storage efficiencies in a graphical format. FlexGroup volume capacity in Active IQ Unified Manager is done from the FlexGroup level and does not show individual member volume capacities.

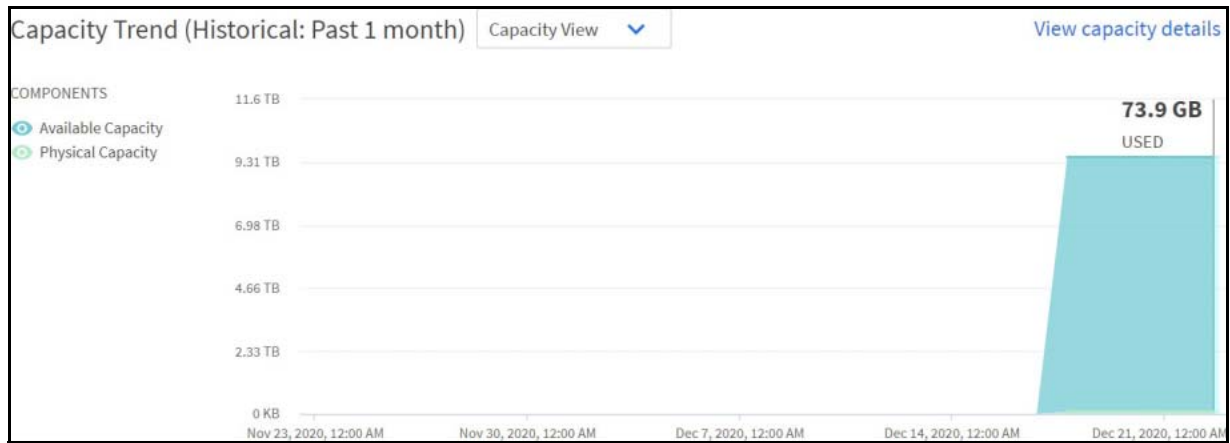
For member volume information about performance, use Active IQ Performance Manager. For capacity information about member volumes, use the command line.

Figure 42 Active IQ Unified Manager; FlexGroup capacity view



Active IQ Unified Manager can also show capacity trends for your volumes via Workload Analysis.

Figure 43 Active IQ Unified Manager Capacity Trend



Note

Currently, you cannot use Active IQ workloads with FlexGroup volumes.

■ Performance monitoring

Active IQ Unified Manager collects an archive of performance statistics for ONTAP, including the FlexGroup as a whole and its member volumes. A granular view of the FlexGroup volume allows storage administrators to evaluate individual member FlexVol volumes for performance anomalies and to take corrective actions as needed, such as the following:

- Adding more space
- Adding more members (`volume expand`)
- Nondisruptive volume move to less allocated nodes

Note

These tasks cannot be carried out in Active IQ. Currently, only the command line and/or the ONTAP System Manager GUI can carry out these tasks.

8. FlexGroup Administration Considerations
Viewing FlexGroup volumes

Figure 44 shows several FlexVol members and their corresponding performance. Each line represents a FlexVol member.

Figure 44 Active IQ Performance Manager FlexGroup volume view

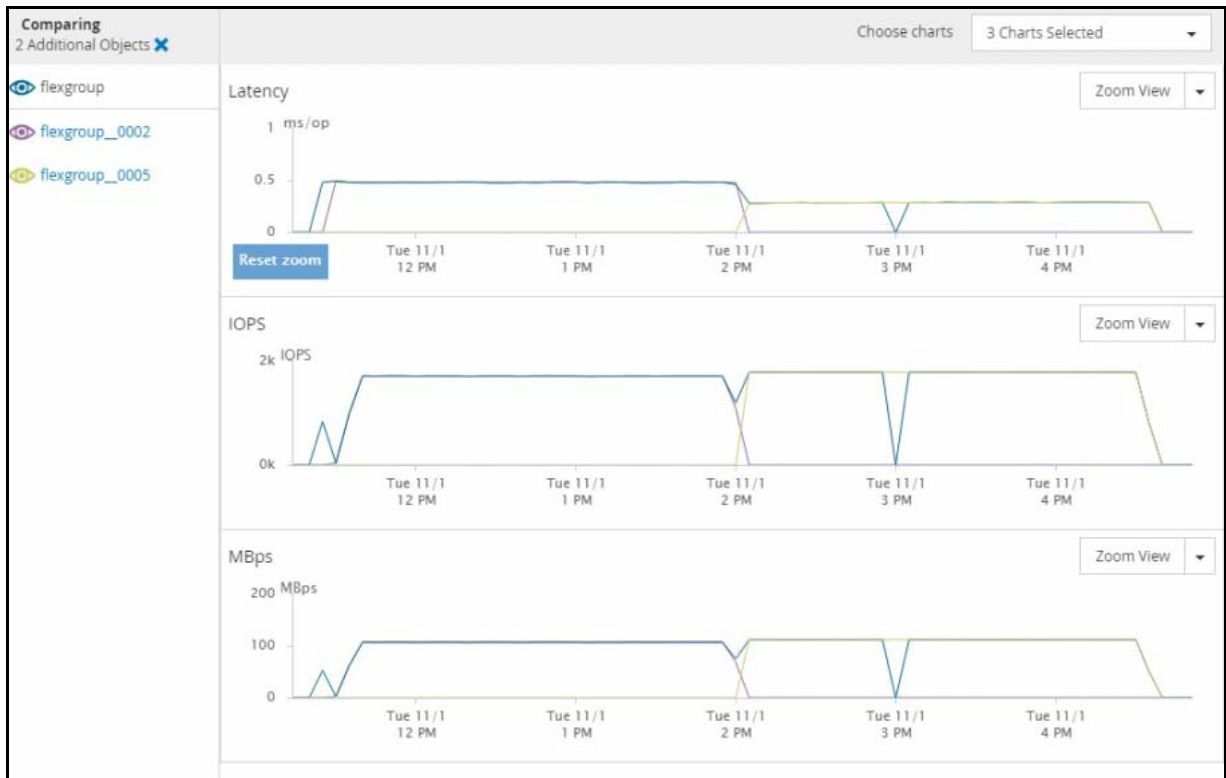


In Figure 45, two 1TB files were written to a FlexGroup volume. In the chart, we can see which member volumes took on that workload (members 2 and 5), and we see a summary of the workload performance. In Figure 46, we can see the IOPS and MBps graphs.

Figure 45 Member volume performance chart

| Volume | Latency | IOPS | MBps | |
|----------------|-------------|------------|-----------|-------|
| flexgroup_0001 | 4 ms/op | < 1 IOPS | 0 MBps | Add → |
| flexgroup_0002 | 0.481 ms/op | 1,581 IOPS | 98.8 MBps | Add → |
| flexgroup_0005 | 0.287 ms/op | 1,743 IOPS | 109 MBps | Add → |
| flexgroup_0006 | N/A | N/A | N/A | Add → |
| flexgroup_0004 | N/A | N/A | N/A | Add → |
| flexgroup_0003 | N/A | N/A | N/A | Add → |
| flexgroup_0008 | N/A | N/A | N/A | Add → |
| flexgroup_0007 | N/A | N/A | N/A | Add → |

Figure 46 Member volume graphs



Command Line

The CLI is another way to view FlexGroup volume information. Each privilege level gives a different set of options for viewing the FlexGroup volume properties.

■ Admin privilege level

- Total capacity (total, available, and used: calculated from all the member volumes), storage efficiencies
- Snapshot reserve or Snapshot policy
- List of aggregates and nodes that the FlexGroup volume spans
- Volume style and extended volume style (tells us whether the volume is a FlexGroup volume)
- Security style, owner, or group
- Junction path
- Maximum files and inodes
- Member volume information (through `-is-constituent true` or `volume show-space`)

■ Advanced privilege level

- Maximum directory size value
- FlexGroup master set ID (MSID)
- Whether the volume was transitioned from 7-Mode (important for FlexVol to FlexGroup volume conversion)
- FlexGroup maximum member volume sizes

■ Diag privilege level

- Detailed member volume information (capacity, used, and so on)
- FlexGroup ingest statistics (`flexgroup show`)

Note

Member volume space information can be seen in the admin privilege level by using the command `volume show-space`. For examples, refer to "[Capacity monitoring and alerting with the command line](#)".

Viewing FlexGroup volume capacity

This section covers various methods for monitoring a FlexGroup volume's capacity, including viewing total storage efficiency savings. Monitoring FlexGroup capacity is also possible with the FPolicy support.

Note

In ONTAP 9.8 and later, capacity views should be focused to the total FlexGroup volume rather than the underlying member volumes, as proactive resizing handles member volume free space balancing. However, if you need to view member volumes, the following sections still apply.

Total FlexGroup capacity

The total FlexGroup capacity is a number that is derived from the following:

- **Total space**
Total combined allocated space for a FlexGroup volume (member volume capacity * number of members).
- **Available space**
The amount of space that is available in the most allocated member volume.

When you provision a 10TB FlexGroup volume, clients see 10TB. ONTAP sees 10TB divided by the number of member volumes created. In most cases, it is not necessary to think about this, especially with [proactive resizing](#) in ONTAP 9.8. However, with smaller FlexGroup volumes and/or larger files, these calculations can become more important.

You can view the total FlexGroup capacity in ONTAP System Manager, in Active IQ Unified Manager, or through the CLI at the admin privilege level.

Overprovisioning or Thin Provisioning in a FlexGroup Volume

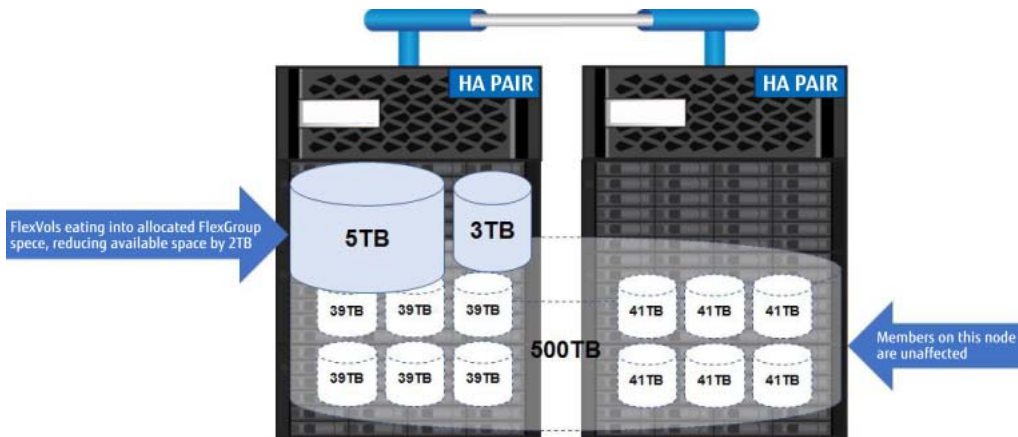
Overprovisioning or thin provisioning with a FlexGroup volume can be useful in scenarios for which you want to minimize capacity management. You can create large FlexGroup volumes that don't take up actual capacity until data is written and instead rely on the physical available space in an aggregate.

Thin provisioning should be used with the following caveats in mind:

- When a volume is out of space, it is truly out of space because the physical capacity has been used. More disk space must be added to remediate space issues, whether on the same node or by adding new nodes in the cluster.
- The space allocated does not necessarily reflect the actual space available. It is possible to allocate volumes that are much larger than their physical space.

- If your system is sharing aggregates with other volumes (FlexGroup or FlexVol volumes) that are thick provisioned, you should use thin provisioning with caution. Existing FlexVol or FlexGroup volumes on the same aggregates as FlexGroup volumes can potentially affect how data is ingested. Existing volumes reduce the amount of space available for member volumes because used space eats into other volume allocations.

Figure 47 Capacity effect when thin-provisioned FlexGroup volumes exist with space-guaranteed FlexVol volumes



Using volume space guarantees can protect against other datasets affecting volume capacities, but they do not offer the most efficient use of your storage capacity.

Best Practice 13: Using Thin Provisioning

If you use thin provisioning with ONTAP, it is important to use tools such as Active IQ Unified Manager or to set up monitoring through the CLI to track the available space in your storage system.

Adding capacity to a FlexGroup volume

A FlexGroup volume can grow to immense capacities, but, as data grows, even a massive container such as a FlexGroup volume might require more capacity.

In some cases, FlexGroup performance can be adversely affected as the capacity of member volumes becomes closer to full due to an increase in remote hardlink creation across member volumes. For more information about capacity, free space, and their effect on FlexGroup volumes, refer to "[Capacity considerations](#)".

As a result, maintaining at least 10% free space in a FlexGroup member volume is generally a good practice. Naturally, that 10% has different meaning depending on the size of the member volumes and file sizes, so be sure to follow the recommendations on provisioning volumes that have [large files or files that grow](#) in this document.

Note

ONTAP 9.8 and later greatly reduces the management overhead for member volume capacities with proactive resizing and other features, so, if possible, move to that release when using FlexGroup volumes.

Recommendations for adding capacity

There are two main ways to add capacity to a FlexGroup volume, in order of preference:

- Grow existing member volumes by using the `volume size` command.
- Add new member volumes by using `volume expand`.

The preferred method for adding capacity to an existing FlexGroup volume is to [grow the FlexGroup volume](#). Since data written to a FlexGroup is static and does not redistribute to other member volumes after it is placed in the FlexGroup, growing the FlexGroup in-place maintains a more consistent level of performance and capacity distribution than adding new member volumes to the workload.

If this approach is not possible because of physical aggregate limitations or the member FlexVol volumes approaching the 100TB limit, or if you are adding new nodes to the cluster and want to balance the workload out more, then you should add new member volumes using `volume expand` instead.

If you are adding new nodes or aggregates to use with the FlexGroup, it might make more sense to first use `volume move` to balance the member volumes across the nodes and then either grow them or add new member volumes in the same multiples per node. Refer to ["Adding disks, aggregates, and nodes"](#) for a visualization of the volume move approach.

Best Practice 14: Increasing Volume Size

- If possible, increase capacity through volume size or resize from ONTAP System Manager rather than adding new members; this approach preserves the existing FlexGroup data balance.
- If you are adding new nodes or aggregates to use with the FlexGroup volume, use `volume move` to rebalance the members across the new hardware and then either grow FlexGroup by using `volume size` or add new members with `volume expand` in the same multiples per node. For example, if each node has four members, add four new members per node.
- Avoid running `volume size` commands on FlexGroup member volumes individually without the guidance of Fujitsu Support. Run `volume size` only on the FlexGroup volume itself.
- Upgrade to ONTAP 9.8 or later to gain the benefits of [proactive resizing](#) and use volume autosize to avoid needing to manage capacity manually.
- Use capacity monitoring to keep track of how full member volumes are becoming. Set warnings at threshold percentages based on the total member volume capacity to give you ample time to address any issues (for example, 80% may be fine to 10TB, but not for 100GB).
- Use thin provisioning to set higher virtual caps of space on volumes without affecting total space allocation.

Note

There currently is no way to reduce capacity by removing member volumes. Only shrinking a volume is allowed.

Growing the volume versus adding new members

If a FlexGroup volume requires capacity or increased file count, there are two main approaches.

■ Growing the FlexGroup volume (volume size)

You can add capacity to existing member volumes by growing the total size of the FlexGroup volume. You do this in the CLI by using the `volume size` command or in ONTAP System Manager. The added size is divided evenly across the member volumes in the FlexGroup volume. For instance, if you add 8TB to a FlexGroup volume with eight member volumes, each member volume grows by 1TB. Therefore, when you add space, it is important to know how many member volumes are in the FlexGroup volume. To find this number, use `volume show -name [flexgroup] -is-constituent true` from the CLI.

In deciding whether to grow the FlexGroup volume or add member volumes, consider your desired result and intent. Grow the FlexGroup volume if:

- You simply want more capacity on existing nodes or aggregates.
- You do not want to increase the total volume count in your cluster.
- Your FlexGroup member volumes are nowhere near the 100TB limit.
- You are not at the two billion file limit for the member volumes.
- You have available physical space where the member volumes currently live.

- You want to preserve the data balance across the member volumes.

These scenarios are by no means exhaustive; there might be other instances in which you want to grow a FlexGroup volume instead of adding new member volumes. If you are unsure, contact Fujitsu Support.

■ Adding member volumes (volume expand)

Another way to add capacity or file count to a FlexGroup volume is by adding member volumes. Currently, Fujitsu officially supports up to 200-member FlexGroup volumes, which offers a maximum of 20PB capacity and 400 billion files. If you need more capacity or higher member volume counts than that, contact your sales representative to begin a qualification process for larger FlexGroup volumes.

To add more member volumes, you must currently use the `volume expand` command in the CLI. This command adds new, empty member volumes of exactly the same size as the existing FlexGroup member volumes. The number of new member volumes is determined by the `aggr-multiplier` and `aggr-list` options in the `volume expand` command. For an example of this command, refer to "[Example of expanding a FlexGroup volume](#)".

Adding new member volumes is not the preferred way to add capacity or file counts in most cases, because the new member volumes will be empty and can throw off the ingest balance of new requests. If you are adding new member volumes, be sure to add them in multiples—preferably the same number as the volumes that already exist in the system.

However, in some use cases, adding member volumes is the best way to add capacity to a FlexGroup volume. For example:

- The FlexGroup member volumes are already at or near 100TB.
- New nodes or aggregates are being added to the cluster.
- More maxfiles are needed and the member volumes are already at their [maximum values for their sizes](#).
- The cluster capacity is at a level where member volumes cannot be grown, and other nodes in the cluster have only enough space for member volumes of the same capacity.

These scenarios are by no means exhaustive; there might be other instances where you want to add members to a FlexGroup volume instead of growing the volume. If you are unsure, contact Fujitsu Support.

■ Volume resizing

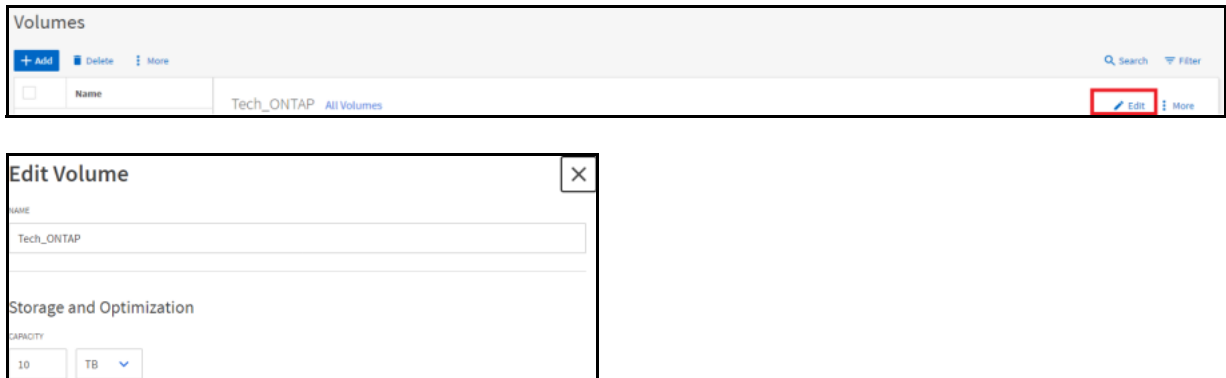
To grow or shrink the volume capacity as necessary, you can run the `volume size` command on the FlexGroup volume or use ONTAP System Manager. Adding capacity affects the ingest heuristics favorably because the member volumes have more available free space, so ONTAP favors local placement of files to a parent folder for better performance.

When you use this command, the member FlexVol volumes are each increased by the total capacity/total number of member volumes. For example, if a FlexGroup volume has eight member volumes and is grown by 80TB, then each member volume increases by 10TB automatically by ONTAP.

In releases prior to ONTAP 9.8, it was important to consider these individual increases when the total FlexGroup size increase is factored in. [Proactive resizing with volume autogrow](#) reduces the importance of considering individual member volume sizes, because ONTAP manages the free space available.

Resizing a volume from ONTAP System Manager

To resize a volume in ONTAP System Manager 9.7 and later, select the volume you want to resize and click Edit. Then type in the new size.



Volume Expand

You can grow FlexGroup volumes nondisruptively with volume size or you can add more capacity dynamically by using the volume expand command, which is available at the admin privilege level. This command adds more FlexVol member volumes in the FlexGroup volume and should be used when one of the following occurs:

- The existing member volumes have reached their maximum capacities (100TB or two billion files)
- The physical limits of the node capacities have been reached and more aggregates or nodes are added to the cluster

Note

If you are simply trying to add more capacity or a higher file count to a FlexGroup volume, either [resize the existing volume](#) or increase the [maxfiles](#) value before adding new member volumes via volume expand.

Volume expand considerations

If you are adding new member volumes to a FlexGroup volume, you should take into consideration the following.

- Use `volume expand` only if increasing the existing volume size or file count is not an option or if the FlexGroup is being expanded across new hardware.
- `Volume expand` is currently done via the CLI only.
- If you must add new members, be sure to add them in the same multiples as the existing FlexGroup volume. That is, if the existing FlexGroup volume has 16 member volumes, eight per node, add 16 new members, eight per node, to promote consistent performance.
- If you add new nodes to an existing cluster and add new members to those nodes, try to maintain a consistent number of member volumes per node as in the existing FlexGroup volume.
- Adding new members to a FlexGroup volume changes the ingest heuristics to favor the new, empty member volumes for new data more often and can affect overall system performance for new data ingest while the new members catch up to the existing members.
- Add member volumes in multiples, preferably equal to the working set of member volumes, if possible. For example, if you have eight member volumes, add eight new member volumes when adding members.
- Adding member volumes adds FlexVols to the cluster, which counts against the maximum volume count allowed for the platform.

- When you add new members to a FlexGroup volume, the existing Snapshot copies and SnapMirror relationships are no longer valid for volume-level SnapRestore operations but can be used for client-driven file restore via previous versions and .snapshot directory access. For more information about snapshot restores, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays Data Protection and Backup for ONTAP FlexGroup Volumes](#).
- Existing data in the FlexGroup has no way to rebalance, but that doesn't necessarily indicate a problem. For more information on data imbalances and their impact, refer to ["Data imbalances in FlexGroup volumes"](#).

If you use `volume expand`, be sure to follow the guidance listed in ["Recommendations for adding capacity"](#).

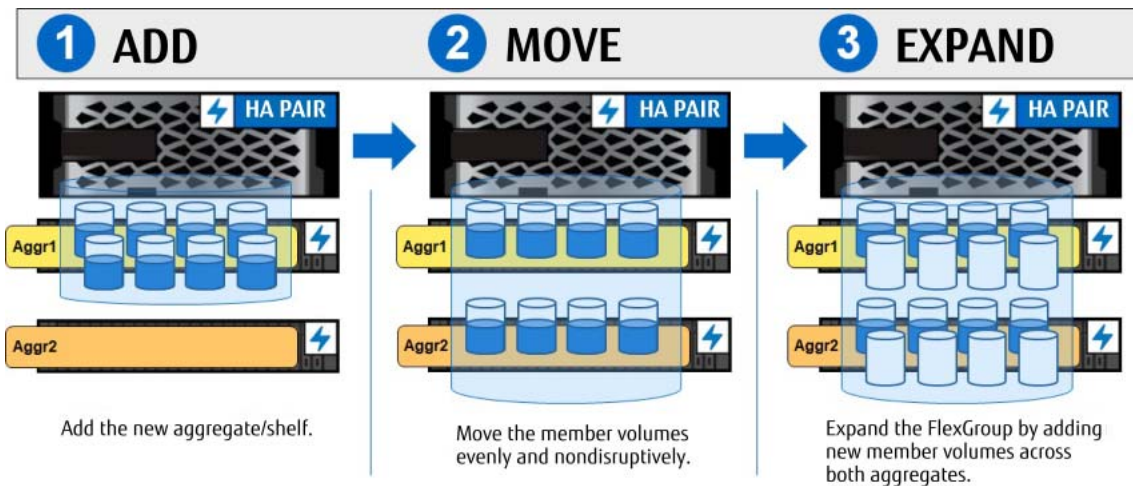
For an example of `volume expand`, refer to ["Command Examples"](#).

Adding disks, aggregates, and nodes

When adding disks to an existing aggregate that contains FlexGroup member volumes, no action is required unless you also want to increase the total volume size.

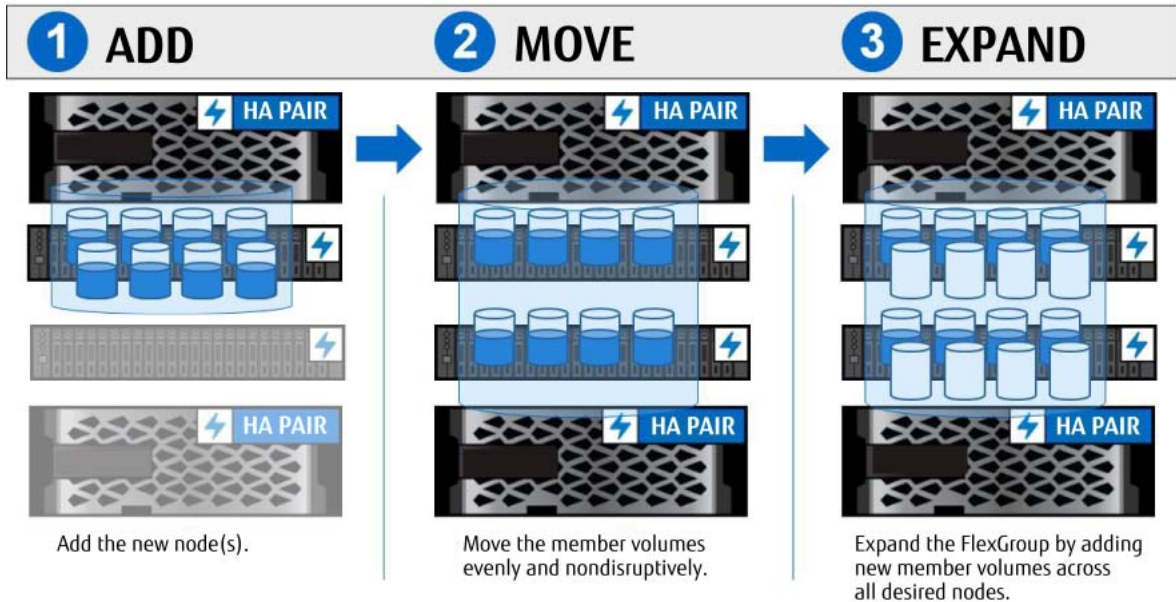
When adding aggregates to nodes, if the FlexGroup volume must span the new aggregates, you can use nondisruptive volume moves to move member volumes to the new aggregates without needing a maintenance window. Then you would create member volumes in the FlexGroup volume spanning new and old aggregates.

Figure 48 Adding aggregates with FlexGroup volumes



When adding new nodes to a cluster, follow the same steps for adding aggregates to a cluster. Use `volume move` and `volume expand` commands to adjust the member volumes.

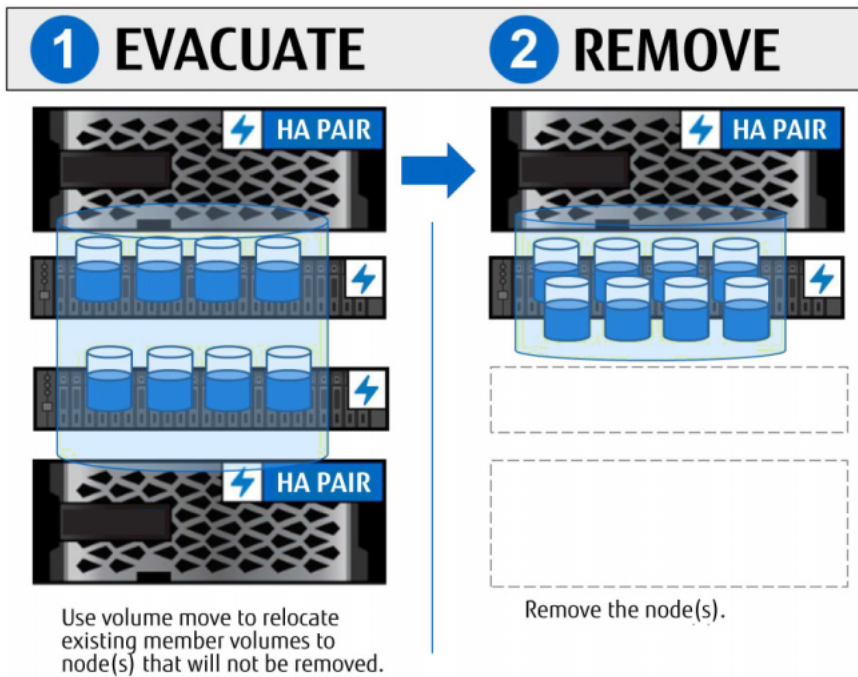
Figure 49 Adding nodes and expanding the FlexGroup volume



Removing or evacuating nodes from a cluster

You can also remove nodes from a cluster that contain FlexGroup member volumes by using nondisruptive volume moves of member volumes. For example, if you want to remove two nodes from an eight-node cluster and each node has 16 member volumes, then you would use `volume move` to distribute 32 member volumes across the remaining six nodes. Because 32 volumes do not evenly divide across six nodes, use the next divisible node count to evenly distribute so that four nodes get eight member volumes per node. If there isn't enough space for four nodes to take on eight member volumes each, place six member volumes on two nodes (12) and five member volumes on four nodes (20).

Figure 50 Removing nodes that contain FlexGroup member volumes



Note

Member volumes cannot be removed from FlexGroup volumes once they are added.

Nondisruptive volume move considerations

ONTAP enables you to perform nondisruptive volume moves between aggregates or nodes in the same cluster. This feature provides flexibility when you are dealing with maintenance windows or attempting to balance performance or capacity allocation in a cluster.

FlexGroup volumes also support this feature, but with even more granularity; you can move each member volume in a FlexGroup volume by using this functionality. Volume move does not move the entire FlexGroup in a single move.

Storage administrators therefore have a way to move workloads around in a cluster if capacity or performance concerns arise. With the [ability of Active IQ to review individual member FlexVol volumes](#), you can quickly identify and resolve issues.

Note

Keep in mind that, although volume moves are nondisruptive, the amount of time that they take depends on the volume size and on the overall load on the node.

When to use nondisruptive volume moves

Nondisruptive volume moves can come in handy in the following scenarios for FlexGroup:

- The member volume is nearing capacity, and no physical storage is available on the current node to grow that volume in place.
- The member volume shares an aggregate with other FlexVol volumes and is being affected by the FlexVol volume's performance or capacity.
- A member volume is overworked in a FlexGroup volume and needs more node resources.
- You want to migrate FlexGroup volumes from spinning disk to SSD for performance or from SSD to spinning disk for archiving.
- New cluster nodes or data aggregates are added.
- You are performing a head swap or other planned maintenance operations (to provide for the least amount of downtime).

Using nondisruptive volume moves

Nondisruptive `volume move` for a FlexGroup member volume is available at the admin privilege level of the command line. Although you can use the ONTAP System Manager GUI to move FlexVol volumes, you currently cannot use it to move FlexGroup member volumes.

To move a FlexGroup member volume, complete the following steps:

Procedure ►►► —————

- 1 Identify the volume that needs to be moved. Use Active IQ Unified Manager or the CLI to determine this information.
- 2 From the command line, run the `volume move start` command. This command can be run at the admin privilege level.

```
cluster::> volume move start -vserver SVM -destination-aggregate aggr1_node2 -volume
flexgroup4TB__000
flexgroup4TB__0001 flexgroup4TB__0002 flexgroup4TB__0003 flexgroup4TB__0004
flexgroup4TB__0005 flexgroup4TB__0006 flexgroup4TB__0007 flexgroup4TB__0008

cluster::> volume move start -vserver SVM -volume flexgroup4TB__0003 -destination-
aggregate aggr1_node2
[Job 2603] Job is queued: Move "flexgroup4TB__0003" in Vserver "SVM" to aggregate
"aggr1_node2". Use the "volume move show -vserver SVM -volume flexgroup4TB__0003"
command to view the status of this operation.

cluster::> volume move show
Vserver   Volume           State      Move Phase Percent-Complete Time-To-Complete
-----
SVM       flexgroup4TB__0003
                healthy replicating
                45%           Tue Dec 06 13:43:01 2016
```

Auto balance aggregate

The Auto Balance Aggregate feature provides ONTAP recommended nondisruptive volume moves when system performance or capacity reaches a point specified by the storage administrator. This feature is not currently supported with FlexGroup volumes.

Considerations when deleting FlexGroup volumes

The volume recovery queue feature helps prevent accidental deletion of volumes by maintaining a recovery queue of deleted volumes for 12 hours. Although the volume is no longer accessible from clients and is hidden from administrator view at admin privilege levels, the space is still allocated, and the remnants of the volume remain in case an emergency recovery is needed. FlexGroup volumes use this recovery queue too, so space is not freed up until the recovery queue expires or is manually purged.

You can see deleted volumes from the command line by specifying `-type DEL` at the `diag` privilege level. Neither volumes nor the recovery queue can be seen from the GUI.

```
cluster::*> volume show -vserver DEMO -type DEL
Vserver  Volume                Aggregate      State      Type      Size Available  Used%
-----  -
DEMO     flexgroup__0001_2321
         aggr1_node1 offline      DEL        5TB       -           -
DEMO     flexgroup__0002_2322
         aggr1_node2 offline      DEL        5TB       -           -
```

Deleted volumes can also be seen with the `volume recovery-queue` command, also with `diag` privileges:

```
cluster::*> volume recovery-queue show
Vserver  Volume                Deletion Request Time      Retention Hours
-----  -
DEMO     flexgroup__0001_2321
         Tue May 01 17:14:14 2018      12
DEMO     flexgroup__0002_2322
         Tue May 01 17:14:13 2018      12
2 entries were displayed.
```

To purge the volumes from the recovery queue manually, run the following commands:

```
cluster::*> volume recovery-queue purge -vserver DEMO -volume flexgroup__0001_2321
Queued private job: 4660

cluster::*> volume recovery-queue purge -vserver DEMO -volume flexgroup__0002_2322
Queued private job: 4661
```

To bypass the recovery queue when deleting a volume, use the `-force true` flag from the CLI with the `volume delete` command at the advanced privilege level. ONTAP System Manager does not support forced deletions or managing the volume recovery queue.

Volume rename considerations

ONTAP supports FlexGroup volume renaming. FlexGroup volume names are meant only for identification of the volumes within the cluster by storage administrators. Client-facing names for volumes are exposed by way of CIFS/SMB shares and volume junction paths (export paths) for NFS, not by how you name a volume in the cluster.

For example, a volume could have an admin name of `vol1` but exported to clients as `/accounting`. Junction paths and SMB shares can be changed at any time in a FlexGroup volume, but doing so causes a disruption for clients, because they must reconnect to the new share or export path through a remount or SMB share reconnection. Volume renames, however, do not cause any client-side disruption.

9. Qtrees

ONTAP introduced support in FlexGroup volumes for logical directories called qtrees. Qtrees allow a storage administrator to create folders from the ONTAP GUI or CLI to provide logical separation of data within a large bucket. Qtrees provide flexibility in data management by enabling unique export policies, unique security styles, and granular statistics.

Qtrees have multiple use cases and are useful for home directory workloads because qtrees can be named to reflect the user names of users accessing data, and dynamic shares can be created to provide access based on a username.

The following bullets give more information regarding qtrees in FlexGroup volumes.

- Qtrees appear as directories to clients.
- Qtrees are able to be created at the volume level; you cannot currently create qtrees below directories to create qtrees that are subdirectories.
- Qtrees are created and managed the same way as a FlexVol qtree is managed.
- Qtrees cannot be replicated using SnapMirror. SnapMirror currently is only performed at the volume level. If you want more granular replication with a FlexGroup, use a combination of FlexGroup volume and [junction paths](#).
- A maximum of 4,995 qtrees is supported per FlexGroup volume. Quota monitoring and enforcement can be applied at the qtree or user level.

Note

ONTAP 9.8 added [qtree QoS](#) support.

Qtrees and file moves

A qtree is considered a unique filesystem in ONTAP. While it looks like a directory from a NAS client perspective, some operations might behave differently than if it were an actual directory. One example of that is moving a file between qtrees in the same volume.

When a file move is done in a volume across directories, the file is simply renamed to a new name and happens within seconds because that is a move inside of the same filesystem.

When a file move occurs between two qtrees, the file is copied to the new location rather than being renamed. This causes the operation to take much longer.

This is a behavior that occurs whether the qtree lives in a FlexVol or a FlexGroup.

Qtree IDs and rename behavior

Once a non-inherited export policy is applied to a qtree, NFS file handles will change slightly when dealing with operations between qtrees. ONTAP will validate qtree IDs in NFS operations, which will impact things like file renames/moves when moving to or from a qtree in the same volume as the source folder or qtree. This is considered a security feature, which helps prevent unwanted access across qtrees, such as in-home directory scenarios. However, simply applying export policy rules and permissions can achieve similar goals.

For example, a move or rename to or from a qtree in the same volume will result in an Access Denied error. The same move or rename to or from a qtree in a different volume results in the file being copied. With larger files, the copy behavior can make it seem like a move operation is taking an unusually long time, where most move operations are near-instantaneous, as they are simple file renames when in the same file system and volume.

This behavior is controlled by the advanced privilege option.

These are the behaviors of different operations.

Assuming that file permissions allow and that client is allowed by export policies to access both source and destination volume/qtrees, these are the current permutations with the 'validate-qtrees-export' flag enabled or disabled:

Enabled:

- Rename in same volume and qtrees: SUCCESS
- Rename in same volume, different qtrees: EACCESS
- Rename between volumes where qtrees IDs differ: EACCESS
- Rename between volumes where qtrees IDs match: XDEV

Disabled:

- Rename in same volume and qtrees: SUCCESS
- Rename in same volume, different qtrees: SUCCESS
- Rename between volumes where qtrees IDs differ: XDEV
- Rename between volumes where qtrees IDs match: XDEV

Note

NFS3ERR_XDEV and NFS3ERR_ACCESS are defined in [RFC-1813](#).

To change the behavior of renames/moves across qtrees, modify `-validate-qtrees-export` to disabled. Refer to [Validating qtrees IDs for qtrees file operations](#) in "FUJITSU Storage ETERNUS AX/HX Series NFS Reference" for more information.

Note

There is no known negative impact to disabling the `-validate-qtrees-export` option, outside of allowing renames across qtrees.

File handle effects for qtrees exports

Normally, the NFS export file handles that are handed out to clients are 32 bits or less in size. However, with qtrees exports, an extra few bits are added to create 40 bit file handles. In most clients, this is not an issue, but older clients (such as [HPUX 10.20, introduced in 1996](#)) might have problems mounting these exports. Be sure to test older client connectivity in a separate test SVM before enabling qtrees exports, because there is currently no way to change file handle behavior after qtrees exports have been enabled.

Managing quotas with FlexGroup

FlexGroup volumes support user/group and tree quotas. The level of support for these can be broken down into the following.

- Support for quota reporting in ONTAP 9.3.
- Support for FPolicy, which can provide quota enforcement from third-party vendors, such as DefendX (formerly NTP) in ONTAP 9.4.
- Enforcement of quotas (that is, setting hard and soft limits for capacity and file count) is supported in ONTAP 9.5 and later.

User and group quota considerations

To implement user or group quotas, the cluster must be able to resolve the specified username or group. This requirement means that the user or group must exist locally on the SVM or within a resolvable name service server, such as Active Directory, LDAP, or NIS. If a user or group cannot be found by the SVM, then the quota rule is not created. If a user or group quota fails to create because of an invalid user, the command line issues this error:

```
Error: command failed: User name user not found. Reason: SecD Error: object not found.
```

ONTAP System Manager delivers a similar message. Use the `event log show` command to investigate the issue further. For more information about configuring name services for identity management in ONTAP, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays How to Configure LDAP in ONTAP Multiprotocol NAS Identity Management](#).

Creating a user or group quota

User and group quotas can be created to report or enforce capacity or file count limits on a per-user basis. These quotas would be used in scenarios where multiple users or groups share the same namespace or qtree. These steps are the same for FlexVol and FlexGroup volumes.

Creating a quota – ONTAP System Manager

To create user or group quota in ONTAP System Manager, navigate the left-hand menu to Storage > Quotas. That takes you to a page with three tabs: Reports, Rules, and Volume Status.

Reports show you the current quota tracking for users, groups, and qtrees.

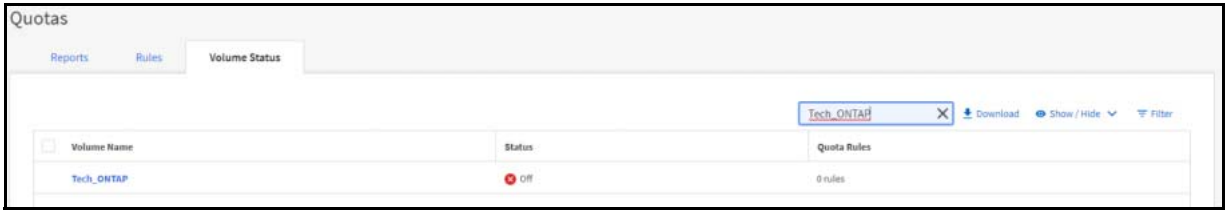
Figure 51 Quota reports – ONTAP System Manager

| Type | Volume | Storage VM | Qtree | Users | Group | % Space Used | % Files Used |
|------|--------|------------|-------|------------------------|-------|------------------------------|-------------------------|
| user | home | DEMO | - | root | - | 4.65 GB used No Hard Limit | 25 used No Hard Limit |
| user | home | DEMO | - | 14 | - | 4 KB used No Hard Limit | 2 used No Hard Limit |
| user | home | DEMO | - | apache | - | 383 MB used No Hard Limit | 2 used No Hard Limit |
| user | home | DEMO | - | Podcast | - | 0 Bytes used No Hard Limit | 2 used No Hard Limit |
| user | home | DEMO | - | admin | - | 4.65 GB used No Hard Limit | 2 used No Hard Limit |
| user | home | DEMO | - | BUILTIN\Administrat... | - | 0 Bytes used No Hard Limit | 15 used No Hard Limit |
| user | home | DEMO | - | squash | - | 0 Bytes used No Hard Limit | 3 used No Hard Limit |
| user | home | DEMO | - | 1003 | - | 12 KB used No Hard Limit | 5 used No Hard Limit |
| user | home | DEMO | - | prof1 | - | 0 Bytes used No Hard Limit | 11 used No Hard Limit |
| user | home | DEMO | - | 1108 | - | 0 Bytes used No Hard Limit | 1 used No Hard Limit |

9. Qtrees
 Managing quotas with FlexGroup

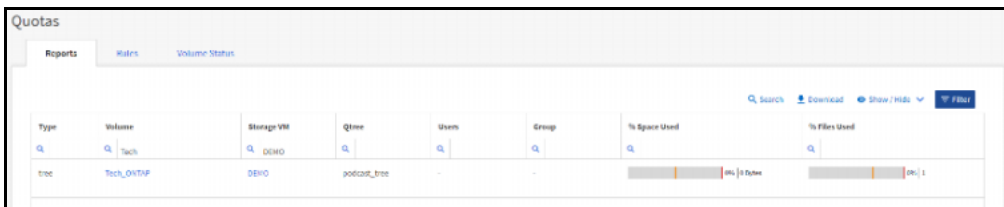
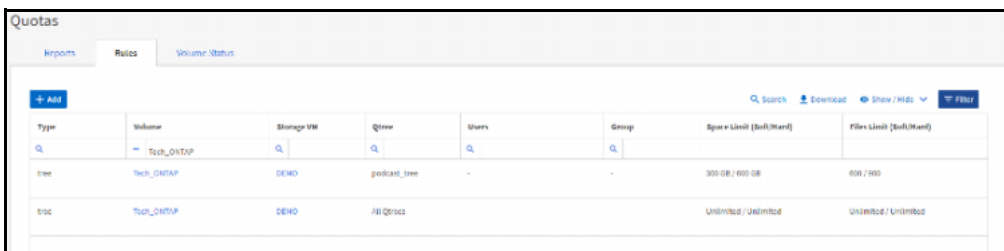
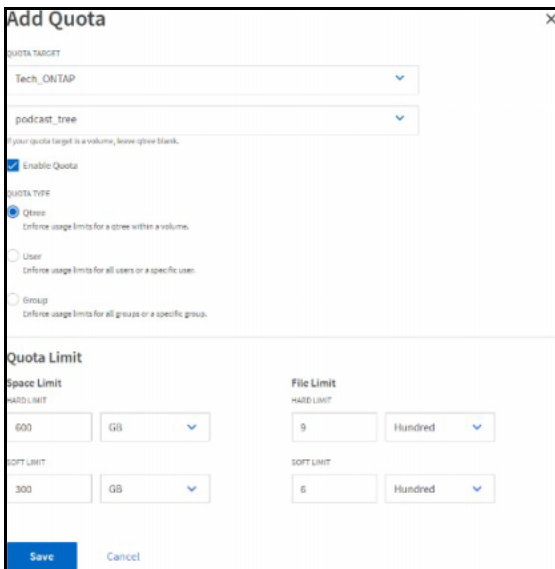
Volume status shows whether quotas are on or off for the volume.

Figure 52 Quota volume status – ONTAP System Manager



Rules is where you would create new quotas for users, groups, or qtrees. Click Add and enter the information for the user, group or qtree quota in the dialog screen. After the rule is created, ONTAP System Manager performs all of the necessary steps to enable and activate the quota.

Figure 53 Quota rules – ONTAP System Manager



Creating a user or group quota – command line

To create a user or group reporting quota with the command line for a specific user or group, use the following command at the admin privilege level:

```
cluster::> quota policy rule create -vserver SVM1 -policy-name default -volume flexgroup -type [user|group] -target [username or groupname] -qtree ""
```

To create a user or group reporting quota with the command line for all users or groups, use the following command at the admin privilege level. The target is provided as an asterisk to indicate all:

```
cluster::> quota policy rule create -vserver SVM1 -policy-name default -volume flexgroup -type [user|group] -target * -qtree ""
```

Creating a tree reporting quota from the command line

To create a tree reporting quota with the command line for a specific user or group, use the following command at the admin privilege level:

```
cluster::> quota policy rule create -vserver DEMO -policy-name tree -volume flexgroup_local -type tree -target qtree
```

To enable quotas, use `quota on` or `quota resize`.

```
cluster::> quota on -vserver DEMO -volume flexgroup_local
[Job 9152] Job is queued: "quota on" performed for quota policy "tree" on volume "flexgroup_local" in Vserver "DEMO".

cluster::> quota resize -vserver DEMO -volume flexgroup_local
[Job 9153] Job is queued: "quota resize" performed for quota policy "tree" on volume "flexgroup_local" in Vserver "DEMO".

cluster::> quota show -vserver DEMO -volume flexgroup_local

      Vserver Name: DEMO
      Volume Name: flexgroup_local
      Quota State: on
      Scan Status: -
      Logging Messages: -
      Logging Interval: -
      Sub Quota Status: none
      Last Quota Error Message: -
      Collection of Quota Errors: -
      User Quota enforced: false
      Group Quota enforced: false
      Tree Quota enforced: true
```

The following example shows a quota report command on a FlexGroup volume with a tree quota specified:

```
cluster::> quota report -vserver DEMO -volume flexgroup_local
Vserver: DEMO

Volume  Tree      Type  ID      ----Disk---  ---Files-----  Quota
-----  -----  -----  -----  -
flexgroup_local
      qtree    tree  1      0B      -      1      -  qtree
```

Files used and disk space used are monitored and increment as new files are created:

```
cluster::> quota report -vserver DEMO -volume flexgroup_local
Vserver: DEMO
Volume Tree      Type  ID      ----Disk---  ---Files-----  Quota
-----  -----  -----  -----  -----  -----  -----
flexgroup_local
      qtree    tree  1
                        13.77MB  -      4      -  qtree
```

Quota enforcement example

When quota enforcement is enabled on a qtree or for a user/group, ONTAP disallows new file creations or writes after a quota is exceeded. This helps storage administrators have greater control over how much data is being written to a volume or qtree.

In addition, when a quota is exceeded, an event management system message is logged at the DEBUG severity level to notify storage administrators of the quota violation. You can configure these messages so that the system forwards them as SNMP traps or as syslog messages.

In this example, a quota has been set with a hard limit of 1GB and 10 files.

```
cluster::*> quota policy rule show -vserver DEMO
Vserver: DEMO          Policy: tree          Volume: flexgroup_local
Type  Target  Qtree  User  Disk  Soft  Soft
-----  -----  -----  -----  -----  -----  -----
tree  qtree   ""     -     1GB  -     10   -     -
```

When a user tries to copy a 1.2GB file to the qtree, ONTAP reports an out of space error.

```
[root@centos7 qtree]# cp /SANscreenServer-x64-7.3.1-444.msi /FGlocal/mtree/
cp: failed to close '/FGlocal/mtree/SANscreenServer-x64-7.3.1-444.msi': No space left
on device
```

The file is partially written, but it is unusable because it is missing data.

```
# ls -alh total 1.1G
drwxr-xr-x 2 root root 4.0K Jul 19 15:44 .
drwxr-xr-x 11 root root 4.0K Jun 28 15:10 ..
-rw-r--r-- 1 root root0 Dec 12 2017 newfile1
-rw-r--r-- 1 root root0 Dec 12 2017 newfile2
-rw-r--r-- 1 root root 1021M Jul 19 2018 SANscreenServer-x64-7.3.1-444.msi
```

ONTAP then reports the quota as exceeded.

```
cluster::*> quota report -vserver DEMO
Vserver: DEMO
Volume Tree      Type  ID      ----Disk---  ---Files-----  Quota
-----  -----  -----  -----  -----  -----  -----
flexgroup_local
      qtree    tree  1
                        1.01GB  1GB    5      10  qtree
```

The same behavior occurs for file count limits. In this example, the file count limit is 10 and the qtree already has five files. An extra five files meet our limit.

```
[root@centos7 /]# su student1
sh-4.2$ cd ~
sh-4.2$ pwd
/home/student1
sh-4.2$ touch file1
sh-4.2$ touch file2
sh-4.2$ touch file3
sh-4.2$ touch file4
sh-4.2$ touch file5
touch: cannot touch 'file5': Disk quota exceeded

cluster::*> quota report -vserver DEMO
Vserver: DEMO
```

| Volume | Tree | Type | ID | ----Disk--- | Used | Limit | ---Files--- | Used | Limit | Quota Specifier |
|-----------------|-------|------|-------------------------|-------------|--------|-------|-------------|------|-------|-----------------|
| flexgroup_local | qtree | tree | 1 | | 1.01GB | 1GB | 5 | 10 | | qtree |
| home | | user | student1, NTAP\student1 | | 4KB | 1GB | 10 | 10 | | student1 |

```
2 entries were displayed.
```

From the event logs, we can see the quota violations.

```
cluster::*> event log show -message-name quota.exceeded
Time                Node                Severity            Event
-----
7/19/2018 16:27:54  node02
                                DEBUG              quota.exceeded: ltype="hard",
volname="home", app="", volident="@vserver:7e3cc08e-d9b3-11e6-85e2-00a0986b1210",
limit_item="file", limit_value="10", user="uid=1301", qtree="treeid=1", vfiler=""
7/19/2018 15:45:02  node01
                                DEBUG              quota.exceeded: ltype="hard",
volname="flexgroup_local", app="", volident="@vserver:7e3cc08e-d9b3-11e6-85e2-
00a0986b1210", limit_item="disk", limit_value="1048576", user="", qtree="treeid=1",
vfiler=""
```

Quotas can be used to control the file counts allowed for FlexGroup volumes if you plan on leaving 64-bit file IDs disabled. For more information, refer to ["Using quota enforcement to limit file count"](#).

Performance effect of using quotas

Performance effects are always a concern when enabling a feature. To alleviate performance concerns when using quotas, we ran a standard NAS benchmark test against FlexGroup volumes in ONTAP 9.5 with and without quotas enabled. We concluded that the performance effect for enabling quotas on a FlexGroup volume is negligible, as shown in [Figure 54](#) and [Figure 55](#).

Figure 54 ONTAP 9.5 performance (operations/sec)—quotas on and off

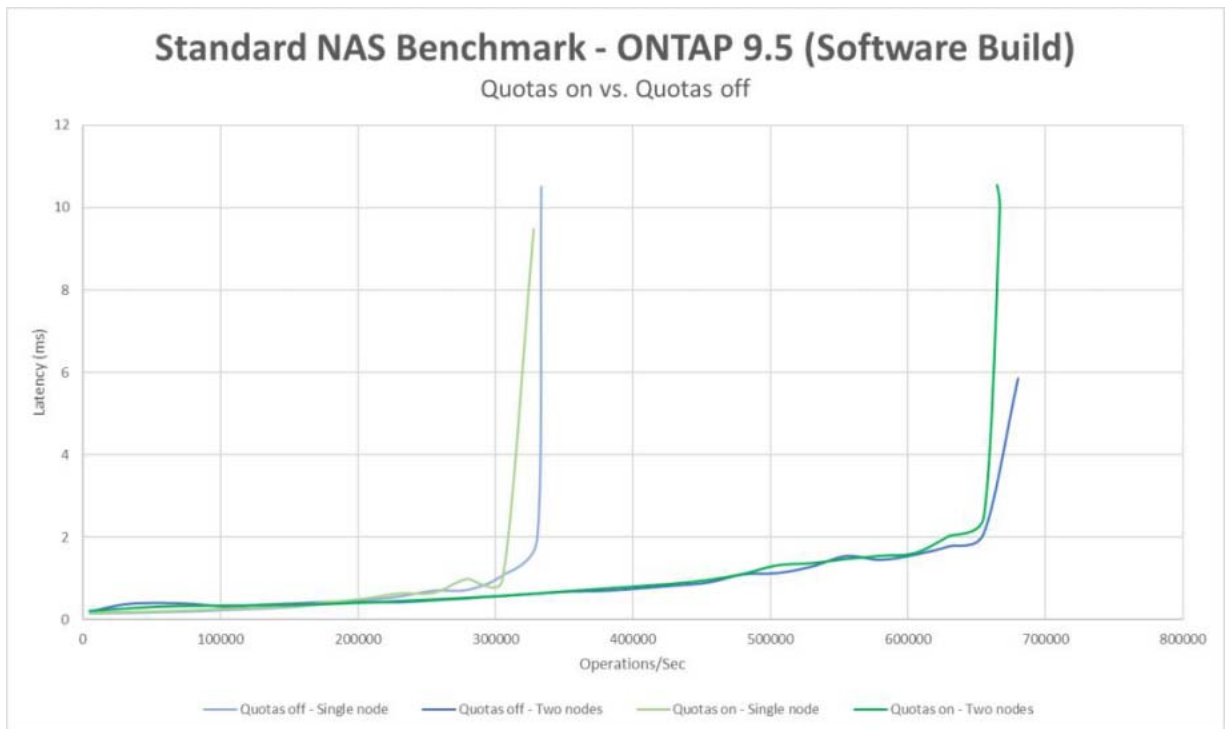
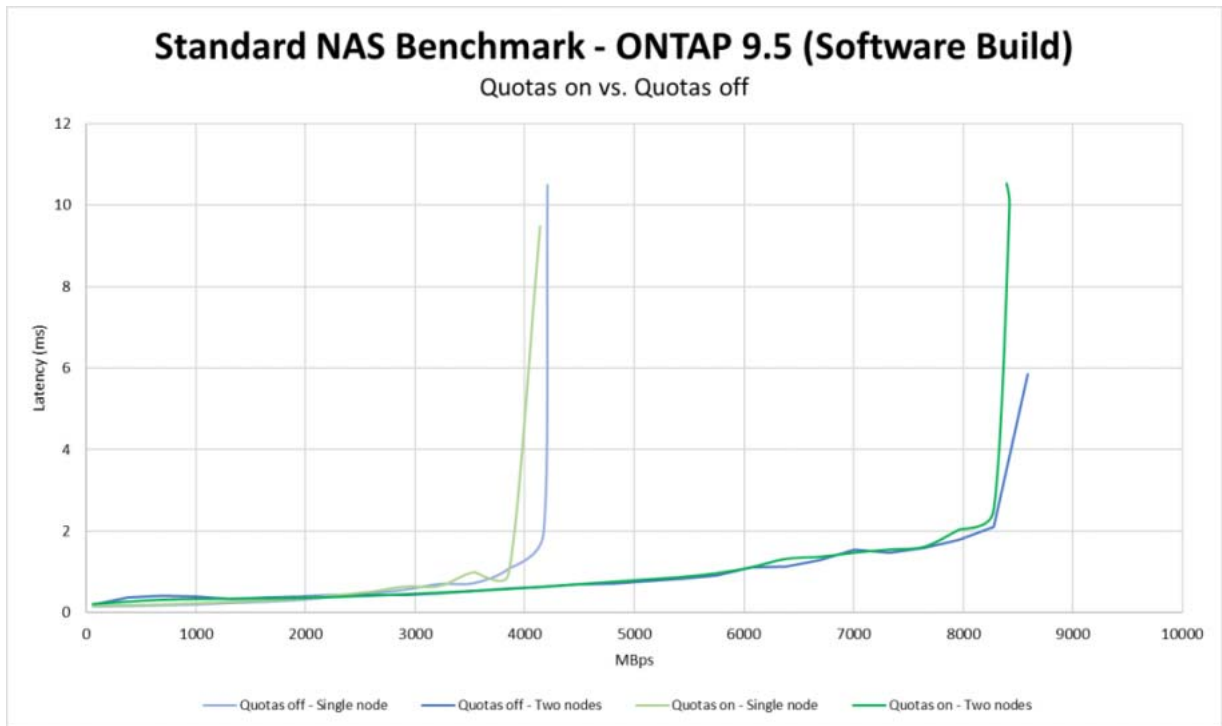


Figure 55 ONTAP 9.5 performance (MBps)–quotas on and off



Quota scan completion times

When a quota initialization or resize takes place, ONTAP must perform some background tasks to complete the necessary work to reflect quota usage accurately. These tasks take time, which depends on a number of factors covered below.

Initialization completion time

The time it takes for quotas to initialize on a volume or qtree depends on the following factors:

- The number of files and folders in a volume. More files mean a longer initialization, while file size do not affect initialization time.
- Type of volume. FlexVol scans can take longer than FlexGroup scans, because FlexGroup quota scans are performed in parallel across the nodes a FlexGroup resides.
- Type of hardware and load on system. Heavily loaded systems with many files can result in scans that take hours.

You can check quota initialization status with the command `quota show -volume volname -instance`.

Quota resize completion time

Quota resize is used when a quota policy is changed. The resize performs a scan with the new limits. This process also has some considerations for time to completion.

- The resize only scans using the newly added rules, so it completes faster than an initialize.
- Resize typically completes in a matter of seconds, since it has to do less than quotas on/off.
- Use resize instead of toggling quotas on/off, because resize completes faster.
- Quota resize can run up to 100 concurrent jobs; after 100 jobs, resize operations must wait in a queue.
- More concurrent scans can impact resize performance and add time to the job completion.

User-mapping considerations with quotas

User mapping in multiprotocol environments (data access from both SMB and NFS) for quotas occurs at the member volume level. Eventually, all member volumes agree on the user mapping. However, sometimes there might be a discrepancy, such as when user mapping fails or times out when doing a name mapping that succeeded on another member. This means that at least one member considers the user to be part of a user-mapped pair, and at least one other member considers it to be a discrete record.

At worst, enforcement of the quota rules can be inconsistent until the issue is resolved. For instance, a user might be able to briefly overrun a quota limit.

An event management system message is sent when user mapping results are coordinated.

```
cluster::*> event route show -message-name fg.quota.usermapping.result -instance

                Message Name: fg.quota.usermapping.result
                Severity: NOTICE
                Corrective Action: (NONE)
                Description: This message occurs when the quota mapper
                decides whether to map the Windows quota record and the UNIX quota record of a user
                into a single multiuser record.
```

Tree quota considerations

SVMs in ONTAP can have a maximum of five quota policies, but only one policy can be active at a time. To see the active policy in an SVM, use the following command:

```
cluster::> vserver show -vserver DEMO -fields quota-policy
vserver quota-policy
-----
DEMO      default
```

Note

Currently, you cannot view this information in ONTAP System Manager.

The default policy is adequate in most cases and does not need to be changed. When `quota on` is issued, the active policy is used—not the policy that was assigned to a volume. Therefore, it's possible to get into a situation where you think you have applied a quota and rules to a volume, but `quota on` fails.

The following example applies a quota policy to a volume:

```
cluster::*> quota policy show -vserver DEMO -policy-name tree

    Vserver: DEMO
    Policy Name: tree
    Last Modified: 10/19/2017 11:25:20
    Policy ID: 42949672962

cluster::*> quota policy rule show -vserver DEMO -policy-name tree -instance

    Vserver: DEMO
    Policy Name: tree
    Volume Name: flexgroup_local
    Type: tree
    Target: tree1
    Qtree Name: ""
    User Mapping: -
    Disk Limit: -
    Files Limit: -
    Threshold for Disk Limit: -
    Soft Disk Limit: -
    Soft Files Limit: -
```

Turning on quotas produces an error because the SVM has default assigned for quotas and does not contain any rules.

```
cluster::*> quota on -vserver DEMO -volume flexgroup_local -foreground true

Error: command failed: No valid quota rules found in quota policy default for volume
flexgroup_local in Vserver DEMO.
```

When you add a rule to default, the quota on command works, but the SVM does not use the new tree policy.

```
cluster::*> quota policy rule create -vserver DEMO -policy-name default -volume
flexgroup_local - type tree -target ""

cluster::*> quota on -vserver DEMO -volume flexgroup_local -foreground true
[Job 8063] Job succeeded: Successful

cluster::*> vserver show -vserver DEMO -fields quota-policy
vserver quota-policy
-----
DEMO      default
```

To use the necessary policy, you must modify the SVM and then turn quotas off and back on.

```
cluster::*> vserver modify -vserver DEMO -quota-policy tree

cluster::*> quota off -vserver DEMO *

cluster::*> quota policy rule delete -vserver DEMO -policy-name default *
1 entry was deleted.

cluster::*> quota on -vserver DEMO -volume flexgroup_local -foreground true
[Job 8084] Job succeeded: Successful
```

This behavior is not unique to FlexGroup volumes; this would happen with FlexVol volumes as well.

How clients see space when quotas are enabled

When quotas are enabled for a qtree in ONTAP, the clients only see the available space as reported by that quota.

For example, this is a quota for qtree1:

```
cluster::*> quota report -vserver DEMO -volume flexgroupDS -tree qtree1
Vserver: DEMO

Volume  Tree      Type  ID      ----Disk---  ---Files-----  Quota
-----  -----  -----  -----  -----  -----  -----
flexgroupDS
      qtree1  tree  1      0B   500GB    1    - qtree1
```

This is how much space that volume actually has:

```
cluster::*> vol show -vserver DEMO -volume flexgroupDS -fields size
vserver volume      size
-----  -----  -----
DEMO    flexgroupDS 10TB
```

This is what the client sees for space for that volume:

```
# df -h /mnt/nas2
Filesystem      Size Used Avail Use% Mounted on
demo:/flexgroupDS 9.5T 4.5G 9.5T 1% /mnt/nas2
```

This is what is reported for that qtree:

```
# df -h /mnt/nas2/qtree1/
Filesystem      Size Used Avail Use% Mounted on
demo:/flexgroupDS 500G 0 500G 0% /mnt/nas2
```

10. General NAS and High-File-Count Considerations

This chapter covers general NAS and high-file-count environment considerations.

High file count considerations

An inode in ONTAP is a pointer to any file or folder within the file system, including Snapshot copies. Each FlexVol volume has a finite number of inodes and has an absolute maximum of 2,040,109,451.

Inodes can be increased after a FlexVol volume has been created and can be decreased only to a number that has not already been allocated.

Default and maximum inode counts

Default and maximum inode counts for volumes (both FlexVol and FlexGroup) are dependent on the total allocated capacity of the volume. For example, a 100GB FlexVol volume would not be able to hold as many inodes as a 8TB FlexVol volume.

[Table 15](#) shows a sample of FlexVol sizes, inode defaults, and maximums.

Table 15 Inode defaults and maximums according to FlexVol size

| FlexVol Size | Default Inode Count | Maximum Inode Count |
|--------------|---------------------|---------------------|
| 20MB* | 566 | 4,855 |
| 1GB* | 31,122 | 249,030 |
| 100GB* | 3,112,959 | 24,903,679 |
| 1TB | 21,251,126 | 255,013,682 |
| 7.8TB | 21,251,126 | 2,040,109,451 |
| 100TB | 21,251,126 | 2,040,109,451 |

*FlexGroup member volumes should not be any smaller than 100GB in size.

Increasing maximum files: considerations

If you would like to avoid monitoring and reacting to out of inode conditions, high file count FlexGroups and FlexVols can be immediately configured with the maximum supported files value, with the following considerations in mind.

The default or maximum number of inodes on a FlexVol volume depends on the volume size and has a ratio of one inode to 4KB of capacity. This means that for every 4KB of allocated space to a volume, you can allocate one inode. Examples of these values are seen in [Table 15](#).

In addition, each inode uses 288 bytes of capacity – this means that having many inodes in a volume can also use up a non-trivial amount of physical space in addition to the capacity of the actual data as well. If a file is less than 64 bytes, it is stored in the inode itself and does not use additional capacity.

This used space counts against the 10% aggregate reserve in ONTAP. Two billion files can use as much as ~585GB of space, and if you have many volumes set to the maximum files limit, then each volume's inode capacity is allocated to that aggregate reserve. This capacity is only used when files are actually allocated to the volume and not by simply setting the maximum files value.

As a result, if you increase the files value to the maximum, you should pay attention to the both the used inodes as well as the used aggregate space. Keeping both values in the 80% range gives the best results for high file count environments.

■ Other considerations:

- FlexGroup volumes are ideally the volume of choice to use for high file count environments due to their ability to nondisruptively scale when a limit has been reached.
- An approximate maximum of one inode per 4KB of allocated size can be configured, so a FlexVol or FlexGroup member volume must be approximately 7.8TB or larger in size in order to configure it with the maximum possible files setting of two billion.
 - In a FlexGroup volume, this means each member volume must be 7.8TB or greater.
- You should still monitor for out of inode conditions in case your environment hits the maximum supported values, and you may need to revisit the files setting any time that you grow or shrink a FlexVol or FlexGroup.
- If you choose to set the maximum files value in your volume, you should also consider setting your monitoring thresholds to 80% of the allocated inodes to give yourself ample time to plan and react before you run out of inodes.
- If the files value is set to the maximum amount on a FlexVol or on the individual member volumes in a FlexGroup and you run out of inodes, you cannot increase them further unless you are using a FlexGroup volume and add new member volumes. Therefore, avoid setting FlexVol volumes to two billion if possible. Rather, use FlexGroup volumes so there is at least the option of adding member volumes in case you hit the two billion maximum.
- Finally, keep in the mind that inode metadata is stored in the underlying aggregate, so aggregate free space should be monitored to ensure that the aggregate does not run out of space.

Default and maximum inode counts – FlexGroup volume considerations

When a default volume inode count reaches 21,251,126, it remains at that default value, regardless of the size of the FlexVol volume. This feature mitigates potential performance issues, but it should be considered when you design a new FlexGroup volume. The FlexGroup volume can handle up to 400 billion files (two billion files x 200 FlexVol member volumes), but the default inode count for 200 FlexVol members in a FlexGroup volume is just 4,250,225,200.

This count is based on the following formula:

```
200 member volumes * 21,251,126 default inodes per member = 4,250,225,200 total default inodes
```

If the FlexGroup volume requires more inodes than what is presented as the default value, the inodes must be increased by using the `volume modify -files` command. As mentioned, this value can be increased to the absolute maximum value allowed if desired, but the guidance in ["Increasing maximum files: considerations"](#) should be followed.

When you use a FlexGroup volume, the total default inode count depends on both the total size of the FlexVol members and the number of FlexVol members in the FlexGroup volume.

[Table 16](#) shows various examples of FlexGroup configurations and the resulting default inode counts.

Table 16 Inode defaults resulting from FlexGroup member sizes and member volume counts

| Member Volume Size | Member Volume Count | Default Inode Count (FlexGroup) |
|--------------------|---------------------|---------------------------------|
| 100GB | 8 | 24,903,672 |
| 100GB | 16 | 49,807,344 |
| 1TB | 8 | 170,009,008 |
| 1TB | 16 | 340,018,016 |

| Member Volume Size | Member Volume Count | Default Inode Count (FlexGroup) |
|--------------------|---------------------|---------------------------------|
| 100TB | 8 | 170,009,008 |
| 100TB | 16 | 340,018,016 |

High file counts, low-capacity needs

As mentioned, ONTAP allocates a default inode and maximum inode count based on volume capacity. In [Table 15](#), member volumes smaller than 7.8TB are not able to achieve the maximum of two billion inodes. To get two billion inodes per member volume, the member volume capacity needs to be 7.8TB or greater. A FlexGroup volume with eight member volumes and space guarantees enabled supports up to 16 billion files, but it also provisions ~62.4TB of reserved storage.

If your dataset consists of very small files, you might never come close to that reserved capacity and would be wasting space that could be used for other workloads. For example, if all files in a workload are 288 bytes each in size, 16 billion files consume only ~4.6TB, which is well below the amount of capacity you'd need to get 16 billion files.

When deploying high file counts that use up little capacity, there are two main options for deploying the FlexGroup volume.

- Deploy the FlexGroup volume with 7.8TB or greater member volumes with thin provisioning**
[Thin provisioning](#) a volume simply means that you are telling ONTAP a volume is a certain size, but that the size is not guaranteed in the file system. This provides flexibility in the file system to limit storage allocation to physical space. However, other volumes in the aggregate can affect the free capacity with their used space and if they have enabled space guarantees, so it's important to monitor available aggregate space when using thin provisioning. Refer to ["Overprovisioning or Thin Provisioning in a FlexGroup Volume"](#) for details.
- Manually create the FlexGroup volume with more member volumes than the default**
 If you want to keep space guarantees for the FlexGroup volume, another option for high-file-count and small capacity environments is to create more member volumes in a FlexGroup volume.

Because inode counts are limited per FlexVol member volume according to capacity, adding more smaller member volumes can provide for higher file counts at the same capacity. The following table shows some possible configurations. For more information about manual creation of FlexGroup volumes, refer to ["When would I need to manually create a FlexGroup volume?"](#).

Table 17 High-file-count/small capacity footprint examples—increasing member volume counts

| Total FlexGroup Size | Member Volume Count (Size) | Maximum Inode Count (Entire Flex-Group) |
|--------------------------------|----------------------------|---|
| 80TB (no space guarantee) | 8 (10TB) | 16,320,875,608 |
| 64TB (space guarantee enabled) | 32 (2TB) | 16,320,875,608 |
| 64TB (space guarantee enabled) | 64 (1TB) | 16,320,875,608 |

Planning for high file counts in ONTAP

With utilities like the ["XCP Migration Tool"](#) (using the scan feature), you can evaluate your file count usage and other file statistics to help you make informed decisions about how to size your inode counts in your new FlexGroup volume. For more information about using XCP to scan files, contact Fujitsu Support.

Viewing used and total inodes

In ONTAP, you can view inode counts per volume by using the following command in advanced privilege:

```
cluster::*> volume show -volume flexgroup -fields files,files-used
vserver volume    files    files-used
-----
SVM      flexgroup 170009008 823
```

You can also use the classic `df -i` command. To show all member volumes, use an asterisk with the volume name in diag privilege:

```
cluster::*> df -i Tech_ONTAP*
Filesystem          iused      ifree      %iused      Mounted on      Vserver
/vol/Tech_ONTAP/    10193      169998815  0%          /techontap      DEMO
/vol/Tech_ONTAP__0001/  923      21250203  0%          /techontap      DEMO
/vol/Tech_ONTAP__0002/  4177      21246949  0%          ---             DEMO
/vol/Tech_ONTAP__0003/  878      21250248  0%          ---             DEMO
/vol/Tech_ONTAP__0004/  848      21250278  0%          ---             DEMO
/vol/Tech_ONTAP__0005/  750      21250376  0%          ---             DEMO
/vol/Tech_ONTAP__0006/  972      21250154  0%          ---             DEMO
/vol/Tech_ONTAP__0007/  879      21250247  0%          ---             DEMO
/vol/Tech_ONTAP__0008/  766      21250360  0%          ---             DEMO
```

What happens when you run out of inodes

When a volume runs out of inodes, no more files can be created in that volume until the number of inodes is increased or existing inodes are freed and the cluster triggers an EMS event (`callhome.no.inodes`). Additionally, an AutoSupport message is triggered. A FlexGroup volume takes per-member inode numbers into account when deciding which member volumes are most optimal for data ingest. For examples, refer to ["Inode-related EMS examples"](#).

EMS messages can be used for monitoring or for triggering scripts that automatically increase inode counts to help avoid space errors before they create production workload problems.

For information on increasing maximum files, refer to ["Increasing maximum files: considerations"](#).

Async delete

ONTAP 9.8 introduces a new feature that allows storage administrators to delete entire directories from the cluster CLI, rather than needing to perform deletions from NAS clients. This provides a way to remove high file count folders much faster than via NAS protocols, as well as removing network and client performance contention. This command works for both FlexVol and FlexGroup volumes.

In testing, `async-delete` performed almost 10 times faster than single threaded `rm` commands and is slightly faster on FlexVol volumes.

Table 18 Async-delete performance

| A300 (24,000 files/folders) | <code>rm -rf *</code> seconds | <code>async-delete</code> seconds | Speed increase |
|-----------------------------|-------------------------------|-----------------------------------|----------------|
| FlexVol | 18.3 | 2 | 9.1x |
| FlexGroup | 32.1 | 3 | 10.7x |

When a directory deletion occurs with `async delete`, a job runs and creates several tasks that run in parallel to delete the directory. By default, the job throttles to 5,000 concurrent tasks, but that amount can be decreased to a minimum of 50 or increased to a maximum of 100,000.

When a delete command is issued, ONTAP scans the specified directory. If subdirectories are found, the contents of those directories are deleted first.

The following caveats apply:

- CLI only
- SVM and volumes must be valid
- Volume must be online and mounted
- Directory path must be valid
- Only one async-delete can be run at a time
- Must be run on a directory; cannot be run on single files

To run a delete job:

```
cluster::*> async-delete start -vserver DEMO -volume FlexGroup1 -path /files
[Job 34214] Job is queued: Asynchronous directory delete job.
```

To check progress:

```
cluster::*> async-delete show -vserver DEMO -instance
```

64-bit file identifiers

By default, NFS in ONTAP uses 32-bit file IDs. File IDs are unique identifiers in the file system that allows ONTAP to keep track of which files are which. 32-bit file IDs are limited to 2,147,483,647 maximum signed integers, which is where the two billion inode limit for FlexVols comes from.

FlexGroup volumes are able to support hundreds of billions of files in a single namespace by linking multiple member volumes together, but to get safely beyond the 32-bit signed integer limit of two billion (and remove the possibility of [file ID collisions](#)), 64-bit file IDs must be enabled.

ONTAP can hand out up to 4,294,967,295 file IDs (the 32-bit unsigned integer) in a FlexGroup volume when 32-bit file IDs are used before file ID collisions are guaranteed to occur. File ID collisions are mathematically impossible when there are 2,147,483,647 files, which is why that is the safest file count to use with 32-bit file IDs. After that value is exceeded, the likelihood of file ID collisions grows the closer the file count gets to the unsigned 32-bit integer value of 4,294,967,295. ONTAP does not prevent you from creating more than two billion files in a FlexGroup volume if you set the maxfiles value to a higher value. To learn more about what happens with file ID collisions, refer to ["Effect of file ID collision"](#).

With 64-bit file IDs, ONTAP can allocate up to 9,223,372,036,854,775,807 unique file IDs to files, although the stated supported limit for maximum files in a FlexGroup volume is 400 billion.

The 64-bit file identifier option is set to off/disabled by default. This was by design to make certain that legacy applications and operating systems that require 32-bit file identifiers were not unexpectedly affected by ONTAP changes before administrators could properly evaluate their environments.

Note

Check with your application or OS vendor for their support for 64-bit file IDs before enabling them, or create a test SVM and enable it to see how applications and clients react with 64-bit file IDs. Most modern applications and operating systems can handle 64-bit file IDs without issue.

This option can be enabled with the following advanced privilege level command and has NFSv3 and NFSv4 options.

```
cluster::*> set advanced
cluster::*> nfs modify -vserver SVM -v3-64bit-identifiers enabled -v4-64bit-identifiers enabled
```

Alternately, you can use [ONTAP System Manager](#) to enable or disable these values.

What happens when I modify this option?

After enabling or disabling this option, you must remount all clients. Otherwise, because the file system IDs change, the clients might receive stale file handle messages when attempting NFS operations on existing mounts. For more information about how enabling or disabling FSID change options can affect SVMs in high-file-count environments, refer to ["How FSIDs operate with SVMs in high-file-count environments"](#).

Do I have to enable 64-bit file IDs?

You might notice that, when you create a new FlexGroup volumes on an SVM that does not have 64-bit file IDs enabled, you get a warning that you should enable the option. However, since enabling the option forces you to remount volumes (and take an outage) and since some applications don't support 64-bit file IDs, you might not want to enable that option.

If your FlexGroup volumes do exceed two billion files, you can leave this value unchanged. However, to prevent any file ID conflicts, the inode maximum on the FlexGroup volume should also be increased to no more than 2,147,483,647.

```
cluster::*> vol show -vserver SVM -volume flexgroup -fields files
```

Note

This option does not affect SMB operations and is unnecessary with volumes that use only SMB.

If your environment has volumes that need 32-bit and other volumes that require more than two billion files, then you can use different SVMs to host those volumes and enable or disable 64-bit file IDs as needed.

Best Practice 15: 64-Bit File Identifiers

Fujitsu strongly recommends enabling the NFS server option `-v3-64bit-identifiers` at the advanced privilege level before you create a FlexGroup volume, especially if your file system exceeds or might exceed the two billion inode threshold.

NFSv3 versus NFSv4.x: File IDs

NFSv3 and NFSv4.x use different file ID semantics. Now that FlexGroup volumes support NFSv4.x, ONTAP 9.7 provides two different options for enabling or disabling 64-bit file IDs.

When you use both NFSv3 and NFSv4.x in an SVM and you want the 64-bit ID option to apply to both protocols, you must set both options.

If only one option is set and volumes are accessed by both protocols, you might see undesired behavior between protocols. For instance, NFSv3 might be able to create and view more than two billion files, whereas NFSv4.x would send an error when a file ID collision occurs.

The options are:

```
-v3-64bit-identifiers [enabled/disabled]  
-v4-64bit-identifiers [enabled/disabled]
```

Note

If you upgrade to ONTAP 9.7 (the first release to support NFSv4.x on FlexGroup volumes), upgrade to 9.7P7 or later.

Using quota enforcement to limit file count

It is possible to set up a quota policy that prevents a Flex Group volume from exceeding two billion files if 32-bit file handles are still being used by way of quota enforcement.

Because quota enforcement policies do not apply to files created below the parent volume (only monitoring/reporting policies), create a qtree inside the FlexGroup volume. Then create a quota tree rule for that qtree with two billion files as the limit to help reduce the risk of users overrunning the 32-bit file ID limitations. Alternately, you can create specific user or group quota rules if you know the user names and group names that will be creating files in the volume.

```
cluster::*> qtree create -vserver DEMO -volume FG4 -qtree twobillionfiles -security-style unix -
oplock-mode enable -unix-permissions 777
cluster::*> quota policy rule create -vserver DEMO -policy-name files -volume FG4 -type tree -
target "" -file-limit 2000000000
cluster::*> quota on -vserver DEMO -volume FG4
[Job 15906] Job is queued: "quota on" performed for quota policy "tree" on volume "FG4" in
Vserver "DEMO".
cluster::*> quota resize -vserver DEMO -volume FG4
[Job 15907] Job is queued: "quota resize" performed for quota policy "tree" on volume "FG4" in
Vserver "DEMO".
cluster::*> quota report -vserver DEMO -volume FG4
Vserver: DEMO
```

| Volume | Tree | Type | ID | ----Disk--- | Used | Limit | ---Files--- | Used | Limit | Quota Specifier |
|--------|-----------------|------|----|-------------|------|-------|-------------|------|-------|-----------------|
| FG4 | twobillionfiles | tree | 1 | 0B | - | 1 | 2000000000 | | | twobillionfiles |
| FG4 | | tree | * | 0B | - | 0 | 2000000000 | | | * |

2 entries were displayed.

After that is done, use file permissions and/or export policy rules to limit access and prevent users from creating files at the volume level. Apply SMB shares to the qtree rather than the volume, and NFS mounts should occur at the qtree level.

Then, as files are created in the qtree, they count against the limit.

```
[root@centos7 home]# cd /FG4/twobillionfiles/
[root@centos7 twobillionfiles]# ls
[root@centos7 twobillionfiles]# touch new1
[root@centos7 twobillionfiles]# touch new2
[root@centos7 twobillionfiles]# touch new3
[root@centos7 twobillionfiles]# ls
new1 new2 new3
cluster::*> quota report -vserver DEMO -volume FG4
Vserver: DEMO
```

| Volume | Tree | Type | ID | ----Disk--- | Used | Limit | ---Files--- | Used | Limit | Quota Specifier |
|--------|-----------------|------|----|-------------|------|-------|-------------|------|-------|-----------------|
| FG4 | twobillionfiles | tree | 1 | 0B | - | 4 | 2000000000 | | | twobillionfiles |
| FG4 | | tree | * | 0B | - | 0 | 2000000000 | | | * |

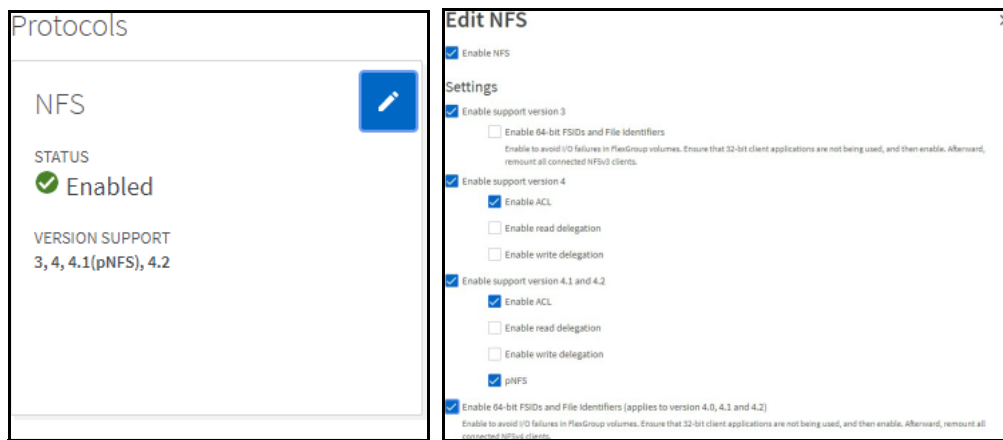
ONTAP System Manager: 9.7

ONTAP 9.7 introduced a new System Manager interface based on REST API capabilities. Because the 64-bit file ID option does not currently exist in the REST API, the only way to modify it in System Manager is to use the CLI.

ONTAP System Manager: 9.8 and later

ONTAP System Manager 9.8 and later includes a GUI method for enabling or disabling the 64-bit file ID value from the Storage > Storage VMs menu option. Click on the desired SVM and select Edit from the NFS protocol menu.

Figure 56 64-bit File IDs in ONTAP System Manager 9.8



Effect of file ID collision

If 64-bit file IDs are not enabled, the risk for file ID collisions increases. When a file ID collision occurs, the effect can range from a stale file handle error on the client, to the failure of directory and file listings, to the entire failure of an application. Usually, it is imperative to enable the 64-bit file ID option when you use FlexGroup volumes.

You can check a file's ID from the client using the `stat` or `ls -i` command. When an inode or file ID collision occurs, it might look like the following. The inode is 3509598283 for both files.

```
# stat libs/
  File: `libs/'
  Size: 12288          Blocks: 24          IO Block: 65536 directory
Device: 4ch/76d Inode: 3509598283 Links: 3
Access: (0755/drwxr-xr-x)  Uid: (60317/ user1)    Gid: (10115/ group1)
Access: 2017-01-06 16:00:28.207087000 -0700
Modify: 2017-01-06 15:46:50.608126000 -0700
Change: 2017-01-06 15:46:50.608126000 -0700

# stat iterable/
  File: `iterable/'
  Size: 4096          Blocks: 8          IO Block: 65536 directory
Device: 4ch/76d Inode: 3509598283 Links: 2
Access: (0755/drwxr-xr-x)  Uid: (60317/ user1)    Gid: (10115/ group1)
Access: 2017-01-06 16:00:44.079145000 -0700
Modify: 2016-05-05 15:12:11.000000000 -0600
Change: 2017-01-06 15:23:58.527329000 -0700

# ls -i libs
3509598283 libs

# ls -i iterable
3509598283 iterable
```

A collision can result in issues such as circular directory structure errors on the Linux client during `find` or `rm` commands and an inability to remove files. In some cases, you may even see stale file handle errors.

```
rm: WARNING: Circular directory structure.
This almost certainly means that you have a corrupted file system.
NOTIFY YOUR SYSTEM MANAGER.
The following directory is part of the cycle:
`/directory/iterable'
rm: cannot remove `/directory': Directory not empty
```

Note

File ID collisions affect NFS only. SMB does not use the same file ID structure.

Effects of file system ID changes in ONTAP

NFS uses a file system ID (FSID) when interacting between client and server. This FSID lets the NFS client know where data lives in the NFS server's file system. Because ONTAP can span multiple file systems across multiple nodes by way of junction paths, this FSID can change depending on where data lives. Some older Linux clients can have problems differentiating these FSID changes, resulting in failures during basic attribute operations, such as `chown` and `chmod`.

If you disable the FSID change option (for NFSv3 or NFSv4), be sure to enable the 64-bit file ID option on the NFS server (refer to "[64-bit file identifiers](#)"), because the total number of file IDs are shared across volumes in the SVM and you run the risk of hitting file ID collisions sooner.

This FSID change option could also affect older legacy applications that require 32-bit file IDs. Perform the appropriate testing with your applications in a separate SVM before toggling FSID change.

How FSIDs operate with SVMs in high-file-count environments

The FSID change option for NFSv3 and NFSv4.x provides FlexVol and FlexGroup volumes with their own unique file systems, which means that the number of files allowed in the SVM is dictated by the number of volumes. However, disabling the FSID change options cause the 32-bit or 64-bit file identifiers to apply to the SVM itself, meaning that the file limits with 32-bit file IDs would apply to all volumes.

For example, if you have 10 billion files in 10 different volumes in your SVM, leaving the FSID change option enabled ensures that each volume can have its own set of unique file IDs. If you disable the FSID change option, then all 10 billion files share the pool of file IDs in the SVM. With 32-bit file IDs, you will likely see file collisions.

Fujitsu recommends leaving the FSID change option enabled with FlexGroup volumes to help prevent file ID collisions.

How FSIDs operate with Snapshot copies

When a Snapshot copy of a volume is created, a copy of a file's inodes is preserved in the file system for access later. The file theoretically exists in two locations.

With NFSv3, even though there are two copies of essentially the same file, the FSIDs of those files are not identical. FSIDs of files are formulated by using a combination of WAFL inode numbers, volume identifiers, and Snapshot IDs. Because every Snapshot copy has a different ID, every Snapshot copy of a file has a different FSID in NFSv3, regardless of the setting of the `-v3-fsid-change` option. The NFS RFC specification does not require FSIDs for a file to be identical across file versions.

Directory size considerations: maxdirsize

In ONTAP, there are limitations to the maximum directory size on disk. This limit is known as `maxdirsize`. The `maxdirsize` value for a volume is capped at 320MB, regardless of platform. This means that the memory allocation for the directory size can reach a maximum of only 320MB before a directory can no longer grow larger. Directory sizes grow when file counts in a single directory increase. Each file entry in a directory counts against the allocated space for the directory. For information on how many files you can have in a single directory, refer to ["Number of files that can fit into a single directory with the default maxdirsize"](#).

Best Practice 16: Recommended ONTAP Version for High-File-Count Environments

For high-file-count environments, use the latest ONTAP release available to gain the benefit of FlexGroup feature enhancements, WAFL enhancements, and performance improvements for high file count workloads.

What directory structures can affect maxdirsize?

The `maxdirsize` value can be a concern when you are using flat directory structures, where a single folder contains millions of files at a single level. Folder structures where files, folders, and subfolders are interspersed have a low impact on `maxdirsize`. There are several directory structure methodologies.

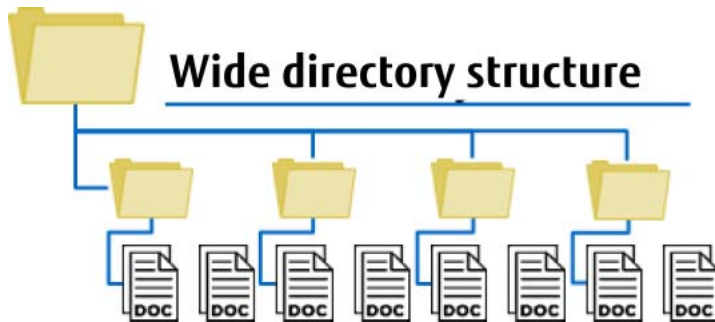
- **Flat directory structure**

A single directory with many files.



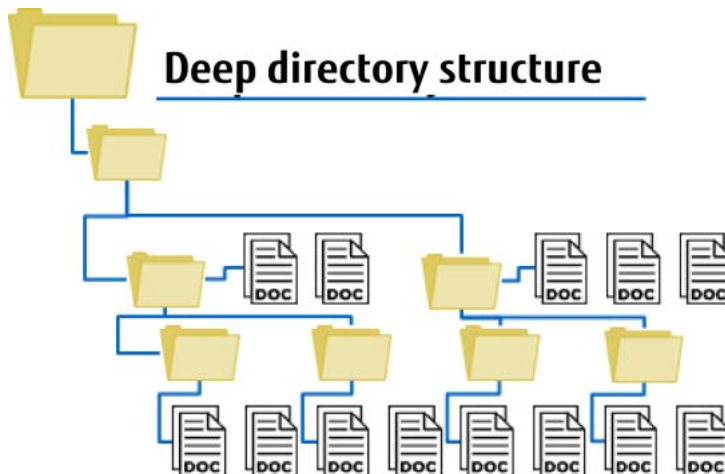
- **Wide directory structure**

Many top-level directories with files spread across directories.



- **Deep directory structures**

Fewer top-level directories, but with many subfolders; files spread across directories.



How flat directory structures can affect FlexGroup volumes

Flat directory structures (many files in a single or few directories) have a negative effect on a wide array of file systems, whether they are Fujitsu systems or not. Potential issues can include, but are not limited to:

- Memory pressure
- CPU utilization
- Network performance/latency (particularly during mass queries of files, `GETATTR` operations, `REaddir` operations, and so on)

FlexGroup volumes can also have an extra effect on `maxdirsize`. Unlike a FlexVol volume, a FlexGroup volume uses remote hard links inside ONTAP to help redirect traffic. These remote hard links are what allow a FlexGroup volume to deliver scale-out performance and capacity in a cluster.

However, in flat directories, a higher ratio of remote hard links to local files is seen. These remote hard links count against the total `maxdirsize` value, so a FlexGroup volume might approach the `maxdirsize` limit faster than a FlexVol will.

For example, if a directory has millions of files in it and generates roughly 85% remote hard links for the file system, you can expect `maxdirsize` to be exhausted at nearly twice the amount as a FlexVol would.

Best Practice 17: Directory Structure Recommendation

- For the best performance, avoid flat directory structures in ONTAP if at all possible. Wide or deep directory structures work best, as long as the path length of the file or folder does not exceed NAS protocol standards.
 - If flat directory structures are unavoidable, pay close attention to the `maxdirsize` values for the volume and increase them as necessary with the guidance of Fujitsu Support.
 - NFS path lengths are defined by the client OS.
 - For information about SMB path lengths, see this [Microsoft Dev Center link](#).
-

Querying for used maxdirsize values

It is important to monitor and evaluate `maxdirsize` allocation in ONTAP. However, there are no commands for this specific to ONTAP.

Instead, `maxdirsize` allocation would need to be queried from the client.

The following command from an NFS client is able to retrieve the directory size information for a folder inside a FlexGroup volume for the 10 largest directories in a given mount point, while omitting Snapshot copies from the search.

```
# find /mountpoint -name .snapshot -prune -o -type d -ls -links 2 -prune | sort -rn -k 7 | head
```

10. General NAS and High-File-Count Considerations

Directory size considerations: maxdirsize

The following example took less than a second on a dataset in folders with millions of files:

```
[root@centos7 ~]# time find /flexgroup/manyfiles/ -name .snapshot -prune -o -type d -ls -links 2
-prune | sort -rn -k 7 | head
787227871 328976 drwxr-xr-x    2 root root    335544320 May 29 21:23
/flexgroup/manyfiles/folder3/topdir_8/subdir_0
384566806 328976 drwxr-xr-x    2 root root    335544320 May 29 13:14
/flexgroup/manyfiles/folder3/topdir_9/subdir_0
3605793347 328976 drwxr-xr-x    2 root root    335544320 May 29 21:23
/flexgroup/manyfiles/folder3/topdir_0/subdir_0
3471151639 328976 drwxr-xr-x    2 root root    335544320 May 29 13:45
/flexgroup/manyfiles/folder3/topdir_4/subdir_0
2532103978 328976 drwxr-xr-x    2 root root    335544320 May 29 14:16
/flexgroup/manyfiles/folder3/topdir_2/subdir_0
2397949155 328976 drwxr-xr-x    2 root root    335544320 May 29 14:15
/flexgroup/manyfiles/folder3/topdir_1/subdir_0
1994984460 328976 drwxr-xr-x    2 root root    335544320 May 29 13:43
/flexgroup/manyfiles/folder3/topdir_6/subdir_0
1860674357 328976 drwxr-xr-x    2 root root    335544320 May 29 13:18
/flexgroup/manyfiles/folder3/topdir_5/subdir_0
1458235096 328976 drwxr-xr-x    2 root root    335544320 May 29 14:25
/flexgroup/manyfiles/folder3/topdir_3/subdir_0
1325327652 328976 drwxr-xr-x    2 root root    335544320 May 29 14:25
/flexgroup/manyfiles/folder3/topdir_7/subdir_0

real    0m0.055s
user    0m0.002s
sys     0m0.035s
```

Using XCP to check maxdirsize

The XCP Migration Tool is mostly considered to be a rapid data mover, but it also derives value in its [robust file scanning capabilities](#). XCP is able to run `find` commands in parallel as well, so the previous examples can be run even faster on the storage system as well as filter results to directories with specified file counts.

The following XCP command example allows you to run `find` only on directories with more than 2,000 entries:

```
# xcp diag find --branch-match True -fmt '{size} {name}'.format(size=x.digest, name=x)
localhost:/usr 2>/dev/null | awk '{if ($1 > 2000) print $1 " " $2}'
```

This XCP command helps you find the directory size values:

```
# xcp -match "type == d" -fmt '{ } {}'.format(used, x) localhost:/usr | awk '{if ($1 > 100000)
print}' | sort -nr
```

When XCP looks for the directory size values, it scans the file system first. Here's an example:

```
[root@XCP flexgroup]# xcp -match "type == d" -fmt '{ } {}'.format(used, x)
10.193.67.219:/flexgroup_16/manyfiles | awk '{if ($1 > 100000) print}' | sort -nr

660,693 scanned, 54 matched, 123 MiB in (24.6 MiB/s), 614 KiB out (122 KiB/s), 5s
1.25M scanned, 58 matched, 234 MiB in (22.1 MiB/s), 1.13 MiB out (109 KiB/s), 10s
...
31.8M scanned, 66 matched, 5.83 GiB in (4.63 MiB/s), 28.8 MiB out (22.8 KiB/s), 7m52s

Filtered: 31816172 did not match
31.8M scanned, 66 matched, 5.83 GiB in (12.6 MiB/s), 28.8 MiB out (62.4 KiB/s), 7m53s.
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_9/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_8/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_7/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_6/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_5/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_4/subdir_0
336871424 10.193.67.219:/flexgroup_16/manyfiles/folder3/topdir_3/subdir_0
```

Number of files that can fit into a single directory with the default maxdirsize

To determine how many files can fit into a single directory with the default `maxdirsize` setting, use this formula:

- Memory in KB \times 53 \times 25%

Since `maxdirsize` is set to 320MB by default on larger systems, the maximum number of files in a single directory is 4,341,760 for SMB and NFS on FlexVol volumes.

FlexGroup volumes use remote hardlinks to redirect I/O to member volumes. These hardlinks count against the total directory size, so the maximum number of files allowed with 320MB `maxdirsize` would depend on the number of hardlinks that were created. The file count per directory might be in the 2 to 2.6 million range for directories in a FlexGroup volume.

Fujitsu strongly recommends that you keep the `maxdirsize` value at the default value.

Event management system messages sent when maxdirsize is exceeded

The following event management system (EMS) messages are triggered when `maxdirsize` is either exceeded or close to being exceeded. Warnings are sent at 90% of the `maxdirsize` value and can be viewed with the event log show command or with the ONTAP System Manager event section. Active IQ Unified Manager can be used to monitor `maxdirsize`, trigger alarms, and send a notification before the 90% threshold (refer to "[Capacity Monitoring and Alerting](#)"). These event management system messages also support SNMP traps.

```
wapl.dir.size.max  
wapl.dir.size.max.warning  
wapl.dir.size.warning
```

Effect of increasing the maxdirsize value

When a single directory contains many files, lookups (such as in a find operation) can consume large amounts of CPU and memory. Directory indexing creates an index file for directory sizes exceeding 2MB to help offset the need to perform so many lookups and avoid cache misses.

Usually, this helps large directory performance. However, for wildcard searches and `readdir` operations, indexing is not of much use. When possible, use the latest version of ONTAP for high file count environments to gain benefits from WAFL improvements.

Best Practice 18: Maxdirsize Maximums

Values for `maxdirsize` are hard coded not to be able to exceed 4GB. To avoid performance issues, Fujitsu recommends setting `maxdirsize` values no higher than 1GB.

Do FlexGroup volumes bypass maxdirsize limitations?

In FlexGroup volumes, each member volume has the same `maxdirsize` setting (which is configured at the FlexGroup level). Even though the files in a directory could potentially span multiple FlexVol member volumes and nodes, the directory itself resides on a single member volume. As a result, the same `maxdirsize` limitations you see in a FlexVol volume still come into play with a FlexGroup. This is because directory size is the key component, not the volume. In a FlexGroup volume, since a directory would reside in a single FlexVol member volume, there is no relief for environments facing `maxdirsize` limitations.

Best Practice 19: Avoiding maxdirsize Issues

Newer platforms offer more memory and CPU capacity, and the ETERNUS AX series systems provide performance benefits for high-file-count environments. However, the best way to reduce the performance effect in directories with large numbers of files is to spread files across more directories in the file system.

Effect of exceeding maxdirsize

When `maxdirsize` is exceeded in ONTAP, an `out of space` error (`ENOSPC`) is issued to the client and an event management system message is triggered. This error can be misleading to storage administrators, because they imply an actual capacity issue when the problem in this case has to do with file count. Always check the ONTAP event log to narrow down problems when clients report seeing capacity issues.

To remediate a directory size issue, a storage administrator must increase the `maxdirsize` setting or move files out of the directory. For more information about remediation, contact Fujitsu Support. For examples of the `maxdirsize` event management system events, refer to ["Example of maxdirsize message"](#).

File system analytics

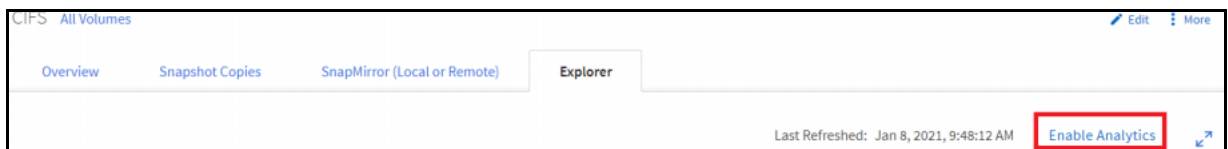
ONTAP 9.8 introduced a new feature that provides a way called File System Analytics for storage administrators to get instant access to file and directory information from ONTAP System Manager.

This initial release of FSA includes information such as:

- File sizes
- Folder sizes
- Atime and mtime histograms
- File and folder listings
- Inactive and active data reporting
- File and directory counts

This information is gathered by ONTAP as the file system is updated after an initial scan is performed and takes minimal system resources to use. File System Analytics are off by default and can be enabled (and disabled) via ONTAP System Manager from the new Explorer tab on the volume page for both FlexVol and FlexGroup volumes, regardless of the NAS protocol in use.

Figure 57 File System Analytics – enable



After analytics are enabled and the initial scan completes (completion time depends on file and folder count), you can browse the entire directory structure by clicking through the directory trees in ONTAP System Manager's Explorer tab.

Figure 58 File System Analytics – directory and file information

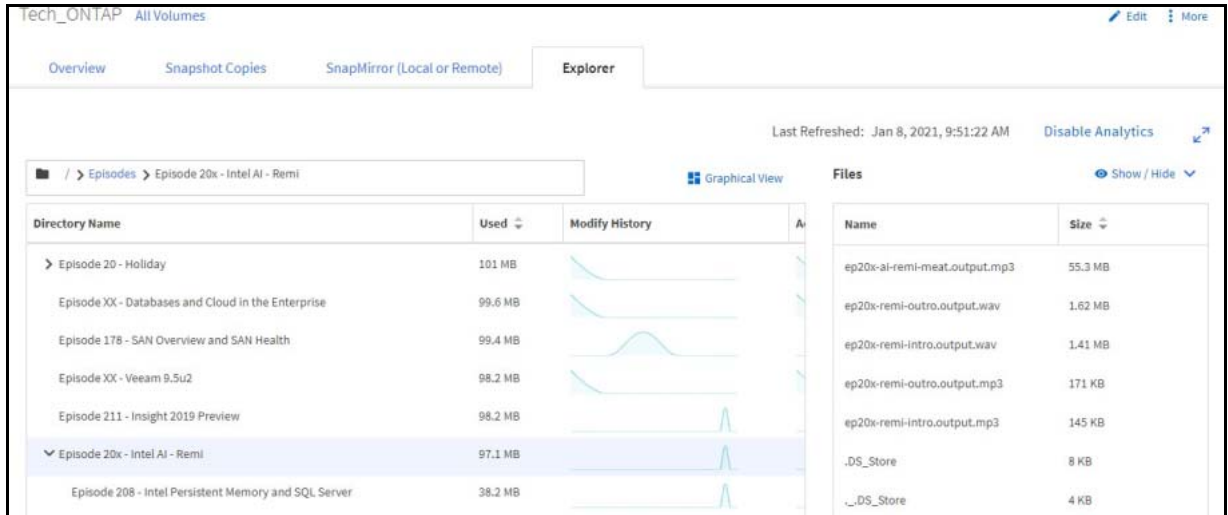
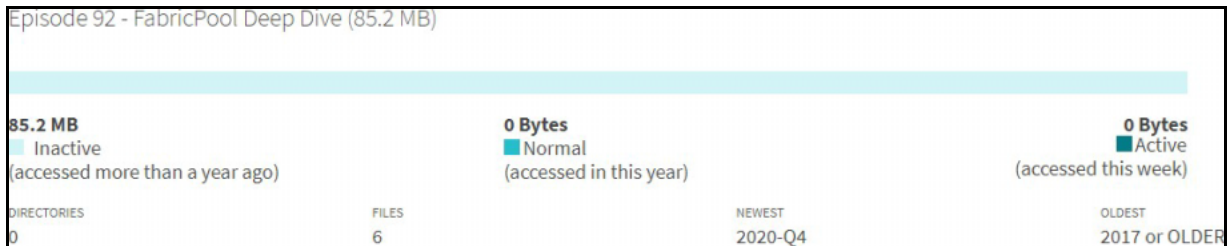


Figure 59 File System Analytics – inactive and active data



When files and folders are created or deleted, File System Analytics updates the tree in seconds with the new information. File System Analytics allows storage administrators to get file and folder information without the need to use off-box utilities or commands such as `du`, `find` and `ls`, which can be time-intensive in high file count environments.

Special character considerations

Most common text characters in Unicode (when they are encoded with UTF-8 format) use encoding that is equal to or smaller than three bytes. This common text includes all modern written languages, such as Chinese, Japanese, and German. However, with the popularity of special characters such as the [emoji](#), some UTF-8-character sizes have grown beyond three bytes. For example, a [trophy symbol](#) is a character that requires four bytes in UTF-8 encoding.

Special characters include, but are not limited to, the following:

- Emojis
- Music symbols
- Mathematical symbols

When a special character is written to a FlexGroup volume, the following behavior occurs:

```
# mkdir /flexgroup4TB/ trophy
mkdir: cannot create directory '/flexgroup4TB/\360\237\217\206': Permission denied
```

In the preceding example, `\360\237\217\206` is hex `0xF0 0x9F 0x8F 0x86` in UTF-8, which is a trophy symbol.

ONTAP software did not natively support UTF-8 sizes that are greater than three bytes in NFS. To handle character sizes that exceed three bytes, ONTAP places the extra bytes into an area in the operating system known as `bagofbits`. These bits are stored until the client requests them. Then the client interprets the character from the raw bits. FlexVol and FlexGroup volumes support `bagofbits`.

Best Practice 20: Special Character Handling in FlexGroup Volumes

For optimal special character handling with FlexGroup volumes, use the `utf8mb4` volume language.

Also, ONTAP has an event management system message for issues with `bagofbits` handling, which includes how to identify the offending file ID.

```
Message Name: wapl.bagofbits.name Severity: ERROR

Corrective Action: Use the "volume file show-inode" command with the file ID and
volume name information to find the file path. Access the parent directory from an
NFSv3 client and rename the entry using Unicode characters.
Description: This message occurs when a read directory request from an NFSv4 client
is made to a Unicode-based directory in which directory entries with no NFS
alternate name contain non-Unicode characters.
```

To test `bagofbits` functionality in FlexGroup, use the following command:

```
# touch "$(echo -e "file\xFC")"
```

Support for `utf8mb4` volume language

As mentioned before, special characters might exceed the supported three bytes UTF-8 encoding that is natively supported. ONTAP then uses the `bagofbits` functionality to allow these characters to work.

This method for storing inode information is not ideal, so, `utf8mb4` volume language is supported. When a volume uses this language, special characters that are four bytes in size are stored properly and not in `bagofbits`.

Volume language is used to convert names sent by NFSv3 clients to Unicode, and to convert on-disk Unicode names to the encoding expected by NFSv3 clients. In legacy situations in which NFS hosts are configured to use non-UTF-8 encodings, you should use the corresponding volume language. Use of UTF-8 has become almost universal these days, so the volume language is likely to be UTF-8.

NFSv4 requires use of UTF-8, so there is no need to use non-UTF-8 encoding for NFSv4 hosts. Similarly, CIFS uses Unicode natively, so it works with any volume language. However, use of `utf8mb4` is recommended because files with Unicode names above the basic plane are not converted properly on non-`utf8mb4` volumes.

Volume language can only be set on a volume at creation by using the `-language` option. You cannot convert a volume's language. To use files with a new volume language, create the volume and migrate the files by using a utility like the ["XCP Migration Tool"](#).

Best Practice 21: UTF-8 or `utf8mb4`?

It is best to use the `utf8mb4` volume language to help prevent issues with filename translation unless clients are unable to support the language.

Managing slow directory listings via NFS in high-file-count environments

Some workflows in high-file-count environments include running `find`, `ls`, or other read metadata-heavy operation on an existing dataset. This type of workload is inefficient and can take a long time to complete. If it is necessary to run these operations, there are a few things you can try to help speed things along.

Generally speaking, the issue with these types of operations are client, protocol, or network related. The storage rarely is the bottleneck for read metadata slowness. ONTAP is able to multithread read metadata operations. With `ls` operations, `getattr` requests are sent one at a time, in serial, which means for millions of `getattr` operations, there might be millions of network requests to the storage. Each network request incurs n milliseconds of latency, which adds up over time.

As such, there are a few ways to speed these up:

- **Send more `getattr` requests at a time**
By itself, `ls` can't send requests in parallel. But with utilities like the XCP Migration Tool, it is possible to send multiple threads across the network to greatly speed up `ls` operations. Using XCP scan can help with speed, depending on what the `ls` output is being used for later. For example, if you need the user permissions/owners of the files, using `ls` by itself might be a better fit. But for sheer listing of file names, XCP scan is preferable.
- **Add more network hardware (for example, 100GB instead of 10GB) to reduce round-trip time (RTT)**
With larger network pipes, more traffic can be pushed over the network, thus reducing load and potentially reducing overall RTT. With millions of operations, even shaving off a millisecond of latency can add up to a large amount of time saved for workloads.
- **Run `ls` without unnecessary options, such as highlighting/colors**
When running `ls`, the default behavior is to add sorting, colors, and highlighting for readability. These add work for the operation, so it might make sense to run `ls` with the `-f` option to avoid those potentially unnecessary features.
- **Cache `getattr` operations on the client more aggressively**
Client-side caching of attributes can help reduce the network traffic for operations, as well as bringing the attributes local to the client for operations. Clients manage NFS caches differently, but in general, avoid setting `noac` on NFS mounts for high-file-count environments. Also, keep `actimeo` to a level no less than 30 seconds.
- **Create FlexCache volumes**
FlexCache volumes are able to create instant caches for read-heavy workloads. Creating FlexCache volumes for workloads that do a lot of read metadata operations, such as `ls`, can have the following benefits:
 - For local clusters, it can help offload the read metadata operations from the origin volume to the cache volumes, and, as a result, frees the origin volume up for regular reads and writes.
 - FlexCache volumes can reside on any node in a cluster, so creating FlexCache volumes makes the use of cluster nodes more efficient by allowing multiple nodes to participate in these operations, in addition to moving the read metadata operations away from the origin node.
 - For remote clusters across a WAN, FlexCache volumes can provide localized NFS caches to help reduce WAN latency, which can greatly improve performance for read-metadata-heavy workloads.

When using FlexCache volumes to help read metadata workloads, be sure to disable `fastreaddir` on the nodes that use FlexCache.

```
cluster::> node run "priv set diag; flexgroup set fastreaddir=false persist
```

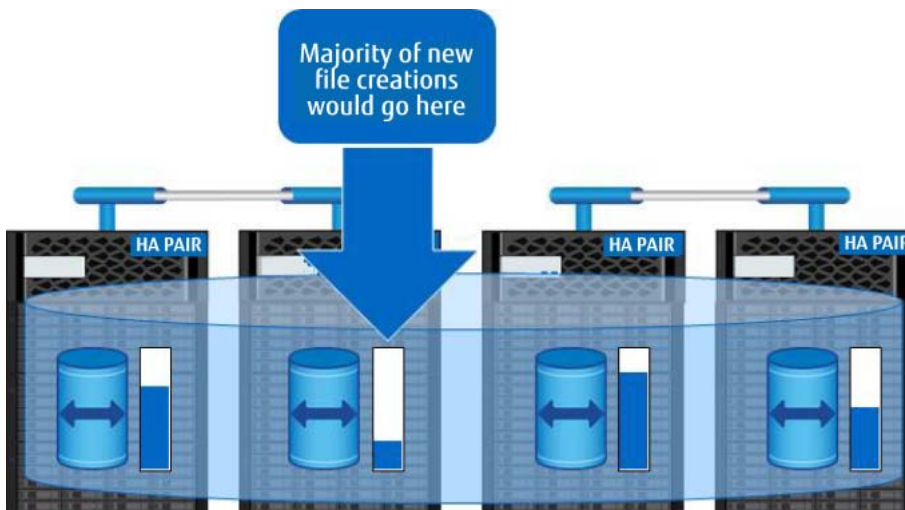
Note

- For this to take effect, a reboot or storage failover is required.
- Starting in ONTAP 9.7, FlexGroup volumes can be origins for FlexCache volumes. For more information about FlexCache volumes, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays FlexCache in ONTAP ONTAP 9.8](#).

File deletions/FlexGroup member volume balancing

A FlexGroup volume spreads data across multiple member volumes relatively evenly on ingest of data. This data layout can help file deletions operate a bit more efficiently on a FlexGroup volume as compared to a FlexVol volume, as the system is able to use more hardware and WAFL affinities to spread out the delete load more efficiently and use less CPU per node for these operations.

Figure 60 Capacity imbalance after deletion of larger files



However, overall performance of file deletions might be slower because of remote access across the FlexGroup volume as compared to FlexVol volumes. In rare cases, the deletion of files (especially sets of large files) can create artificial hot spots in a FlexGroup volume by way of capacity imbalances.

A FlexGroup volume's workload balance can be viewed with the following `diag-privilege-level` command:

```
cluster::*> set diag
cluster::*> node run * flexgroup show [flexgroup name]
```

This displays the following output:

- Member volume dataset ID (DSID)
- Member volume capacities (used and available, in blocks)
- Member volume used %
- Urgency, target, and probability percentages (used in ingest calculations)

For more information, refer to ["flexgroup show"](#).

Rebalancing data within a FlexGroup volume

It is currently not possible to rebalance existing data in a FlexGroup volume to even out capacities, but in most cases, it is not necessary to. ONTAP generally does a good job of balancing the ingest load so that new writes redirect to less full member volumes, and, with ONTAP 9.8's new [proactive resizing](#) feature, ONTAP grows and shrinks member volumes as needed to maintain an even buffer of available free space, so a rebalance is not necessary. A data imbalance does not mean there will be a performance issue unless the data imbalance is also accompanied by very full member volumes. In those cases, performance issues are only seen during new file creation.

In the rare case in which a member volume grows significantly larger than other member volumes, you should analyze the workload to see if anything has changed (for example, the workload went from creating 1MB files to 100GB files). You can use the XCP Migration Tool to scan folders and files to identify file sizes and anomalies. One common scenario that can overallocate a single member volume is if an end user zips up a large amount of data in the FlexGroup. That single zip file might grow to be very large and can fill up a member volume.

For an example of scanning files with XCP, refer to ["Using XCP to scan files before migration"](#).

After the files are identified, either delete them, move them to other volumes, add space to the member volumes, or add new member volumes to help balance the ingest load in a FlexGroup volume. Ideally, upgrade the cluster to ONTAP 9.8 to gain the benefits of proactive resizing, which helps remove the management overhead for member volume capacity.

Why doesn't a FlexGroup volume rebalance existing data?

As a FlexGroup volume ingests data, it has three goals:

- The volume should encourage all its member FlexVol volumes to participate in hosting the workload in parallel. If only a subset of member volumes is active, the FlexGroup volume should distribute more new data toward the underactive members.
- The FlexGroup volume should prevent any member FlexVol volume from running out of free space, unless all other members are also out of free space. When one member has more data than others, the FlexGroup volume should align the underused members by placing new data on them at a higher- than-average rate.
- The FlexGroup volume must minimize the performance losses caused by pursuing the previous two goals. If the FlexGroup volume were to carefully and accurately place each new file where it could be most beneficial, then the previous two goals could be easily achieved. However, the cost of all that careful placement would appear as increased service latency. An ideal FlexGroup volume blends performance with capacity balance, but favors performance.

Some of these goals are in conflict, so ONTAP employs a sophisticated set of algorithms and heuristics to maintain a balance in the FlexGroup volumes. However, in some scenarios, imbalances such as the following might occur:

- Large files or files that grow over time might be present in a FlexVol member volume.
- A workload changes from smaller files to large files (such as a change in how video surveillance cameras record from 4K resolution to 8K resolution).
- Many files might be zipped or tarred into a single file in the same FlexGroup volume as the files themselves.
- A large amount of data might be deleted, and most of that data could be from the same member volume (rare).

In scenarios where FlexGroup member volumes have an imbalance of capacity or files, ONTAP takes extra measures to help the less-allocated member volumes catch up to the filled members. As a result, performance can be affected for new file creations. Existing data should see little to no effect.

■ Performance issues when member volumes reach 80% used capacity

ONTAP 9.8 and proactive resizing further mitigates performance impact when member volumes reach a capacity threshold and is the preferred ONTAP version for FlexGroup volumes.

Listing files when a member volume is out of space

If a FlexGroup member volume runs out of space, the entire FlexGroup volume reports that it is out of space. Even read operations, such as listing the contents of a folder, can fail when a FlexGroup member is out of space.

Although `ls` is a read-only operation, FlexGroup volumes still require a small amount of writable space to allow it to work properly. ONTAP uses that storage to establish metadata caches. For example, suppose the name `foo` points to an inode with X properties, and the name `bar` points to an inode with Y properties. The amount of space used is negligible—a few kilobytes, or maybe a few megabytes on large systems—and this space is used and released repeatedly. Internally, this space is called the RAL reserve.

Under normal circumstances, even if you manage to fill up a member volume, a bit of space is left for the FlexGroup volume to use as it performs read-only operations like `ls`. However, ONTAP prioritizes other operations over the RAL reserve. If a member volume is 100% full, for example, and you create a Snapshot copy and then try to continue using the volume, the WAFL Snapshot reserve is used as you overwrite blocks and therefore consumes more space. ONTAP prioritizes the Snapshot space and takes space from things like the RAL reserve. This scenario rarely occurs, but it explains why an operation like `ls` might fail because of lack of space.

File rename considerations

FlexGroup volumes handle most high-metadata workloads well. However, with workloads that do a large amount of file renames at a time (for example, hundreds of thousands), performance of these operations suffers in comparison to FlexVol volumes. This is because a file rename does not move the file in the file system; instead, it just moves the file name to a new location. In a FlexGroup volume, moving this name would likely take place as a remote operation and create a remote hard link. Subsequent renames would create more remote hard links to the file's location, which would keep adding latency to operations that occur on that file. If an application's workflow is mostly file renames, you should consider using FlexVol volumes instead of FlexGroup volumes. If the desired final landing spot is a FlexGroup volume after the rename occurs, consider moving the files from the FlexVol volume to the FlexGroup volume after the rename process.

Symlink considerations

If your workload contains many symlinks (that is, symlink counts in the millions) in a single FlexGroup volume, attempts to resolve that many symlinks might have a negative effect on performance. The negative effect is caused by creating remote hard links artificially in addition to the remote hard links ONTAP creates.

Best Practice 22: Symlinks in FlexGroup volumes

Try to keep the number of symlinks below a few thousand per FlexGroup if possible.

NFS version considerations

When a client using NFS attempts to mount a volume in ONTAP without specifying the NFS version (for example, `-o nfsvers=3`), a protocol version negotiation between the client and server takes place. The client asks for the highest versions of NFS supported by the server. If the server (in the case of ONTAP, an SVM serving NFS) has NFSv4.x enabled, the client attempts to mount with that version.

In ONTAP 9.7 and later, NFSv4.x is supported. This can create a different set of issues, however. Clients still mount the latest NFS version advertised by the NFS server (in this case, the ONTAP SVM). If NFSv4.x versions are enabled, clients might mount through NFSv4.x when NFSv3 is desired or expected. When NFSv4.x mounts, performance and access permissions behave differently than in NFSv3.

Network connection concurrency: NFSv3

In addition to the preceding considerations, it is worth noting that ONTAP has a limit of 128 concurrent operations per TCP connection for NFSv3 operations. This limit means that for every IP address, the system can handle only up to 128 concurrent operations. Therefore, it's possible that an NFSv3 client would not be able to push the storage system hard enough to reach the full potential of the FlexGroup technology. Clients can be configured to control the number of concurrent operations (by using RPC slot tables) that are sent through NFSv3, which can help avoid hard-to-track performance issues.

Identifying potential issues with RPC slot tables

Many modern NFSv3 clients use dynamic values for RPC slot tables, which means that the client sends as many concurrent operations on a single TCP thread as possible—up to 65,336. However, ONTAP allows only 128 concurrent operations per TCP connection, so if a client sends more than 128, ONTAP enacts a form of flow control on NFSv3 operations to prevent rogue clients from overrunning storage systems by blocking the NFS operation (exec contexts in ONTAP) until resources free up. This flow control may manifest as performance issues that cause extra latency and slower job completion times that might not have a readily apparent reason from the general storage system statistics. These issues can appear to be network related, which can send storage administrators down the wrong troubleshooting path.

To investigate whether RPC slot tables might be involved, use the ONTAP performance counter. You can check whether the number of exec contexts blocked by the connection being overrun is incrementing.

To gather those statistics, run the following command:

```
statistics start -object cid -instance cid
```

Then, review the statistics over a period of time to see if they are incrementing.

```
statistics show -object cid -instance cid -counter execs_blocked_on_cid
```

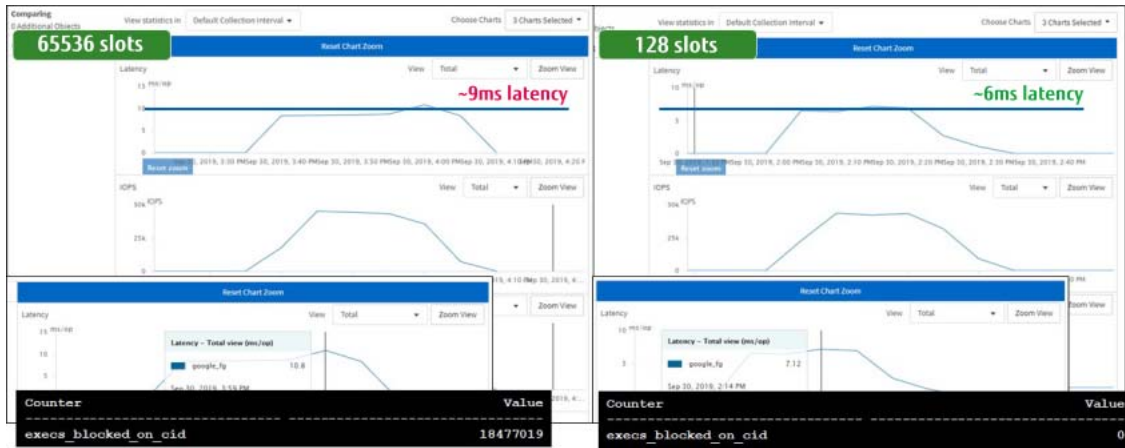
In ONTAP 9.8 and later, a new EMS message (`nblade.execsOverLimit`) has been added to help identify RPC slot table issues. This EMS triggers when the `execs_blocked_on_cid` counters exceed a certain amount over a set period of time. If you see this message in your events, contact Fujitsu Support or look into reducing the number of slot tables used on your NFSv3 clients.

Example of RPC slot table effect on performance

In the following example, a script was run to create 18 million files across 180,000 subdirectories. This load generation was performed from three clients to the same NFS mount. The goal was to generate enough NFS operations with clients that had the default RPC slot table settings to cause ONTAP to enter a flow-control scenario. Then the same scripts were run again on the same clients—but with the RPC slot tables set to 128.

The result was that the default slot tables (65,536) generated 18 million `execs_blocked_on_cid` events and added 3ms of latency to the workload versus the run with the lower RPC slot table setting (128).

Figure 61 Effect of RPC slot tables on NFSv3 performance



Although 3ms might not seem like a lot of latency, it can add up over millions of operations, considerably slowing down job completion.

Resolving issues with RPC slot tables

ONTAP cannot control the number of slot tables a client sends per TCP connection for NFSv3 operations. Therefore, clients must be configured to limit the maximum slot tables sent through NFS to 128. This setting varies depending on the client OS version. Contact the client vendor for more information.

It is possible to get more performance out of a client's NFS connectivity by connecting more mount points to different IP addresses in the cluster on the same client, but that approach can create complexity. For example, rather than mounting a volume at `SVM:/volumename`, multiple mount points on the same client across different folders and IP addresses in the volume could be created.

For example:

```
LIF1:/volumename/folder1
LIF2:/volumename/folder2
LIF3:/volumename/folder3
```

Another possible option is to use the `nconnect` option available for some Linux distributions that can perform multiplexing of NFSv3 over the same TCP connection. This option provides more available concurrent sessions and better overall performance.

Does the RPC slot table limit affect other NAS protocols?

RPC slot table limits only affect NFSv3 traffic.

- SMB clients use different connection methodologies for concurrency, such as SMB multichannel, SMB multiplex, and SMB credits. The SMB connection methodology depends on the client and server configuration and protocol version. For example, SMB 1.0 uses SMB multiplex (mpx), whereas SMB2.x uses SMB credits.
- NFSv4.x clients do not use RPC slot tables—instead, they use state IDs and session tables to control the flow of concurrent traffic from clients.

NFS write appends

ONTAP provides parallel processing of these write appends to improve performance on write appends regardless of the file sizes involved.

Nconnect

Nconnect is a mount option available in some Linux distributions. This option specifies how many TCP connections should be used per mount and offers substantial performance benefits in some workloads per client—generally only when the network threads are the bottleneck in a workload. This also provides benefits to ONTAP by allowing clients to leverage more RPC slot tables per mount. See "[Network connection concurrency: NFSv3](#)" for details on RPC slot tables.

ONTAP 9.8 offers support for the use of nconnect with NFS mounts, provided the NFS client also supports it. If you wish to use nconnect, check to see if your client version provides it and use ONTAP 9.8 or later.

[Table 19](#) shows results from a single Ubuntu client using different nconnect thread values.

Table 19 nconnect performance results

| Nconnect Value | Threads per process | Throughput | Difference |
|----------------|---------------------|------------|------------|
| 1 | 128 | 1.45GB/s | - |
| 2 | 128 | 2.4GB/s | +66% |
| 4 | 128 | 3.9GB/s | +169% |
| 8 | 256 | 4.07GB/s | +181% |

Mapping NFS connected clients to volume names

To check what version of NFS is being mounted from the cluster, use the `nfs connected-clients show` command available in ONTAP 9.7:

```
cluster::> nfs connected-clients show -node * -vserver DEMO

Node: node1
Vserver: DEMO
Data-IP: 10.x.x.x
Client-IP      Volume-Name      Protocol  Idle-Time      Local-Reqs  Remote-Reqs
-----
10.x.x.x      CIFS             nfs4.1    2d 0h 9m 3s   153         0
10.x.x.x      vsroot           nfs4.1    2d 0h 9m 3s   0           72
10.x.x.x      flexgroup_16__0001
                nfs3             0s        0              0           212087
10.x.x.x      flexgroup_16__0002
                nfs3             0s        0              0           192339
10.x.x.x      flexgroup_16__0003
                nfs3             0s        0              0           212491
10.x.x.x      flexgroup_16__0004
                nfs3             0s        0              0           192345
10.x.x.x      flexgroup_16__0005
                nfs3             0s        212289         0
```

To avoid issues with mounting a FlexGroup volume in environments in which NFSv4.x is enabled, either configure clients to use a default mount version of NFSv3 through `fstab` or explicitly specify the NFS version when mounting.

For example:

```
# mount -o nfsvers=3 demo:/flexgroup /flexgroup
# mount | grep flexgroup
demo:/flexgroup on /flexgroup type nfs (rw,nfsvers=3,addr=10.193.67.237)
```

Enabling and using NFSv4.x with FlexGroup volumes

FlexGroup volumes function identically to FlexVol volumes when you configure NFSv4.x in your environment. Rather than focusing on performance, the benefits of using NFSv4.x with workloads include:

- **Security**

NFSv4.x greatly improves security with NFS through integration of ancillary protocols (such as NLM, NSM, mountd, and portmapper) into a single port over 2049. Fewer firewall ports being open helps reduce the threat vectors available.

Additionally, NFSv4.x includes Kerberos encryption (krb5, krb5i, and krb5p) as part of its [RFC requirements](#), meaning that a client/server is not compliant with the RFC unless it includes Kerberos support.

NFSv4.x also provides better masking of UID/GID information by requiring the client and server matching domain IDs in their configurations, which helps make spoofing users harder—particularly when using Kerberos encryption.

Finally, NFSv4.x offers granular ACL support that mimics the functionality of Windows NTFS ACLs. This provides the ability to add more users and groups to an ACL than NFSv3 offered with mode bits, as well as allowing more ACL functionality beyond basic read/write/execute (rwx). [NFSv4.x ACLs can even be applied to datasets that will mount only NFSv3](#), which can offer granular security on files and folders even if NFSv4.x isn't being used.

- **Improved locking**

NFSv3 locking was performed outside the NFS protocol, using ancillary protocols like NSM and NLM. This often resulted in stale locks when clients or servers had outages, which prevented access to files until those stale locks were cleared.

NFSv4.x provides locking enhancements by way of a leasing mechanism that holds a lease for a specified time and keeps that lease if the client/server communication is intact. If there are any issues with that communication (whether network or server outage), the lease expires and releases the lock until it is reestablished.

Additionally, locking in NFSv4.x is integrated within the NFS packets, providing more reliable and efficient locking concepts than NFSv3.

- **Data locality and parallel access**

NFSv4.x offers data locality functionality for scale-out NAS environments, such as NFSv4.x referrals, which can redirect mount requests to volumes in ONTAP according to the location on a node to ensure local access to the mount.

NFSv4.1 also offers parallel NFS support, which establishes a metadata server on mount and then redirects data I/O across the namespace. To do this, it uses a client/server communication that keeps track of data according to node and data LIF location. This concept is similar to that of asymmetric logical unit access (ALUA) for SAN. For more information, refer to "[pNFS with FlexGroup volumes](#)".

NFSv4.x performance enhancements in ONTAP

In general, NFSv4.x is less performant than NFSv3 because NFSv4.x is stateful, so it has more to do for each protocol operation. NFSv4.x overhead comes in the form of locking and leasing, ACLs, compound calls, and communication of state IDs between the client and server, as well as the processing of each packet.

One of the weak points for performance with NFSv4.x includes workloads with high metadata ingest. FlexGroup volumes work best with these types of workloads, so if you're considering NFSv4.x for these workloads, Fujitsu strongly recommends using FlexGroup volumes.

One of the benefits of using NFSv4.x is that it does not use RPC slot tables in its operations, so it is not susceptible to [RPC slot exhaustion](#).

If you are using Kerberos with NFS, there is also a small performance effect to operations for processing overhead of the encrypted packets. The effect varies depending on several factors, including:

- ONTAP version
- Hardware being used
- Network latency, WAN latency, and cloud region
- Performance headroom on the cluster
- Kerberos encryption being used (krb5, krb5i, or krb5p)

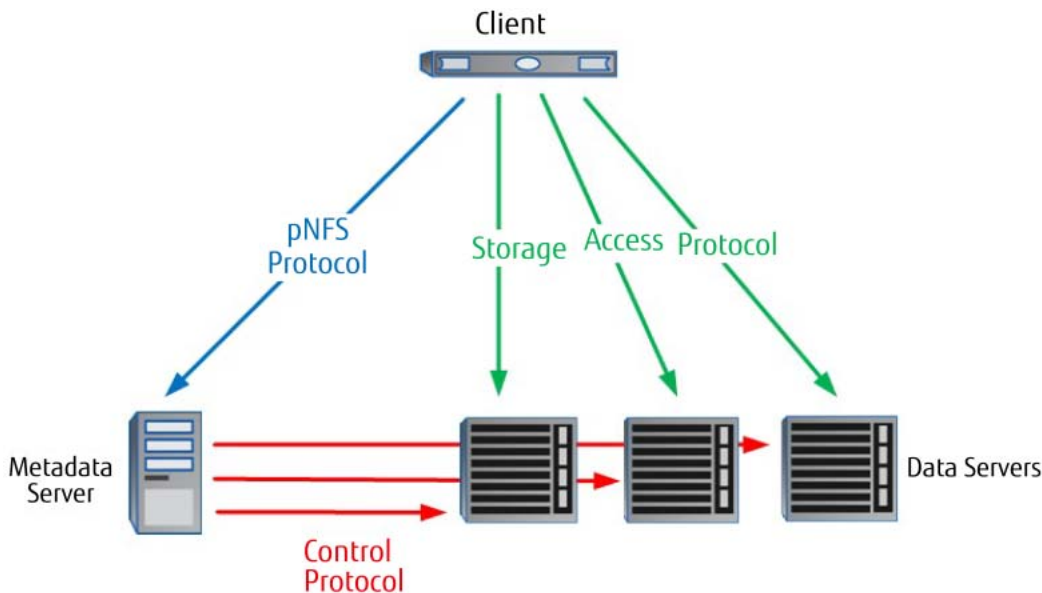
pNFS with FlexGroup volumes

ONTAP offers support for NFSv4.x, which includes NFSv4.1 and its RFC mandatory features. Included in those features is [parallel NFS \(pNFS\)](#), which provides localization of reads and writes across multiple volumes and nodes in a cluster. ONTAP provides the file version of pNFS and does not use the striping or block versions of the feature.

■ How pNFS works in ONTAP

If pNFS has been enabled on the NFS server in an SVM, clients that support pNFS and mount by using NFSv4.1 will first connect to a specific node in the cluster with a single TCP connection that acts as a metadata server. This connection will service pNFS operations, such as client/server communications for data layout, LIF location, and pNFS mappings to help redirect I/O traffic to the local volumes and data LIFs in the cluster. The metadata server also services NFS metadata operations such as `getattr` operations and `setattr` operations.

Figure 62 pNFS diagram



The pNFS architecture includes three main components:

The metadata server that handles all nondata I/O traffic. It is responsible for all metadata operations, such as `GETATTR`, `SETATTR`, `LOOKUP`, `ACCESS`, `REMOVE`, and `RENAME` operations. The metadata server also provides information about the layout of files.

- **Data servers that store file data and respond directly to client read and write requests**
Data servers handle pure `READ` and `WRITE` I/O.
- **One or more clients that are able to access data servers directly**
This access is based on metadata received from the metadata server.

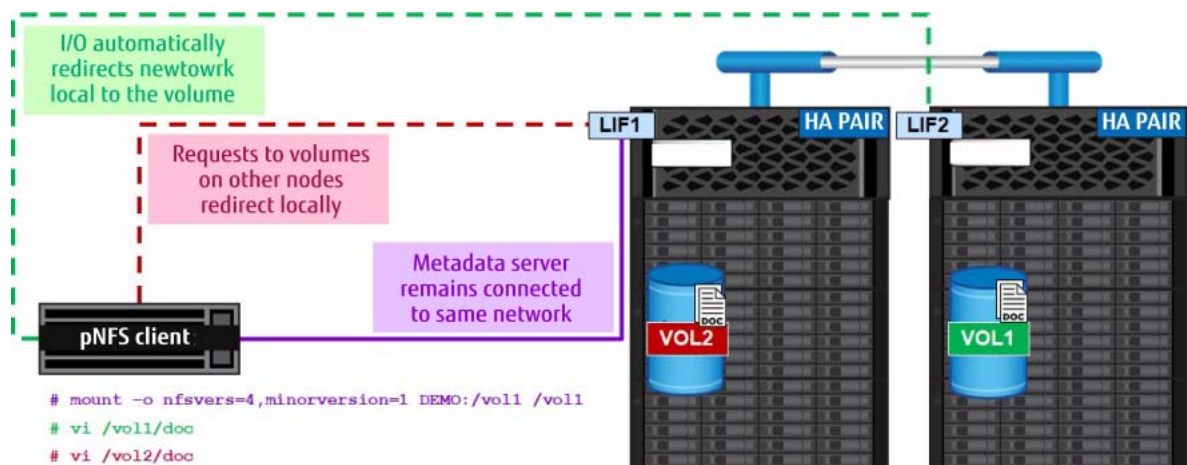
There are three types of protocols used between the clients, metadata server, and data servers:

- **A control protocol used between the metadata server and data servers**
This protocol synchronizes file system data.
- **The pNFS protocol, used between clients and the metadata server**
This is essentially the NFSv4.1 protocol with a few pNFS-specific extensions. It is used to retrieve and manipulate layouts that contain the metadata that describes the location and storage access protocol required to access files stored on numerous data servers.
- **A set of storage access protocols used by clients to access data servers directly**
The pNFS specification currently has three categories of storage protocols: file based, block based, and object based.

When a read or write request is performed by a client over pNFS, the client and server negotiate where to send those requests by using the data layout mappings. For example, if a file lives on volume1 (which lives on node1) in a cluster, but the metadata server is connected to node2, then the data layout mapping informs the client to perform the reads/writes over a network connection local to node1.

If a volume is moved (for example, with a nondisruptive volume move operation), the data layout table is updated and ONTAP redirects local traffic to the volume on the next request. This process is similar to how ALUA works in SAN environments, where a path can switch based on locality of the block device.

Figure 63 pNFS operations diagram



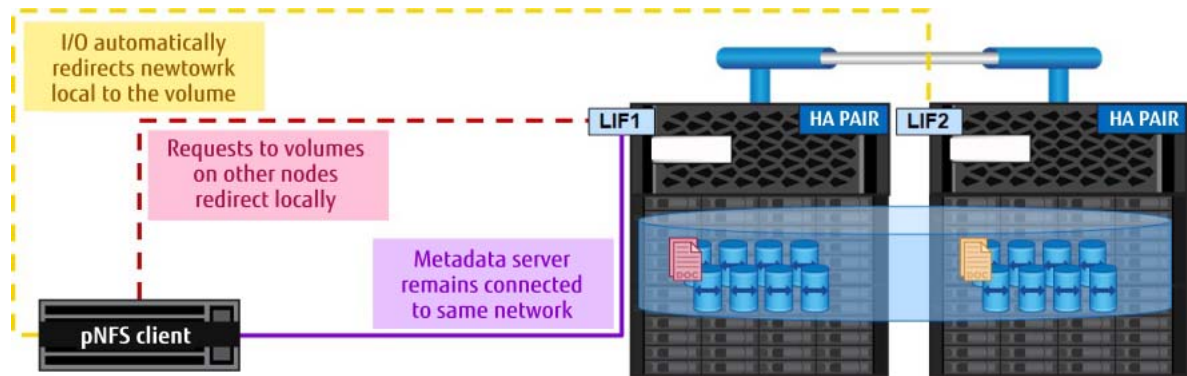
■ How pNFS works with FlexGroup volumes

A FlexGroup volume operates as a single entity, but is constructed of multiple FlexVol member volumes. Each member volume contains unique files that are not striped across volumes. When NFS operations connect to FlexGroup volumes, ONTAP handles the redirection of operations over a cluster network.

With pNFS, these remote operations are reduced, because the data layout mappings track the member volume locations and local network interfaces; they also redirect reads/writes to the local member volume inside a FlexGroup volume, even though the client only sees a single namespace. This approach enables a scale-out NFS solution that is more seamless and easier to manage, and it also reduces cluster network traffic and balances data network traffic more evenly across nodes.

FlexGroup pNFS differs a bit from FlexVol pNFS. Even though FlexGroup load-balances between metadata servers for file opens, pNFS uses a different algorithm. pNFS tries to direct traffic to the node on which the target file is located. If multiple data LIFs per node are given, connections can be made to each of the LIFs, but only one of the LIFs of the set is used to direct traffic to volumes per network interface.

Figure 64 pNFS operations diagram—FlexGroup volumes



■ pNFS best practices

pNFS best practices in ONTAP don't differ much from normal NAS best practices, but here are a few to keep in mind. In general:

- Use the latest supported client OS version.
- Use the latest supported ONTAP patch release.
- Create a data LIF per node, per SVM to ensure data locality for all nodes.
- Avoid using LIF migration on the metadata server data LIF, because NFSv4.1 is a stateful protocol and LIF migrations can cause brief outages as the NFS states are reestablished.
- In environments with multiple NFSv4.1 clients mounting, balance the metadata server connections across multiple nodes to avoid piling up metadata operations on a single node or network interface.
- If possible, avoid using multiple data LIFs on the same node in an SVM.
- In general, avoid mounting NFSv3 and NFSv4.x on the same datasets. If you can't avoid this, check with the application vendor to ensure that locking can be managed properly.
- If you're using NFS referrals with pNFS, keep in mind that referrals establish a local metadata server, but data I/O still redirect. With FlexGroup volumes, the member volumes might live on multiple nodes, so NFS referrals aren't of much use. Instead, use DNS load balancing to spread out connections.

NFSv4.x general considerations

When considering NFSv4.x for your SVM, be sure to factor in performance, client and application support, name services infrastructure, and locking mechanisms before deploying. Also consider whether applications can use both NFSv3 and NFSv4.x on the same datasets. For instance, [VMware recommends against service datastores over both protocol versions](#).

If possible, set up a separate SVM to conduct functionality and performance testing before deploying in production.

NFSv4.x configuration generally requires the following to work properly:

- NFS clients that support NFSv4.x.
- NFS mounts that specify NFSv4.x.
- NFS server configuration (NFSv4.x and desired features enabled—such as referrals, pNFS, ACL support, NFSv4 ID domain configured to be identical on client and NFS server).
- Matching user names and groups on client and server (case sensitive; for example, user1@domain.com should exist on both server and client; USER1 and user1 are not considered matches).
- Optional: Name services for UNIX identities, such as NIS or LDAP, can greatly simplify NFSv4.x implementation and functionality.

NAS metadata effect in a FlexGroup volume

The overhead for metadata operations affects how a workload performs, which can be anywhere from a 10% to 30% performance hit for remote operations. Most of the metadata effect is related to write metadata. Most read metadata has little to no effect.

- `getattr`, `access`, `statfs`, `lock`, `unlock`. Little to no FlexGroup overhead.
- `nfs create`, `unlink`, `lookup`. Little to no FlexGroup overhead under heavy load.
- `nfs mkdir`, `rmdir`, `lookup dir`. 50% to 100% remote access, so high overhead.
- CIFS `open/close`. High overhead.

CIFS/SMB considerations

FlexGroup volumes support both NFS and SMB workloads. ONTAP SMB servers offer some features that can help improve the overall performance experience for SMB workloads on both FlexGroup and FlexVol volumes. The following section covers some of those features, as well as some caveats that apply to FlexGroup volumes or high-file-count environments.

SMB version considerations

FlexGroup volumes support SMB 2.x and SMB 3.x versions only. SMB 1 versions are not able to access CIFS/SMB shares pointing to FlexGroup volumes. As SMB 1 is deprecated by Microsoft, there are no future plans to add SMB 1 support to FlexGroup volumes. For full SMB support information, refer to ["5. FlexGroup Feature Support and Maximums"](#).

Before you migrate a CIFS/SMB workload to a FlexGroup, you should verify that no SMB 1 clients are connected to the existing workloads. In ONTAP, you can do that with the following command:

```
cluster::> cifs session show -protocol-version SMB1
```

If SMB 1 access is attempted to a FlexGroup volume, the `Nblade.flexgroupStatefulProtocolAccess` EMS event is logged.

Use of change notifications with SMB

[SMB change notifications](#) are how SMB clients are informed of a file's existence in a SMB share without needing to close a session or refresh a window (such as pressing the F5 key). SMB clients are in constant communication with the SMB server during SMB sessions, and the SMB server sends periodic updates to the client regarding any file changes in the share. This feature is most useful for applications that must write files and then be able to immediately read the files in SMB shares. Change notifications are controlled through the `changenotify` share property. ONTAP automatically sets this share property on new SMB shares, even if change notifications are not needed.

Best Practice 23: SMB Change Notification Recommendation

If you require the use of SMB change notifications, use ONTAP 9.7 or later.

Large MTU

Large MTU allows SMB's maximum transmission unit (MTU) to be increased from 64KB to 1MB, significantly improving the speed and efficiency of large file transfers by reducing the number of packets that need to be processed. You can enable large MTU with the advanced privilege command `cifs options modify -is-large-mtu-enabled true`.

When this is enabled on the CIFS/SMB server in ONTAP, if the client and SMB protocol version support it (SMB 2.1 and later), then the negotiation for MTU size happens automatically.

You can check to see if your ONTAP SVM is using large MTU with the following command:

```
cluster::> cifs session show -is-large-mtu-enabled true
```

Note

Large MTU refers to large read/writes allowed by SMB 2.1 and later servers. It does not refer to MTU sizes by the network layer.

SMB multichannel

ONTAP supports SMB multichannel, which is an SMB 3.0 protocol feature that enables an SMB 3.x client to establish a pool of connections over a single network interface card (NIC) or multiple NICs and use them to send requests for a single SMB session. This is similar to the [nconnect](#) functionality for NFS.

By doing this, single-client performance can be drastically improved over clients that don't make use of this functionality.

SMB multichannel can be enabled with the following advanced privilege command and takes effect on new SMB sessions:

```
cluster::*> cifs options modify -is-multichannel-enabled true
```

You can check to see if your ONTAP SVM's CIFS/SMB sessions are using SMB multichannel with the following command:

```
cluster::> cifs session show -connection-count >1
```

On a Windows client, you can see if multichannel is in use with the [Get-SmbMultichannelConnection](#) PowerShell cmdlet.

Continuously available shares (CA shares)

CA shares provide a way for SMB connections to survive storage failovers without disruption by using SMB 3.x functionality such as scale-out, persistent handles, witness, and transparent failover. CA shares are officially supported only for SQL and Hyper-V workloads.

CA shares are set at the CIFS/SMB share level with the following command:

```
cluster::*> cifs share properties add -share-name SQL -share-properties continuously-available
```

FlexGroup volumes support CA shares and are only officially qualified for Hyper-V and SQL workloads. However, there are caveats to that support.

- SQL Server workloads with only a few large database files might not be a good fit for FlexGroup volumes. However, SQL Server workloads that have many files (logs or databases) are an appropriate use case for FlexGroup volumes and CA shares. Also refer to ["Databases on FlexGroup volumes"](#).

- Hyper-V workloads are listed as officially supported for CA shares, but, as of ONTAP 9.8, only VMware virtualization workloads are officially supported with FlexGroup volumes. Hyper-V workloads can be used on FlexGroup volumes with CA shares, but there has not been the same testing and qualification done as with VMware workloads.
- Virtual hard disk workloads (such as FSLogix VHDx profiles) can be used on FlexGroup volumes and work with CA shares but have not been officially tested or qualified. In some cases, CA shares aren't necessary to host these workloads, so testing should be performed before deploying in production.

In general, CA shares should not be used with metadata-intensive SMB workloads (such as home directories), as this can cause performance issues. If you are using CA shares, other share properties such as homedirectory, branchcache, access-based enumeration, and attribute caching should not be set.

You can see which CIFS/SMB sessions are using CA shares with the following command:

```
cluster::*> cifs session show -continuously-available Yes|Partial
```

Other considerations

There are a few other potential issues you might see in certain scenarios while performing specific tasks like renaming using SMB 8.3 short names or using CIFS symlinks. The following is a list of these issues. The links show which ONTAP release these issues are resolved. The general recommendation is to run the latest patch release of ONTAP for best results.

Virtualization workload considerations

ONTAP 9.8 is the first release that offers official support for VMware virtualization workloads. That means you can provision a VMware NFS datastore using FlexGroup volumes to scale across multiple nodes in a cluster and provide more than 100TB for virtual machines.

Scalable VMware datastores using FlexGroup volumes offer some advantages over FlexVol volumes.

- Up to 20PB and 400 billion files in a single NFS datastore (VMware limits may reduce that amount)
- Rapid VM cloning using ONTAP sis clone and template caches

However, VMware datastores on FlexGroup volumes do not currently support VMware vVols, nor the VMFS file system (FlexGroups are NAS only).

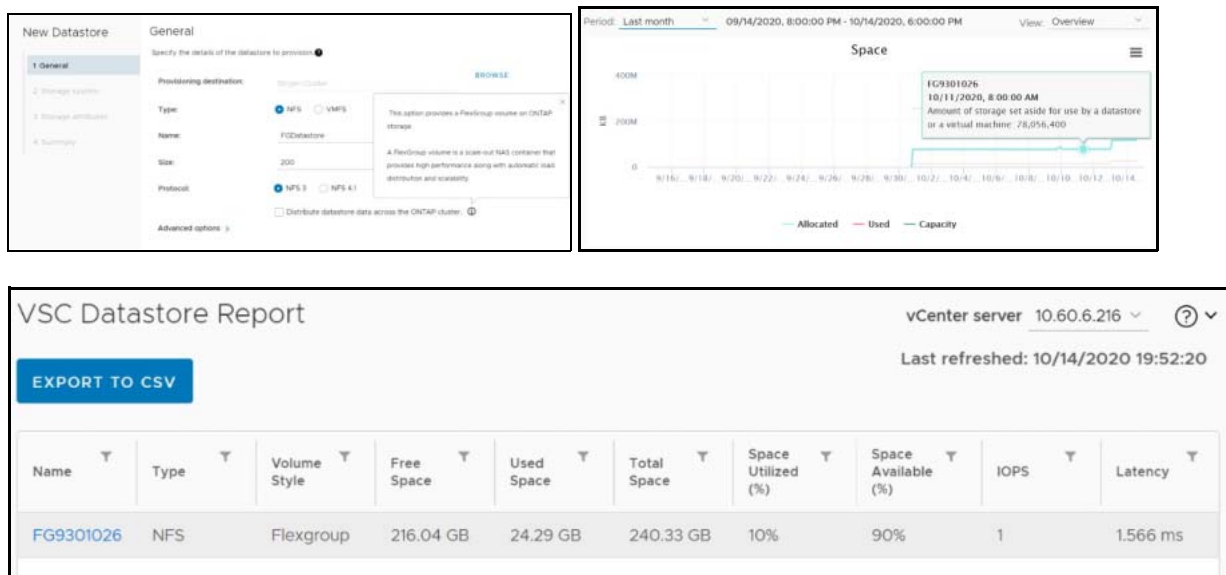
ONTAP tools for VMware vSphere support (formerly Virtual Storage Console)

The new release of ONTAP tools for VMware vSphere provide ways to provision and manage datastores as well. Some of that functionality includes:

- Datastore provisioning as a FlexGroup or FlexVol
- QoS policy management down to the VM level
- Performance metrics at the VM level
- SnapCenter for vSphere support

The following images highlight some of the functionality available.

Figure 65 ONTAP tools for VMware vSphere Support – FlexGroup Datasets



Copy offload

Although ONTAP 9.8 supports an optimized copy offload (VAAI) for faster cloning—even faster than for FlexVol datasets—there are some limitations to keep in mind.

- Copy-offload operations take a variable amount of time (proportional to file size).
- There is a limit of 50 parallel on-demand jobs per node. If jobs exceed that limit, they are placed into a queue.
- There is additional space usage overhead because the template file has to be copied in each of the members. As a result, offload operations in FlexGroup volumes that don't have enough free space may fail.
- Snapshot creation is disallowed until the copy finishes.
- Replication is disallowed until the data pull scan finishes.

Considerations

Although ONTAP has tested and qualified virtualization workloads and ONTAP 9.8 offers features such as [proactive resizing](#) to better accommodate these types of datasets, there are a few considerations you should keep in mind.

- FlexGroup datasets can be provisioned using ONTAP tools for VMware vSphere (preferred) or manually created using the CLI/ONTAP System Manager and then mounted using VMware vSphere.
- Virtual machines and Snapshot copies start out as small files and grow over time. ONTAP resizes individual member volumes as they reach capacity thresholds to automatically maintain an even balance of available free space. This results in some member volumes being larger than others and is normal.
- Qualification testing was done up to 1500 VMs in a FlexGroup dataset. This is not a hard limit, but going beyond the tested limit might create unpredictable results.
- When sizing a FlexGroup dataset, keep in mind that the FlexGroup consists of multiple smaller FlexVol volumes that create a larger namespace. As such, size the dataset to be at least 8x the size of your largest virtual machine. For example, if you have a 6TB VM in your environment, size the FlexGroup dataset no smaller than 48TB.

- FlexGroup volumes support VAAI starting in ONTAP 9.7, which is used to offload copy operations from vSphere to storage. Note that copy offload is not always faster than host copy, and vSphere only offloads operations on cold VMs for NFS storage.
- With virtualization workloads, VMDK files in the same FlexGroup datastore could live in multiple FlexVol member volumes across the cluster. As a result, use SnapCenter for vSphere to coordinate snapshots and replication.
- It is possible to use FlexGroup volumes with virtualization backup products such as Veeam or Rubrik. Check with these vendors for their level of support and interaction with SnapDiff 2.0 or later.
- FlexGroup volumes have only been tested and qualified for VMware datastores. Use of Hyper-V, Citrix Xen, RedHat KVM, and so on has not been tested or qualified and is not officially supported.
- Placing virtual hard disks (VHD) files on FlexGroup volumes is supported regardless of the virtualization provider.
- Snapshot support was added in ONTAP 9.8 but is only available for use via the VMware vSphere APIs.
- VMware support limitations apply (for example, no pNFS).
- VMware and Fujitsu do not currently support a common multipath networking approach. For NFSv4.1, Fujitsu supports pNFS, whereas VMware supports session trunking. NFSv3 does not support multiple physical paths to a volume. For FlexGroup with ONTAP 9.8, our recommended best practice is to let ONTAP tools for VMware vSphere make the single mount, because the effect of indirect access is typically minimal (microseconds). It is possible to use round-robin DNS to distribute ESXi hosts across LIFs on different nodes in the FlexGroup, but this would require the FlexGroup to be created and mounted without ONTAP tools for VMware vSphere, and then the performance management features would not be available.
- Use ONTAP tools for VMware vSphere 9.8 to monitor performance of FlexGroup VMs using ONTAP metrics (dashboard and VM reports), and to manage QoS on individual VMs. These metrics are not currently available through ONTAP commands or APIs.
- SnapCenter for vSphere release 4.4 supports backup and recovery of VMs in a FlexGroup datastore on the primary storage system. Although SnapMirror can be used manually to replicate a FlexGroup to a secondary system, SCV 4.4 does not manage the secondary copies.

Databases on FlexGroup volumes

Usually, databases (such as Oracle) create a few small files when they are deployed. In a FlexGroup volume, small numbers of small files tend to favor local placement to their parent folder. This means that an Oracle deployment of eight databases might all land inside the same FlexGroup member volume. Not only does this provide no benefits from load distribution across nodes in a cluster, it can also present a problem as the files grow over time. Eventually, the files start to fill the member volume to capacity, and there is a need for remediation steps to move around data.

Database workloads, in theory, work well in a single namespace that can span a cluster. However, because the files are likely to grow over time and latency-sensitive databases might run on volumes that traverse the cluster network, Fujitsu currently recommends placing database files in FlexVol volumes.

Note

ONTAP 9.8 provides proactive resizing, which makes hosting large files/files that grow less of a concern, so database workloads on FlexGroup volumes become more realistic.

FlexCache volume considerations

ONTAP supports FlexCache. This feature provides a sparse volume that can accelerate performance for NAS workloads and prevent volume hot spots in a cluster or across a WAN. The FlexCache cache volume is powered by FlexGroup volumes, and the underlying protocol that redirects the pointers and blocks is the remote access layer (RAL). The RAL is also what makes a FlexGroup volume a FlexGroup volume. ONTAP 9.7 added support for FlexGroup origin volumes for FlexCache.

ONTAP 9.8 adds additional functionality for FlexCache volumes, including:

- SMB and multiprotocol NAS support
- 1:100 origin to cache ratio
- SnapMirror secondary origins
- Block-level invalidation
- Pre-population of a FlexCache

For more information about FlexCache, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays, ETERNUS HX series Hybrid Arrays FlexCache in ONTAP ONTAP 9.8](#).

FlexClone

Starting in ONTAP 9.7, FlexClone is supported for use with FlexGroup volumes. This feature provides storage administrators with a way to create instant, space-efficient copies (backed by Snapshot technology) of volumes to use for testing, development, backup verification, and a variety of other use cases. There are no specific considerations for use with FlexGroup volumes, except that a FlexClone copy of a FlexGroup volume uses the same number of member volumes as the FlexGroup parent volume. As a result, the volume count on a node can start to add up as FlexClone copies are created.

For example, if you have a FlexGroup volume that contains 16 member volumes and then create a FlexClone copy of that FlexGroup volume, you now have used 32 volumes in the system. Each new clone of the volume uses 16 member FlexVol volumes as well.

```
cluster::*> volume clone create -vserver DEMO -flexclone FGclone -type RW -parent-vserver DEMO -
parent-volume flexgroup_16

cluster::*> vol show -vserver DEMO -volume flexgroup_16*,FGclone* -fields name -sort-by name
vserver volume name-ordinal
-----
DEMO FGclone -
DEMO flexgroup_16 -
DEMO FGclone_0001 base
DEMO FGclone_0002 base
DEMO FGclone_0003 base
DEMO FGclone_0004 base
DEMO FGclone_0005 base
DEMO FGclone_0006 base
DEMO FGclone_0007 base
DEMO FGclone_0008 base
DEMO FGclone_0009 base
DEMO FGclone_0010 base
DEMO FGclone_0011 base
DEMO FGclone_0012 base
DEMO FGclone_0013 base
DEMO FGclone_0014 base
DEMO FGclone_0015 base
DEMO FGclone_0016 base
DEMO flexgroup_16_0001 base
DEMO flexgroup_16_0002 base
DEMO flexgroup_16_0003 base
DEMO flexgroup_16_0004 base
DEMO flexgroup_16_0005 base
DEMO flexgroup_16_0006 base
DEMO flexgroup_16_0007 base
DEMO flexgroup_16_0008 base
DEMO flexgroup_16_0009 base
DEMO flexgroup_16_0010 base
DEMO flexgroup_16_0011 base
DEMO flexgroup_16_0012 base
DEMO flexgroup_16_0013 base
DEMO flexgroup_16_0014 base
DEMO flexgroup_16_0015 base
DEMO flexgroup_16_0016 base
```

FlexClone to different storage virtual machine (SVM)

ONTAP allows you to create a FlexClone volume that spans different SVMs than the parent volume. This is done using the `-vserver` and `-parent-vserver` command options. This allows you to use the same export path for clients if you need to maintain mount paths.

Example:

```
cluster::*> vol clone create -vserver NFS -flexclone clone -type RW -parent-vserver DEMO
-parent-volume flexgroup -junction-path /flexgroup

cluster::*> vol show -junction-path /flexgroup -fields junction-path,volume,size
vserver volume size junction-path
-----
DEMO flexgroup 1PB /flexgroup
NFS clone 1PB /flexgroup
```

Volume rehost

ONTAP provides a method to quickly change the owning SVM for a volume via the volume rehost command. This is currently unsupported for use with FlexGroup volumes.

FlexClone deletion

When a FlexClone that is also a FlexGroup is deleted, that deletes multiple volumes in parallel. In most cases, this is not an issue, but bug 1368356 can potentially create an issue that causes new volume creations to fail until the condition is cleared. No data outages or access issues are caused by this, but for best results when using FlexClones with FlexGroup volumes, check which ONTAP release the bug is fixed in and use that release.

11. Encryption At-Rest

ONTAP supports Volume Encryption (VE) for FlexGroup volumes. Implementing this feature with FlexGroup volumes follows the same recommendations and best practices as stated for FlexVol volumes. Re-keying an existing FlexGroup volume is possible. Refer to ["Rekeying a FlexGroup volume or encrypting existing FlexGroup volumes"](#) for details.

Generally speaking, VE requires the following:

- A valid VE license
- A key management server
- A cluster-wide passphrase (32 to 256 characters)
- ETERNUS AX/HX series that supports AES-NI offloading

For information about implementing and managing VE with FlexGroup and FlexVol volumes, refer to [FUJITSU Storage ETERNUS AX/HX series Encryption Power Guide](#) and [FUJITSU Storage ETERNUS AX/HX series Scalability and Performance Using FlexGroup Volumes Power Guide](#).

ONTAP supports Aggregate Encryption (AE), which allows you to encrypt at the aggregate level. FlexGroup volumes can use AE, provided all aggregates that contain member volumes belonging to the same FlexGroup volume are encrypted.

Rekeying a FlexGroup volume or encrypting existing FlexGroup volumes

ONTAP supports both rekeying FlexGroup volumes and encrypting FlexGroup volumes that have not yet been encrypted. The process is the same as for a FlexVol volume.

Drive-level encryption (NSE and SED)

FlexGroup volumes can use NSE and SED disks, provided the FlexGroup volume spans only encrypted drives.

■ MetroCluster considerations

As described in ["MetroCluster"](#), if you plan on using VE/AE on a MetroCluster, you must complete the MetroCluster configuration before setting VE/AE.

12. FlexGroup Sample Designs

An ONTAP FlexGroup offers multiple benefits and can be managed like a normal FlexVol volume. The following design variations are examples of what is allowed with a FlexGroup volume.

■ FlexGroup volumes can:

- Share SVM as a FlexVol volume
- Share the same physical disks and aggregates as a FlexVol volume
- Be mounted to other FlexGroup or FlexVol volumes
- Be mounted below the FlexGroup level, similar to FlexVol volumes
- Share export policies and rules with FlexVol volumes
- Enforce quotas

FlexGroup volumes ideally should not:

- Be configured to span mixed disk or aggregate types (for example, member volumes of the same FlexGroup volume on SATA and SSD)
- Span nodes of different hardware types
- Span aggregates with uneven free capacity

Volume affinity and CPU saturation

To support concurrent processing, ONTAP assesses its available hardware at startup and divides its aggregates and volumes into separate classes called affinities. In general terms, volumes that belong to one affinity can be serviced in parallel with volumes that are in other affinities. In contrast, two volumes that are in the same affinity often must take turns waiting for scheduling time (serial processing) on the node's CPU.

A node's affinities are viewed with the advanced privilege nodeshell command `waffinity_stats -g`.

```
cluster::> set -privilege advanced

cluster::*> node run * waffinity_stats -g

Waffinity configured with:

# AGGR_VBN_RANGE affinities / AGGR_VBN affinity : 4
# VOL_VBN_RANGE affinities / VOL_VBN affinity : 4
# STRIPE affinities / STRIPEGROUP affinity : 9
# STRIPEGROUP affinities / VOL affinity : 1
# total AGGR_VBN_RANGE affinities : 8
# total VOL_VBN_RANGE affinities : 32
# total STRIPE affinities : 72
# total affinities : 149
# threads : 19
```

The sample NetApp FAS8080 EX node above is reporting that it can support fully concurrent operations on eight separate volumes simultaneously. It also says that to reach that maximum potential, it works best with at least two separate aggregates hosting four constituents each. Therefore, when you are building a new FlexGroup volume that is served by this node, that new FlexGroup volume should include eight constituents on this node evenly distributed across two local aggregates. Provisioning tools such as ONTAP System Manager attempts to take these affinities into account when creating new FlexGroup volumes, provided the FlexGroup size is adequate to span the available affinities and stay above the minimum 100GB member volume size.

The number of available affinities increased to eight per aggregate (two aggregates, 16 per node) for high-end platforms like the ETERNUS AX series:

```
cluster::*> node run * waffinity_stats -g

Waffinity configured with:
# AGGR_VBN_RANGE affinities / AGGR_VBN affinity :      8
# VOL_VBN_RANGE affinities / VOL_VBN affinity :      4
# STRIPE affinities / STRIPEGROUP affinity :      3
# STRIPEGROUP affinities / VOL affinity :      3
# total AGGR_VBN_RANGE affinities :      16
# total VOL_VBN_RANGE affinities :      64
# total STRIPE affinities :      144
# total affinities :      325
# threads :      18
# pinned :      0
# leaf sched pools :      18
# sched pools :      21
```

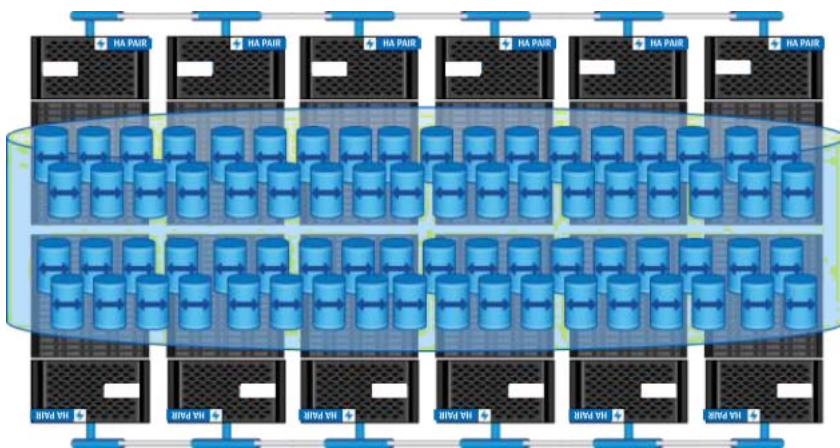
However, storage administrators usually do not need to worry about volume affinities, because ONTAP deploys a FlexGroup volume according to best practices for most use cases. For guidance on when you might need to manually create a FlexGroup volume, see the section above.

To simplify the experience, the `vol create -auto-provision-as flexgroup` command, the `flexgroup deploy` command, and the ONTAP System Manager GUI handle this setup for the storage administrator.

FlexGroup sample design 1: FlexGroup volume, entire cluster (24 nodes)

A FlexGroup volume can span an entire 24-node cluster, thus gaining the benefits of using all of the available hardware in the cluster with a single distributed namespace. In addition to using all your available hardware, you get the added benefit of gaining more potential capacity and more volume affinities in workloads.

Figure 66 FlexGroup volume, entire cluster (24 nodes)



Considerations

If you use an entire cluster to host a FlexGroup volume, keep in mind the information in ["Cluster considerations"](#).

Use cases

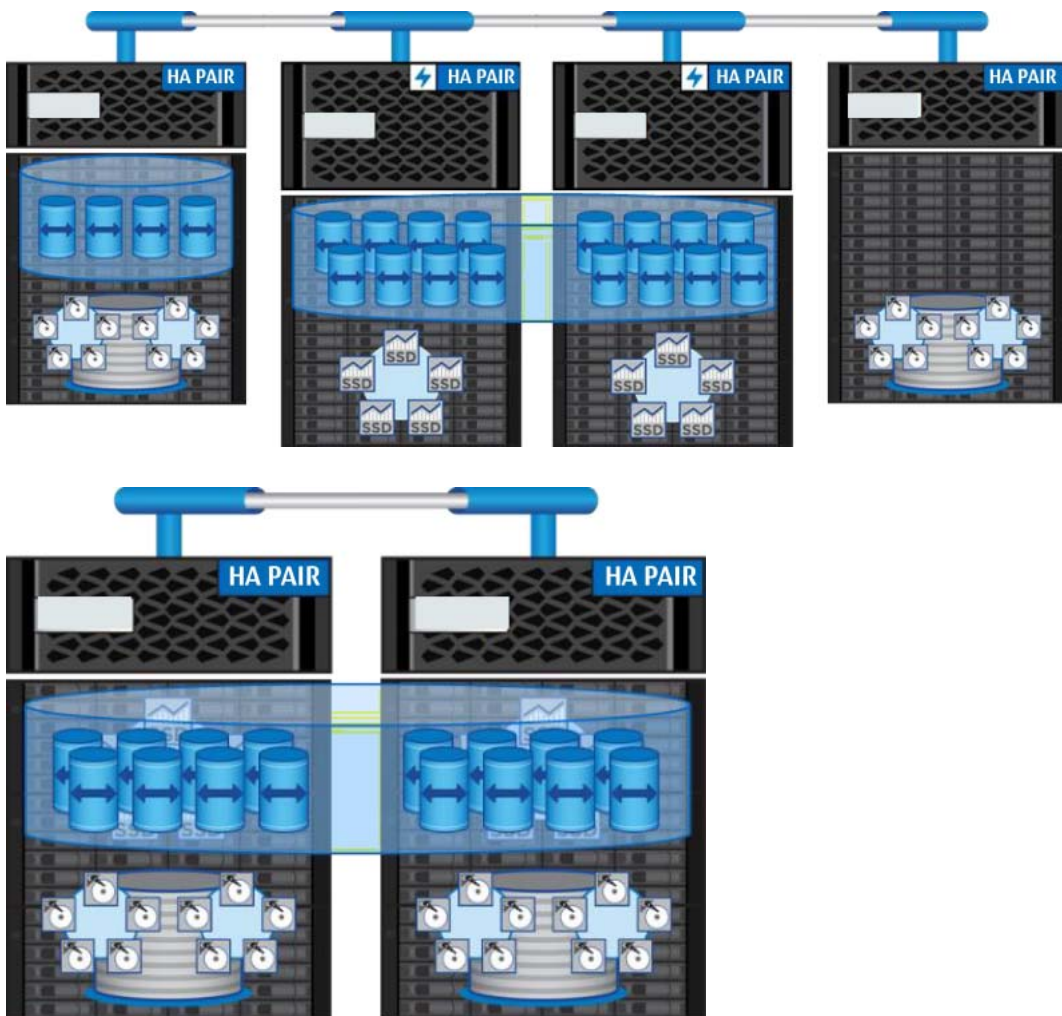
- Immense capacity (archives, scratch space, and media repositories)
- Workloads that require immense compute power in addition to storage (EDA)

FlexGroup sample design 2: multiple nodes, aggregates, partial cluster

Sometimes, storage administrators might not want to span a FlexGroup volume across the nodes of an entire cluster. The reasons include, but are not limited to, the following:

- Mix of hardware or ETERNUS HX series (some nodes are ETERNUS AX series)
- Mix of aggregate or disk types (that is, hybrid aggregates on the same node)
- Desire to dedicate nodes to specific tasks, storage tiers, or tenants
- In these scenarios, the FlexGroup volume can be created to use only specific aggregates, whether on the same node or on multiple nodes. If a FlexGroup volume has already been created, the member FlexVol volumes can be moved nondisruptively to the desired nodes and aggregates. For details, refer to "[When to use nondisruptive volume moves](#)".

Figure 67 Multiple nodes, partial cluster



Considerations

When you try to create a FlexGroup volume on a mix of nodes and aggregates, the automated commands are not of much use. Instead, use `volume create` or the GUI, where it is possible to specify aggregates on FlexGroup creation. For already-created FlexGroup volumes, the command line is the only option.

Use cases

- Mixed workloads (high performance and archive)
- Mixed cluster hardware
- Nodes with hybrid aggregates

FlexGroup sample design 3: FlexGroup, single node

An ONTAP cluster uses a robust back-end cluster network to pass reads and writes from a node that receives an I/O request on a data LIF to the node that owns the physical data. When traffic is remote, a small latency penalty is incurred (about 5% to 10%) for remote I/O as these packets are processed. When traffic is all local to the node that owns the data, no cluster back end is used. Also, NAS operations get special bypass consideration to direct requests to disk even faster, so there is a benefit to going locally to a node.

With FlexGroup, there is no manual intervention of control over where a data requests lands; ONTAP controls that portion for simplicity's sake. Because of this aspect, if a FlexGroup volume spans multiple nodes in a cluster, there is going to be indirect traffic over the cluster interconnects.

Although FlexGroup concurrency often more than outweighs any performance penalty for remote traffic, you can achieve some performance gains by isolating a FlexGroup volume to a single node. In addition, some deployments are performed on a single node to reduce the overall [failure domains](#) of the cluster.

[Figure 68](#) shows a single FlexVol volume that is accessed 100% locally on an AX4100 node versus a single FlexGroup volume with eight FlexVol members that is also accessed 100% locally. The test used was a Git clone during a compilation of the GCC library. The same testing equipment and data described in ["AFF A700 testing"](#) were used.

This test shows that a cluster-wide FlexGroup volume provides marginally better completion times because more hardware can be used. As extra threads are added to a local FlexGroup volume, the completion times start to get longer because the hardware cannot keep up as well. However, both FlexGroup volumes are two to three times faster than a local FlexVol volume and have a more gradual performance curve.

Figure 68 Git clone completion times comparison

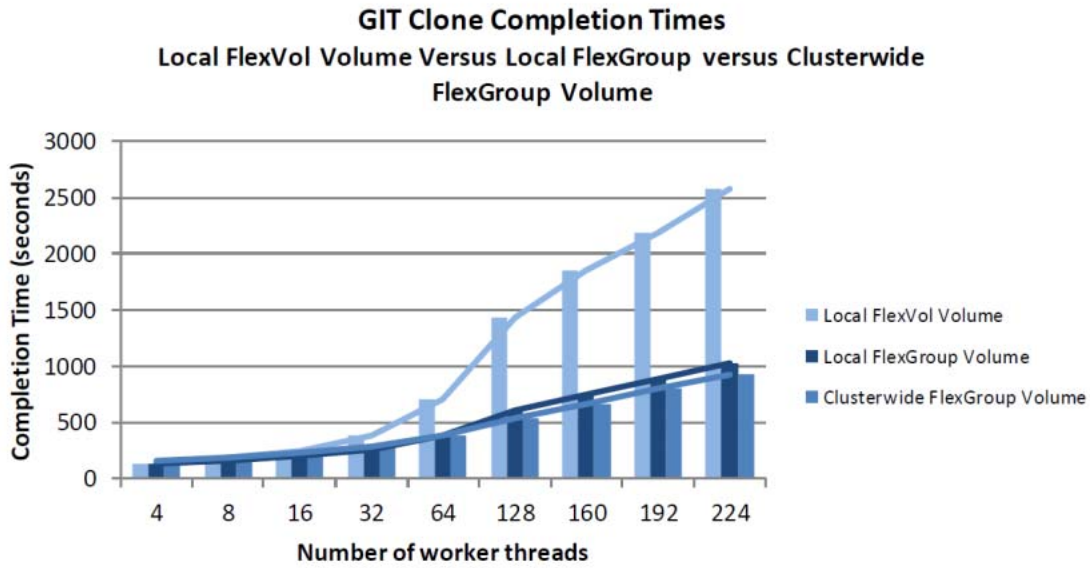
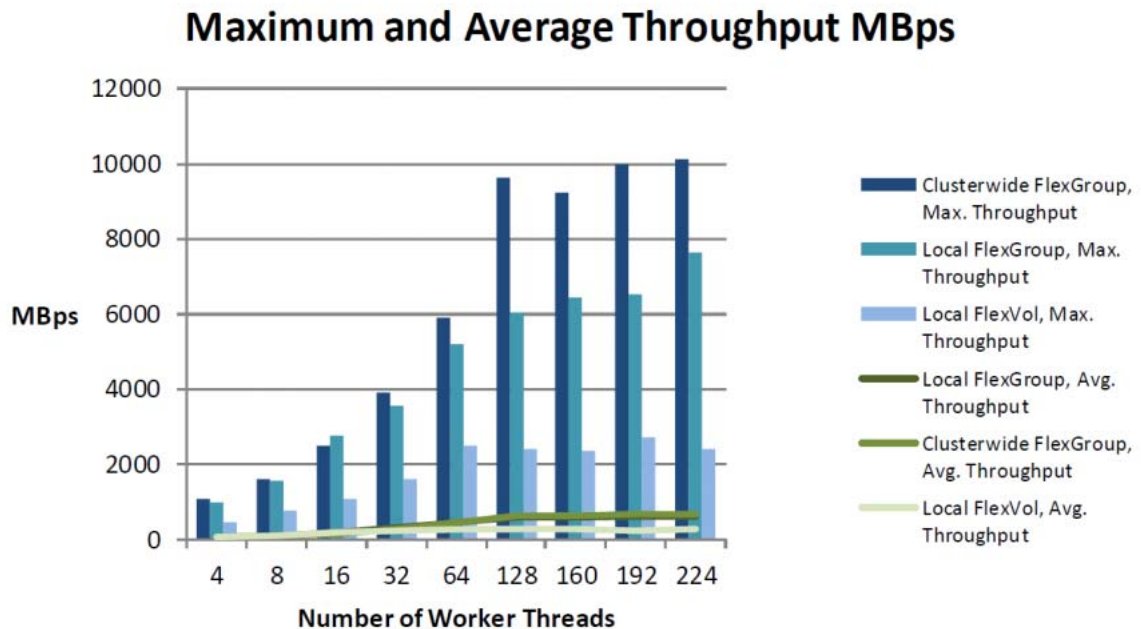


Figure 69 shows average and maximum throughput for the local FlexVol volume versus the local FlexGroup volume. For good measure, the cluster-wide FlexGroup volume was also added for comparison. The local FlexGroup volume shows better overall throughput than the cluster-wide FlexGroup volume until it reaches 16 threads. Then the all-local FlexGroup volume starts to lag behind slightly because the additional hardware allows the workload to push past the limits of a single node.

Figure 69 Average and maximum throughput comparison



In [Figure 70](#) and [Figure 71](#), we compare read and write throughput, respectively, with the local and cluster-wide FlexGroup volumes. At the 64-thread tipping point, the local FlexGroup volume starts to show a shift. Read throughput increases, while write throughput decreases. The cluster-wide FlexGroup volume shows the opposite trend.

Figure 70 Maximum read throughput comparison

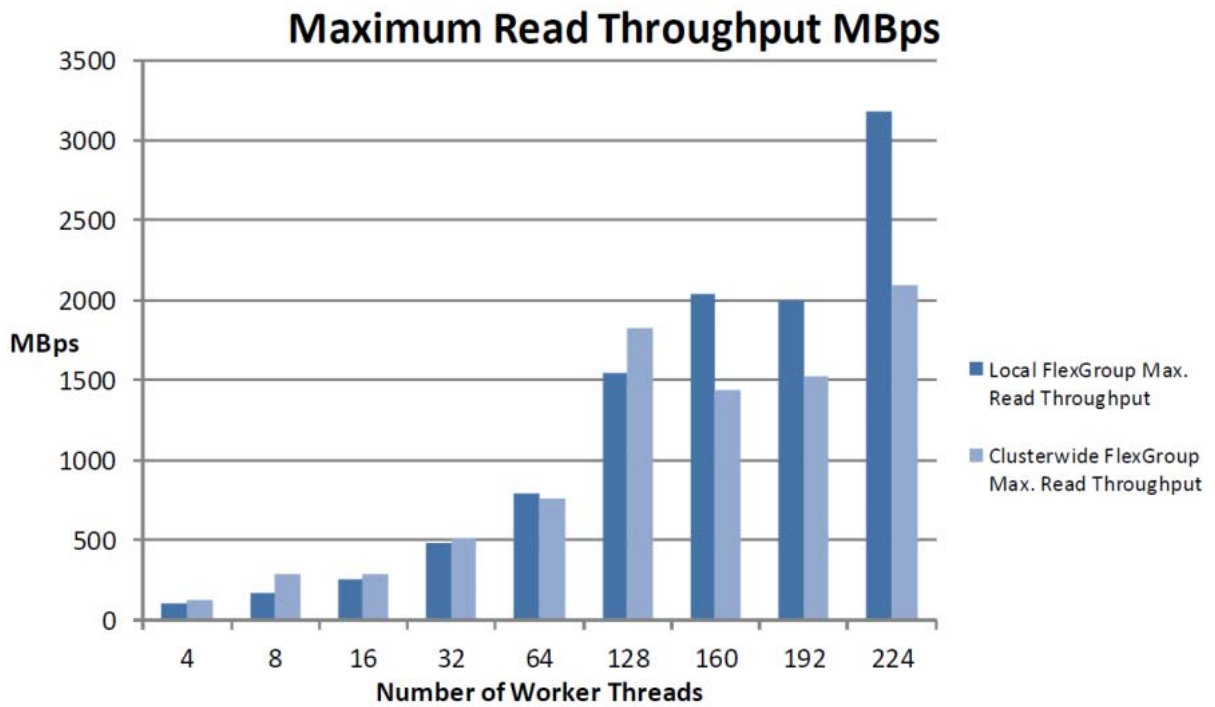


Figure 71 Maximum write throughput comparison

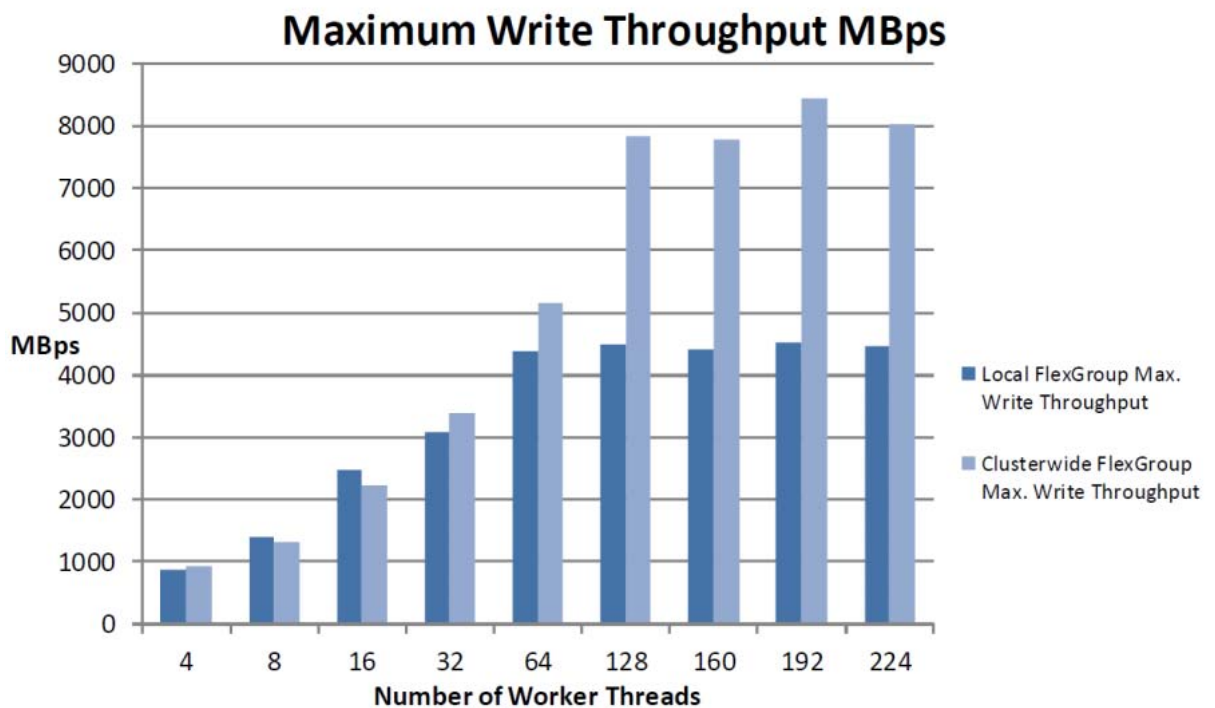
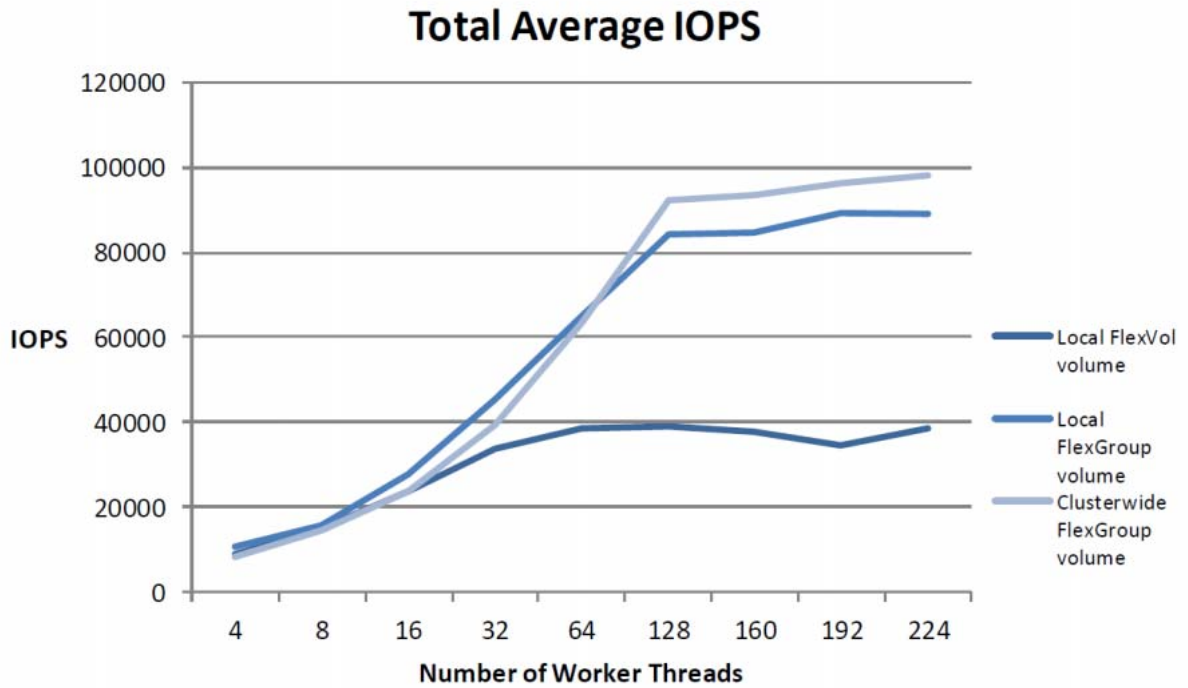


Figure 72 displays the total average IOPS for a local FlexVol volume versus the local and cluster-wide FlexGroup configurations. The FlexGroup configurations produce twice the IOPS that the FlexVol volume does, with the local FlexGroup volume outperforming the cluster-wide FlexGroup volume until the 64- thread tipping point.

Figure 72 Total average IOPS comparison



In this test, 64 worker threads appear to be a sweet spot. Let’s look at the average CPU utilization for a single-node FlexGroup volume versus a FlexGroup volume that spans the HA pair at just above 64 threads. Keep in mind that using more CPUs is a good thing; it means that work is being performed. That work is evidenced by the greater number of IOPS and the higher throughput for a FlexGroup volume that spans multiple nodes under the same workload.

Figure 73 Average CPU utilization, throughput, and IOPS for a FlexGroup volume—ETERNUS AX4100 HA pair, 128 threads



Figure 74 Average CPU utilization, throughput, and IOPS for a FlexGroup volume–single-node ETERNUS AX4100, 128 threads

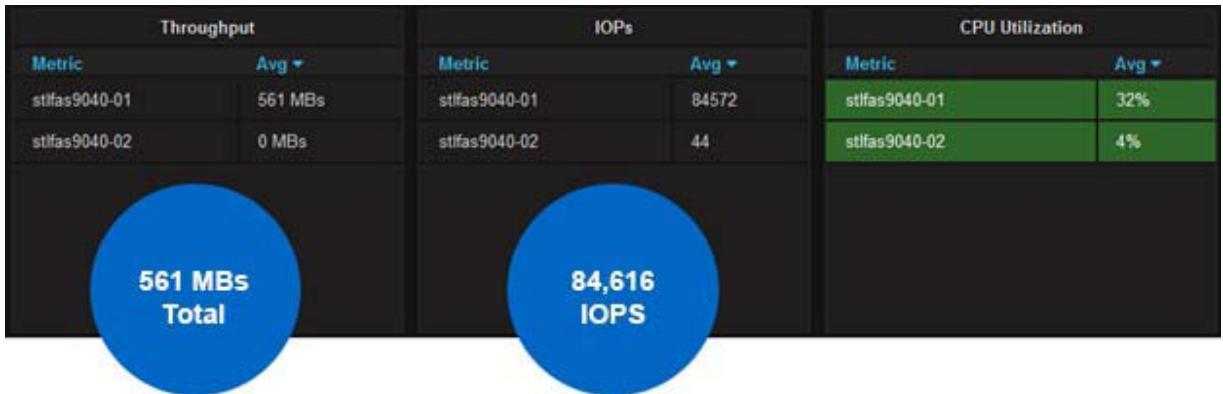
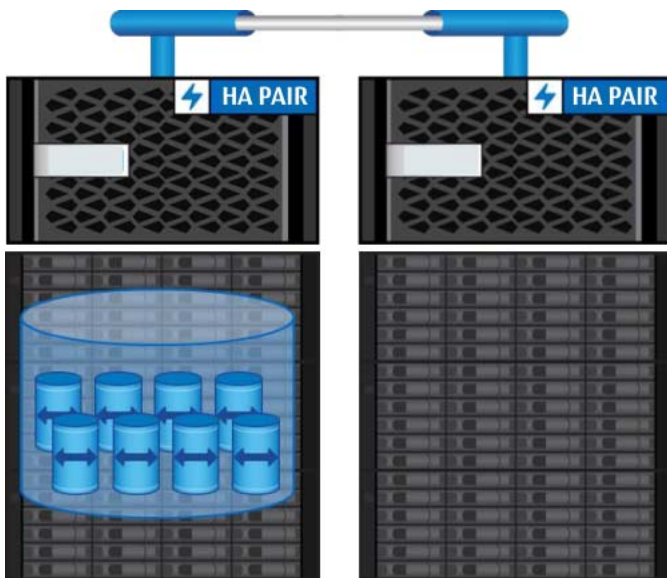


Figure 75 FlexGroup volume, single node



Considerations

When you use a single node for a FlexGroup volume, the gains that are realized by removing cluster interconnect traversal from the equation disappear relatively quickly. They disappear as load is added to the node and CPU, RAM, network bandwidth, and disk utilization becomes an issue. Usually, it makes more sense to spread the FlexGroup volume across multiple nodes rather than trying to save minimal cluster interconnect bandwidth.

Use Cases

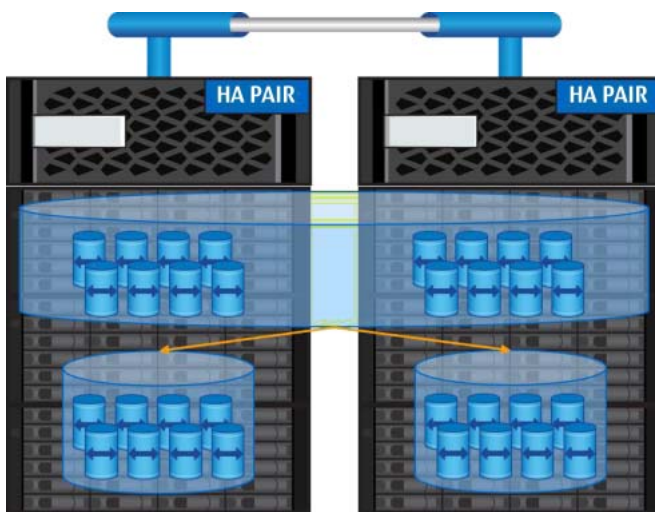
- High read workloads
- Need to isolate workloads to nodes
- Need to keep traffic off the cluster network

FlexGroup sample design 4: FlexGroup volumes mounted to FlexGroup volumes

With FlexVol volumes in ONTAP, you can mount volumes to other volumes to span the cluster and get >100TB in capacity, which was not possible with a single FlexVol volume. This method of designing a file system compares favorably with FlexGroup in terms of performance. However, the management overhead of creating multiple FlexVol volumes across multiple nodes and mounting them to each other in the namespace takes valuable personnel hours. In addition, scaling out capacity and performance can carry similar management headaches.

FlexGroup volumes can be managed like FlexVol volumes in the same way, by mounting a FlexGroup to another FlexGroup to create a folder structure with more granular data management.

Figure 76 FlexGroup volume mounted to FlexGroup volume



Considerations

Mounting FlexGroup volumes to other FlexGroup volumes offers flexibility, but at the cost of management overhead and additional member volume counts.

Use Cases

- More granular control over export policies and rules
- Greater control over the physical location of data
- Granular SnapMirror at the volume level for smaller datasets

FlexVol Volumes Mounted to FlexGroup Volumes

FlexVol volumes can also mount to FlexGroup volumes, and conversely. This configuration is another possibility with a FlexGroup solution.

Use Cases

- More granular control over export policies and rules
- Greater control over the physical location of data
- Features and functionality that are supported in FlexVol that aren't supported in FlexGroup use cases where a workload might occasionally create a large file or a small file that grows over time (for example, if a set of files gets zipped up to a larger zip file)

13. General Troubleshooting and Remediation

Failure scenarios

This section covers some failure scenarios and how a FlexGroup reacts.

Storage failovers

FlexGroup volumes are built on FlexVol volumes, so storage failover operations are functionally the same as for a FlexVol volume. Takeovers have no noticeable disruption. Nondisruptive upgrades, head swaps, rolling upgrades, and so on all perform normally. Givebacks of stateful protocols, such as SMB, are slightly disruptive, because of the transfer of locking states.

One caveat is that if an aggregate is not at home (not on the node that owns it, such as in a partial giveback state or if an aggregate has been relocated), FlexGroup volumes cannot be created until the aggregates are at home.

Note

In rare cases, if a node is powered off (dirty shutdown) in a cluster with four or more nodes, then the member volume cache entries on other nodes that host the FlexGroup volume might not flush properly and the FlexGroup appears to be hung because the surviving nodes are not made aware that the other member volumes have changed their locations. This issue is most commonly encountered during failover and resiliency testing. This is fixed in ONTAP 9.7P9 and later.

Network failures

If a network connection that is accessing a FlexGroup volume has an interruption or failure, the behavior for a FlexGroup volume mirrors that of a FlexVol volume. The cluster attempts to migrate the data LIF to a port or node that can access the network successfully. Clients may experience a brief disruption, as expected with network issues and depending on the protocol version in use.

Snapshot failures

If a FlexGroup Snapshot copy fails, ONTAP considers that Snapshot copy to be partial and invalidates it for SnapRestore operations. The Snapshot set is cleaned up by ONTAP and an EMS event is logged (`mgmtgwd.snapshot.partCreate`).

Hardware failures

Disk failures on aggregates hosting FlexGroup volumes operate the same as with a FlexVol; ONTAP fails the disk and selects a spare to use in a rebuild operation. If more disks in an aggregate fail than are allowed in a RAID configuration, then the aggregate is considered offline and the member volumes that live on the offline aggregate are inaccessible.

The main difference between a FlexVol and a FlexGroup here is that in a FlexGroup volume that spans multiple aggregates, access to other member volumes are fenced off to prevent data inconsistencies until the hardware issue has been addressed and the other member volumes are back online.

Node failures result in a storage failover event, where aggregates owned by the node that fails transfer ownership to the HA partner node, and the FlexGroup volume continues operations normally. If two nodes fail in the same HA pair, then the FlexGroup volume has member volumes that are considered to be offline, and data access is fenced off until the nodes are repaired and back in working order.

Time synchronization

If there is a time skew between the nodes hosting the members of a ONTAP FlexGroup, disruption might occur on the FlexGroup volumes. A time skew is a relative difference in local time between the nodes hosting a FlexGroup.

SMB/NFS protocol operations such as renaming a directory or symlink or unlinking a symlink trigger an internal cache invalidation resulting in a time-based calculation between nodes. If the time skew is large enough, incorrect repeat triggering of cache invalidation operations might occur. The repeat cache invalidation operations prevent the completion of the rename and unlink operations. The messages continue to retry and might affect operations across the FlexGroup and cause disruption.

This issue is fixed in ONTAP 9.8. If you are running an ONTAP release prior to ONTAP 9.8, make sure that the cluster node times are in sync, ideally with a Network Time Protocol (NTP) configuration. ONTAP 9.7 P6 introduces an EMS notification to alert storage administrators if this issue is occurring. To obtain an ONTAP release, use the [Fujitsu download site](#).

14. Capacity Monitoring and Alerting

This chapter covers various methods of monitoring a FlexGroup volume's capacity, including viewing total storage efficiency savings. Monitoring FlexGroup capacity is also possible with the FPolicy support.

Capacity monitoring and alerting becomes less of a concern with ONTAP 9.8's [proactive resizing](#) feature, because the total FlexGroup free space should more closely mirror the individual member volume free space.

Capacity monitoring and alerting with the command line

When you use thin provisioning, use the command `storage aggregate show-space with volume show -is-constituent true, volume show-space, and storage aggregate show` to get better total visibility into FlexGroup volume space usage overall. In the command line, you can also use the `-sort-by` option to organize the list.

Note

To get an accurate portrayal of the space that is being used, pay attention to the `Physical Used` portion of the `volume show-space` command. You can find an example in ["Command Examples"](#).

Event management system messages

Event management system messages alert storage administrators about the capacity of volumes in ONTAP. The messages are listed in this section. You can view them in the command line with the command `event route show -messagename [message] -instance`. For an example of these messages, refer to ["Examples of capacity-related event management system messages"](#).

- Unmodifiable values:
 - Severity level
 - Corrective actions
 - Description
 - SNMP support
- Modifiable values:
 - Destinations
 - Allowed drops or intervals between transmissions

When an event management system message that has SNMP support is triggered, an SNMP trap fires to the configured SNMP server. This action is specified through the `destinations` value. For more information about configuring event management system destinations, see the Express Guide for your specific version of ONTAP.

The default values for Nearly Full (Warning) and Full (Error) are as follows:

```
cluster::*> vol show -vserver SVM -volume flexgroup -fields space-nearly-full-threshold-
percent,space-full-threshold-percent
vserver volume      space-nearly-full-threshold-percent  space-full-threshold-percent
-----
SVM      flexgroup  95%                                98%
```

Event management system messages for `volume.full` look like the following:

```
11/28/2016 18:26:34 cluster-01
DEBUG monitor.volume.full: Volume flexgroup@vserver:05e7ab78-2d84-11e6-a796-00a098696ec7
is full (using or reserving 99% of space and 0% of inodes).
```

In the preceding example, the following values are provided:

- The type of object
- The name of the volume
- The SVM (called `vserver` in the CLI) universal unique identifier (UUID)
- Percentage of space used
- Percentage of inodes used

You can use these values when testing event management system messages. When you look for which SVM is affected by the errors, use the UUID string at the advanced privilege level:

```
cluster::*> vserver show -uuid 05e7ab78-2d84-11e6-a796-00a098696ec7
Vserver      Type      Subtype    Admin      Operational  Root
-----
SVM          data     default    running    running      vsroot      aggr1_node1
```

■ Testing event management system messages

To test an event management system message, use the `event generate` command (available at the `diag` privilege level). Each message has a unique string of values. The values for `volume.full` and `volume.nearlyFull` are listed in the preceding section. The following example shows how to construct a test message for a `volume.nearlyFull` event and the resulting event management system message:

```
cluster::*> event generate -message-name monitor.volume.nearlyFull -values Volume flexgroup
@vserver:05e7ab78-2d84-11e6-a796-00a098696ec7 95 0

cluster::*> event log show -message-name monitor.volume.nearlyFull
Time          Node          Severity      Event
-----
11/28/2016 18:36:35 cluster-01
                                     ALERT          monitor.volume.nearlyFull: Volume
flexgroup@vserver:05e7ab78-2d84-11e6-a796-00a098696ec7 is nearly full (using or reserving 95% of
space and 0% of inodes).
```

■ Modifying the Volume Full and Nearly Full thresholds

With a FlexVol volume, the default values for full and nearlyFull are fine because the volume is isolated to a single container. With a FlexGroup volume, by the time a member FlexVol volume reaches the full or nearly full threshold, the application or end user might already be seeing a performance degradation. This decreased performance is due to increased remote file allocation or a FlexGroup volume that is already reporting to be out of space because of a full or nearly full member volume.

ONTAP 9.8 and later versions don't need the same level of aggressiveness, as features such as [proactive resizing](#) take much of the guesswork out of capacity management.

To help monitor for these scenarios, the volume full and nearly full thresholds might require adjustment to deliver warnings and errors before a volume fills up. Volumes have options to adjust these thresholds at the admin privilege level.

```
-space-nearly-full-threshold-percent
-space-full-threshold-percent
```

Use the `volume modify` command to adjust these thresholds.

Best Practice 24: Volume Space Threshold Recommendations for FlexGroup

Generally speaking, the nearly full threshold of a FlexGroup volume should be set to 80%, and full should be set to 90%. With these settings, you have enough time to remediate space issues by increasing the FlexGroup volume size or adding more capacity in the form of additional member volumes through [volume expand](#). These values can vary based on the average file size and the FlexGroup member volume size.

For instance, a 1TB FlexGroup member can reach 80% immediately with an average file size of 800GB, but a 100TB FlexGroup member would take longer to hit that threshold. Refer to ["Nonideal workloads – large files"](#) for guidance on large file workloads.

For examples using Active IQ to monitor and alert for capacity and inodes, refer to ["Capacity monitoring and alerting examples in Active IQ Unified Manager"](#).

Client-side capacity considerations with thin provisioning

When using a FlexGroup volume, the client usually reports the available space, the used space, and so on in a way that reflects what the storage administrator has provisioned. This reporting is especially true when the volume space guarantee is set to `volume`, because ONTAP returns the expected capacities to the client.

However, when you use thin provisioning and overprovisioning for your physical storage, the client values do not reflect the expected used capacity of the FlexGroup volume. Instead, they reflect the used capacity in the physical aggregate. This approach is no different from the behavior of FlexVol volumes.

In the following example, there are three FlexGroup volumes:

- `flexgroup` has 80TB allocated and is thin provisioned across two aggregates with about 10TB available.
- `flexgroup4TB` has 4TB allocated with a space guarantee of `volume`.
- `flexgroup4TB_thin` has 4TB allocated and is thin provisioned across two aggregates with about 4TB available.

The following output shows that the cluster sees the proper used space in the volumes.

```
cluster::> vol show -fields size,used,percent-used,space-guarantee,available -vserver SVM
-volume flexgroup*,!*__0* -sort-by size
vserver volume          size available used      percent-used space-guarantee
-----
SVM    flexgroup4TB          4TB  3.77TB  30.65GB  5%          volume
SVM    flexgroup4TB_thin    4TB  3.80TB  457.8MB  5%          none
SVM    flexgroup             80TB 10.13TB  5.08GB  87%          none
3 entries were displayed.
```

However, the client sees the used capacity of the overprovisioned FlexGroup volume named `flexgroup` as 66TB, rather than the 5GB that is seen on the cluster. This total includes the total available size of the physical aggregate (5.05TB + 5.08TB = ~10TB) and subtracts that from the total size.

The volumes that are not overprovisioned report space normally.

```
# df -h
Filesystem                Size      Used Avail Use% Mounted on
10.193.67.220:/flexgroup  76T       66T   11T   87% /flexgroup
10.193.67.220:/flexgroup4TB  3.9T      31G    3.8T    1% /flexgroup4TB
10.193.67.220:/flexgroup4TB_thin 3.9T      230M    3.8T    1% /flexgroup4TB_thin

cluster::*> aggr show -aggregate aggr1* -fields usedsize,availsize,percent-used,size
aggregate  availsize percent-used size  usedsize
-----
aggr1_node1 5.05TB    36%          7.86TB 2.80TB
aggr1_node2 5.08TB    35%          7.86TB 2.78TB
2 entries were displayed.
```

14. Capacity Monitoring and Alerting

Viewing FlexVol member capacity from the ONTAP command line

The ~11TB of available space comes from the way that the Linux client calculates the space. This client does 1K blocks, so the number 10881745216 is divided into factors of 1,000. ONTAP uses factors of 1,024 to calculate space.

```
# df | grep flexg
10.193.67.220:/flexgroup          85899345920 75017600704 10881745216 88% /flexgroup
10.193.67.220:/flexgroup4TB      4080218944   32143296   4048075648  1% /flexgroup4TB
10.193.67.220:/flexgroup4TB_thin 4080218944    468736   4079750208  1% /flexgroup4TB_thin
```

Also, the size portion of the output considers the default 5% that is allocated for Snapshot space. That's why 80TB becomes 76TB in the preceding `df` output.

```
cluster::> vol show -fields size,percent-snapshot-space -vserver SVM -volume flexgroup*,!*__0*
-sort-by size
vserver volume          size percent-snapshot-space
-----
SVM    flexgroup4TB          4TB  5%
SVM    flexgroup4TB_thin    4TB  5%
SVM    flexgroup             80TB 5%
3 entries were displayed.
```

When the Snapshot space allocation is reduced to 0, `df` reports a more normalized version of the actual size (but still has the strangeness of the `used` space).

```
cluster::> vol modify -vserver SVM -volume flexgroup -percent-snapshot-space 0
[Job 2502] Job succeeded: volume modify succeeded

# df -h | grep flexgroup
Filesystem          Size Used Avail Use% Mounted on
10.193.67.220:/flexgroup  80T  70T  11T  88% /flexgroup
```

Windows capacity reporting

Windows reports in very much the same way as the Linux clients. The difference is that Windows uses a factor of 1,024, so the numbers are closer to the ONTAP values.

Viewing FlexVol member capacity from the ONTAP command line

When FlexGroup volumes are created, each member is evenly divided according to the total capacity and the number of FlexVol members. For example, in the case of an 80TB FlexGroup volume, the FlexVol members are 10TB apiece. To view member volume capacity, use the `volume show` command at the `diag` privilege level; use `volume show -is-constituent true` or use the `volume show-space` command at the `admin` privilege level.

Viewing FlexVol member capacity is useful when you are trying to determine the true available space in a FlexGroup volume. When a FlexGroup volume reports total available space, it considers the total available space on all member volumes. However, when an individual member volume fills to capacity, the entire FlexGroup volume reports as out of space, even if other member volumes show available space. To mitigate this scenario, the FlexGroup ingest algorithms attempt to direct traffic away from a volume that becomes more heavily used than other volumes.

FlexGroup capacity viewer

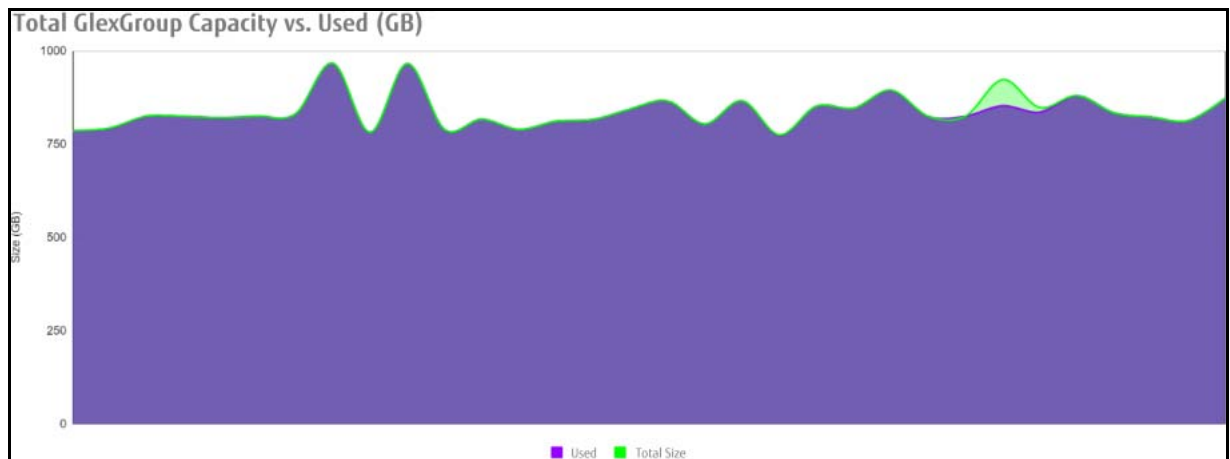
You can also view capacity and inode usage for a FlexGroup volume via custom Google sheets. You can request access for your own copy with an email to Fujitsu Support and including your cluster serial number and the name of a FlexGroup volume you would like to see graphed out.

The Google sheet graphs include:

- Holistic views of the FlexGroup used and total space
- Snapshot reserve and used graphs
- Member volume level views of used and total capacity
- Inode and used capacity trend lines

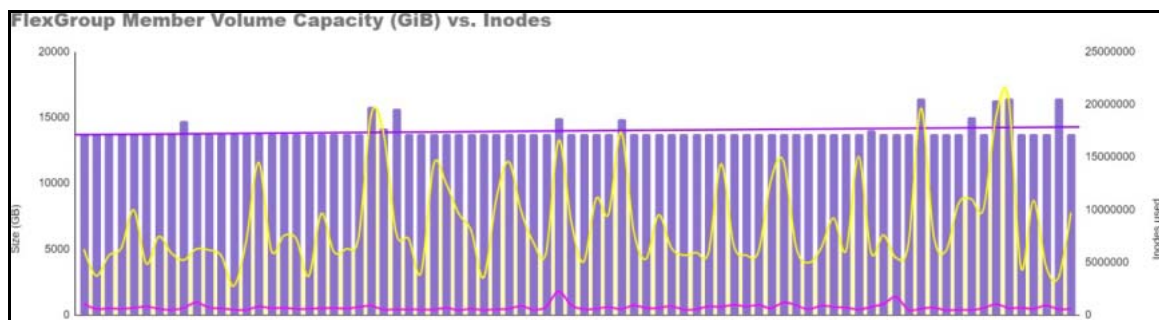
Here is a view of a FlexGroup volume with an uneven capacity usage that also has elastic resizing in action. In this case, the FlexGroup needs total capacity to be increased.

Figure 77 Google Sheet – FlexGroup capacity view



This view shows a FlexGroup volume capacity at the member volume level and includes a trend line of used inodes. Ideally, the inodes used trend in parallel with the used capacity. In this case, we see spikes in capacity per member volume that do not correspond with counts for used inodes, meaning some member volumes might have larger files than others.

Figure 78 Google Sheet – FlexGroup member capacity view



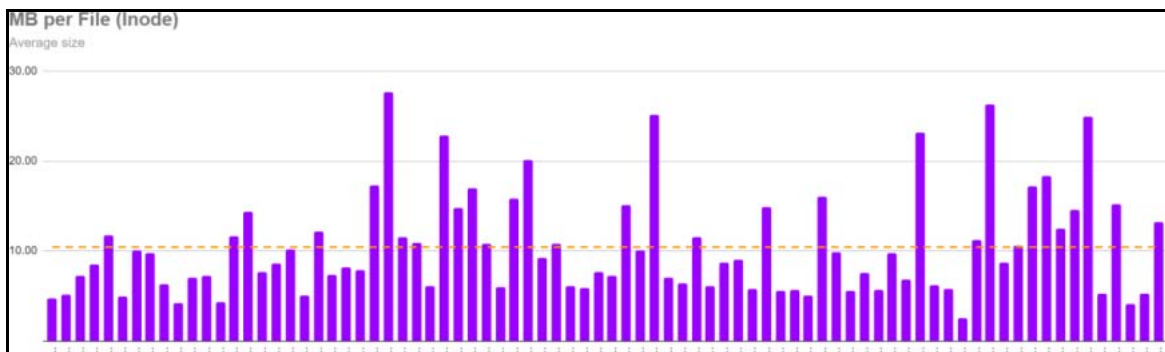
There are also graphs that attempt to take the total used capacity and divide by the used inodes to attempt to show an average file size. It's not an exact science, but gets fairly close.

In the following example, we can see that the purple bar graphs have widely disparate values, which means that the average file sizes have a large range. The orange line is the average inode size for the entire FlexGroup volume. Further investigation may be needed for this FlexGroup to discover why there is such a wide range of average inode sizes.

Note

File size disparity and capacity imbalances do not necessarily indicate a problem, but these graphs can be used if there is a problem and it needs to be identified. Refer to ["Data imbalances in FlexGroup volumes"](#) for more information.

Figure 79 Google Sheet – Average inode size



Note

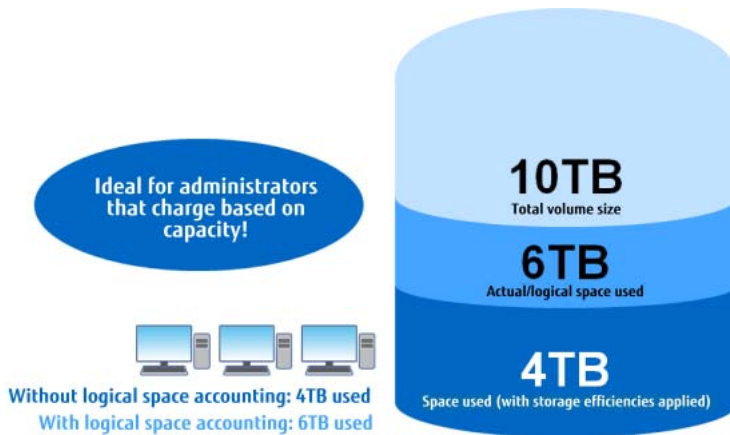
Request a copy of the FlexGroup Capacity Viewer by emailing Fujitsu Support with your cluster serial number and FlexGroup volume name.

Logical space accounting

Logical space accounting enables storage administrators to mask storage efficiency savings so that end users avoid overallocating their designated storage quotas.

For example, if a user writes 6TB to a 10TB volume and storage efficiencies save 2TB, logical space accounting can control whether the user sees 6TB or 4TB.

Figure 80 How logical space accounting works



Currently, FlexGroup volumes do not support this functionality; it is only available for FlexVol volumes.

Monitoring FlexGroup performance

FlexGroup performance can be monitored in many of the same ways that a normal FlexVol volume's performance can be monitored. The same concepts of CPU utilization, disk saturation, NVRAM bottlenecks, and other WAFL-related performance characteristics apply. Also, NAS performance monitoring doesn't change. You still use the basic CIFS/SMB and NFS statistics that you always have.

The main difference with monitoring FlexGroup performance is that you must consider multiple nodes, aggregates, member FlexVol constituent volumes, and the notion of remote placement of files and folders. These elements add another layer to consider when you want to monitor and isolate performance issues.

Monitoring performance from the command line

From the command line, you have several ways to view performance statistics.

Real-time performance monitoring

To monitor system performance in real time, use the `statistics show-periodic` command.

```
cluster::*> statistics show-periodic ?
[[-object] <text>]          *Object
[ -instance <text> ]       *Instance
[ -counter <text> ]        *Counter
[ -preset <text> ]         *Preset
[ -node <nodename> ]      *Node
[ -vserver <vserver name> ] *Vserver
[ -interval <integer> ]   *Interval in Seconds (default: 2)
[ -iterations <integer> ] *Number of Iterations (default: 0)
[ -summary {true|false} ] *Print Summary (default: true)
[ -filter <text> ]         *Filter Data
```

This command provides an up-to-date glimpse into system performance. Leaving the default values alone gives you a cluster-wide view. Specifying an SVM gives you a more granular look, but mainly at the counters that would be specific to an SVM, such as NAS counters, rather than to CPU or disk. When you use SVM-specific statistics, defining the counters that are provided for the object helps reduce the noise on the CLI. You can also get real-time FlexGroup statistics for the ratios of local to remote top-level directories (tld), high-level directories (hld), regular directories, and files.

For examples of these commands, refer to ["Command Examples"](#).

The FlexGroup statistics also can show various other information and can be gathered over a period of time if you initiate a `statistics start -object flexgroup` command. This command collects statistics over time that can be captured in iterations through an automated tool such as Perfstat or perfarchives.

```
cluster::*> statistics start -object flexgroup
Statistics collection is being started for sample-id: sample_69197
```

Use the following to view the statistics:

```
cluster::*> statistics show -object flexgroup -instance 0

Object: flexgroup
Instance: 0
Start-time: 11/30/2016 16:44:42
End-time: 11/30/2016 17:42:57
Elapsed-time: 3495s
Scope: cluster-01

Counter                                     Value
-----
cat1_tld_remote                             2
cat2_hld_local                               180
cat2_hld_remote                             1292
cat3_dir_local                              146804
cat3_dir_remote                              283
cat4_fil_local                              734252
cat4_fil_remote                              1124
groupstate_analyze                          12232
groupstate_update                           86242
instance_name                                0
node_name                                    cluster-01
process_name                                 -
refreshclient_create                         5241
refreshclient_delete                         5241
refreshserver_create                         5244
refreshserver_delete                         5244
```

The statistics capture gives a nice summary of the percentages of remote file and directory placement in the Flex-Group volume when it spans multiple nodes. In the following example, the values are 14% remote directories and 1% remote files.

| | |
|--------------|----|
| remote_dirs | 14 |
| remote_files | 1 |

Protocol statistics

It's also possible to get a glimpse of how individual NAS protocols are influencing performance. Simply use the statistics start command to include NFS or SMB performance counters in the capture. You'll get more options with diag privileges.

```
cluster::*> statistics start -object nfs
nfs_credstore          nfs_exports_access_cache
nfs_exports_cache      nfs_exports_match
nfs_file_session_cache nfs_file_session_cache:constituent
nfs_generic            nfs_idle_conn
nfs_idle_total_conn    nfs_qtree_export
nfs_server_byname      nfserr
nfsv3                  nfsv3:constituent
nfsv3:cpu              nfsv3:node
nfsv4                  nfsv4:constituent
nfsv4:cpu              nfsv4:node
nfsv4_1                nfsv4_1:constituent
nfsv4_1:cpu            nfsv4_1:node
nfsv4_1_diag           nfsv4_1_error
nfsv4_diag             nfsv4_error
nfsv4_spinnp_errors

cluster::*> statistics start -object smb
smb1  smb1:node      smb1:vserver smb1_ctx  smb1_ctx:node
smb2  smb2:node      smb2:vserver smb2_ctx  smb2_ctx:node

cluster::*> statistics start -object cifs
cifs          cifs:node
cifs:vserver  cifs_cap
cifs_cap:constituent cifs_client
cifs_client:constituent cifs_ctx
cifs_ctx:node      cifs_shadowcopy
cifs_unsupp_ioctl  cifs_unsupp_ioctl:constituent
cifs_watch
```

flexgroup show

During FlexGroup I/O, you can also view the member constituent usage and balance through the nodeshell command `flexgroup show`. The command also provides other information that can be useful, such as how often a member volume might be avoided for new files. Be sure to capture this command output if you run into a FlexGroup issue and need to open a support case.

```
cluster::*> node run * flexgroup show Tech_ONTAP
FlexGroup 0x80F03868 (Tech_ONTAP)
* next snapshot cleanup due in 9334 msec
* next refresh message due in 334 msec (last to member 0x80F0386E)
* spinnp version negotiated as 10.13, capability 0x3F7F
* Ref count is 8
* ShouldEnforceQuotas true
* IsAnyMemberInNvfailedState false
* reaction +0.0, workload +0.0, activity level 0, cv 0%
Idx  Member L      Used      Avail Urgc  Target      Probabilities  D-Ingest Alloc  F-Ingest Alloc
-----
1    4503 L 2238647  0% 318698244  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
2    4369 R 3239783  1% 318638088  0% 12.49%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
3    4674 L 2011415  0% 318697586  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
4    4477 R 2334885  0% 318694396  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
5    4329 L 2250619  0% 318697596  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
6    4370 R 2255368  0% 318697148  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
7    4675 L 2252390  0% 318697125  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
8    4478 R 2201995  0% 318698611  0% 12.50%  [100% 100% 87% 87%]  0+ 0  0  0+ 0  0
```

Output breakdown for the flexgroup show command

The `flexgroup show` command has a series of values that might not be intuitive at first glance. [Table 20](#) describes those values and how to interpret them.

Table 20 flexgroup show output column definitions

| Column | Definition |
|----------------------|---|
| Idx | Index number of the member volume. |
| Member | DSID of the FlexGroup member. |
| L | Local or remote to the node. |
| Used | Number and overall percentage of 4K blocks used. |
| Urgc | Urgency: Probability of a file or directory creation being allocated to a remote member volume to avoid premature ENOSPC in a member volume. This value increases according to how close to 100% used a volume's capacity is. |
| Targ | Target: Percentage of what new content should be placed on a member volume as related to its peers. The total summation of all target percentages equal ~100%. |
| Probabilities | The likelihood that a member volume is avoided for use. This number increases according to how full a member volume becomes in relation to other member volumes (tolerance). |
| D-Ingest and D-Alloc | Directory ingest and directory allocation, respectively; how many directories have been allocated to a local member volume. |
| F-Ingest and F-Alloc | File ingest and file allocation, respectively; how many files have been allocated to a local member volume. |

You should run the `flexgroup show` command during a period of I/O activity on a FlexGroup volume. This command gives the following useful information:

- How evenly the traffic is distributed across members
- How evenly distributed the space is on members
- How likely a member volume is to be used for ingesting
- The ratio of directory to file creation in a workload
- The member volume's node locality

Performance archiver

Performance data is captured for support issues through the performance archiver, which runs by default in ONTAP.

Monitoring performance (Active IQ Unified Manager)

A more palatable and widely available tool for monitoring the performance of a FlexGroup volume is Active IQ Unified Manager. This tool is available as a free .ova file or as a Linux installation from the [Fujitsu download site](#).

Active IQ Unified Manager offers both real-time and historical performance information to provide a single monitoring point. Active IQ Unified Manager can give granular performance views for the entire FlexGroup volume or for individual member constituent FlexVol volumes.

[Figure 81](#) is a capture of a simple file creation script on a single Linux VM, so the performance benefits of FlexGroup are not seen here. However, the figure does provide a sense of what Unified Manager can deliver.

Figure 81 Active IQ Performance Manager graphs



15. FlexGroup Data Protection Best Practices

For the FlexGroup data protection best practices, refer to [FUJITSU Storage ETERNUS AX series All-Flash Arrays](#), [ETERNUS HX series Hybrid Arrays Data Protection and Backup for ONTAP FlexGroup Volumes](#).

16. Migrating to ONTAP FlexGroup

One challenge of having many files or a massive amount of capacity is deciding how to effectively move the data as quickly and as nondisruptively as possible. This challenge is greatest in high-file-count, high-metadata-operation workloads. Copies of data at the file level require file-system crawls of the attributes and the file lists, which can greatly affect the time that it takes to copy files from one location to another. That duration does not account for other aspects such as network latency, WANs, system performance bottlenecks, or other things that can make a data migration painful.

With ONTAP FlexGroup, the benefits of performance, scale, and manageability are apparent. Data migrations can take three general forms when dealing with FlexGroup:

- Migrating from non-Fujitsu (third-party) storage to FlexGroup
- Migrating from Data ONTAP operating in 7-Mode to FlexGroup
- Migrating from FlexVol volumes or SAN LUNs in ONTAP to FlexGroup

Data migrations to FlexGroup volumes are the best way to migrate. FlexGroup volume migrations currently cannot be performed with the following methods:

- FlexVol to FlexGroup volume move
- SnapMirror or SnapVault between FlexVol and FlexGroup
- 7-Mode Transition Tool (CBT and CFT)

The following sections cover different migration use cases and how to approach them.

Migration using NDMP

In ONTAP 9.7 and later, FlexGroup volumes now support NDMP operations. These include the `ndmpcopy` command, which can be used to migrate data from a FlexVol to a FlexGroup volume. For information about setting up `ndmpcopy`, see:

[How to run ndmpcopy in Clustered Data ONTAP](#)

In the following example, `ndmpcopy` was used to migrate around five million folders and files from a FlexVol to a FlexGroup volume. The process took around 51 minutes:

```
cluster::*> system node run -node ontap9-tme-8040-01 ndmpcopy -sa ndmpuser:AcDjtsU827tputjN -da
ndmpuser:AcDjtsU827tputjN 10.x.x.x:/DEMO/flexvol/nfs 10.x.x.x:/DEMO/flexgroup_16/ndmpcopy
Ndmpcopy: Starting copy [ 2 ] ...
Ndmpcopy: 10.x.x.x: Notify: Connection established
Ndmpcopy: 10.x.x.x: Notify: Connection established
Ndmpcopy: 10.x.x.x: Connect: Authentication successful
Ndmpcopy: 10.x.x.x: Connect: Authentication successful
Ndmpcopy: 10.x.x.x: Log: Session identifier: 12584
Ndmpcopy: 10.x.x.x: Log: Session identifier: 12589
Ndmpcopy: 10.x.x.x: Log: Session identifier for Restore : 12589
Ndmpcopy: 10.x.x.x: Log: Session identifier for Backup : 12584
Ndmpcopy: 10.x.x.x: Log: DUMP: creating "/DEMO/flexvol/./snapshot_for_backup.1" snapshot.
Ndmpcopy: 10.x.x.x: Log: DUMP: Using subtree dump
Ndmpcopy: 10.x.x.x: Log: DUMP: Using snapshot_for_backup.1 snapshot
Ndmpcopy: 10.x.x.x: Log: DUMP: Date of this level 0 dump snapshot: Thu Jan 9 11:53:18 2020.
Ndmpcopy: 10.x.x.x: Log: DUMP: Date of last level 0 dump: the epoch.
Ndmpcopy: 10.x.x.x: Log: DUMP: Dumping /DEMO/flexvol/nfs to NDMP connection
... (output omitted for length)
Ndmpcopy: 10.x.x.x: Notify: dump successful
Ndmpcopy: 10.x.x.x: Log: RESTORE: RESTORE IS DONE
Ndmpcopy: 10.x.x.x: Notify: restore successful
Ndmpcopy: Transfer successful [ 0 hours, 50 minutes, 53 seconds ]
Ndmpcopy: Done
```

The same dataset using `cp` over NFS took 316 minutes—six times as long as `ndmccopy`:

```
# time cp -R /flexvol/nfs/* /flexgroup/nfscp/

real    316m26.531s
user    0m35.327s
sys     14m8.927s
```

Using the XCP Migration Tool, that dataset took just under 20 minutes—or around 60% faster than `ndmccopy`:

```
# xcp copy 10.193.67.219:/flexvol/nfs 10.193.67.219:/flexgroup_16/xcp Sending statistics...
5.49M scanned, 5.49M copied, 5.49M indexed, 5.60 GiB in (4.81 MiB/s), 4.55 GiB out (3.91 MiB/s),
19m52s.
```

Note

This XCP copy was done on a VM with a 1GB network and not much RAM or CPU; more robust servers will perform even better.

FlexVol to FlexGroup conversion

In ONTAP 9.7 and later, you can convert a single FlexVol to a FlexGroup volume containing a single member volume, in place, with less than 40 seconds disruption. This is regardless of how much data capacity or number of files reside in the volume. There is no need to remount clients, copy data, or make any other modifications that could create a maintenance window. After the FlexVol volume is converted to a FlexGroup volume, you can add new member volumes to the converted FlexGroup volume to expand the capacity.

Why convert a FlexVol volume to a FlexGroup volume?

FlexGroup volumes offer a few advantages over FlexVol volumes, such as:

- Ability to expand beyond 100TB and two billion files in a single volume
- Ability to scale out capacity or performance nondisruptively
- Multi-threaded performance for high-ingest workloads
- Simplification of volume management and deployment

For example, perhaps you have a workload that is growing rapidly and you don't want to have to migrate the data, but you still want to provide more capacity. Or perhaps a workload's performance isn't good enough on a FlexVol volume, so you want to provide better performance handling with a FlexGroup volume. Converting can help here.

When not to convert a FlexVol volume

Converting a FlexVol volume to a FlexGroup volume might not always be the best option. If you require FlexVol features that are not available in FlexGroup volumes, then you should hold off. For example, SVM-DR and cascading SnapMirror relationships are not supported in ONTAP 9.7, so if you need them, you should stay with FlexVol volumes.

Also, if you have a FlexVol volume that is already very large (80–100TB) and already very full (80–90%), you might want to copy the data rather than convert, because the converted FlexGroup volume would have a very large, very full member volume. This could create performance issues and doesn't fully resolve your capacity issues, particularly if that dataset contains files that grow over time.

Figure 82 Converting a FlexVol volume that is nearly full and at maximum capacity



If you were to convert this 90% full volume to a FlexGroup volume, you would have a 90% full member volume. If you add new member volumes, they would be 100TB each and 0% full, so they would take on a majority of new workloads. The data would not rebalance and if the original files grew over time, you could still run out of space with nowhere to go (because 100TB is the maximum member volume size).

Things that can block a conversion

ONTAP blocks conversion of a FlexVol for the following reasons:

- The ONTAP version isn't 9.7 or later on all nodes.
- ONTAP upgrade issues are preventing conversion.
- A FlexVol volume was transitioned from 7-Mode using 7MTT (ONTAP 9.7).
 - Transitioned volumes can be converted as of ONTAP 9.8.
- Something is enabled on the volume that is not supported with FlexGroup yet (SAN LUNs, Windows NFS, SMB1, part of a fan-out/cascade SnapMirror, SVM-DR, Snapshot naming/autodelete, `vmalign` set, SnapLock, space SLO, logical space enforcement/reporting, and so on)
- FlexClone volumes are present, and the FlexVol is the parent volume (the volume being converted can't be a parent or a clone).
- The volume is a FlexCache origin volume.
- Snapshot copies with Snap IDs greater than 255 (ONTAP 9.7).
 - ONTAP 9.8 adds support for 1023 snapshots, so this limit does not apply in that release.
- Storage efficiencies are enabled (can be reenabled after).
- The volume is a source of a SnapMirror relationship, and the destination has not been converted yet.
- The volume is part of an active (not quiesced) SnapMirror relationship.
- Quotas are enabled (they must be disabled first, then reenabled after).
- Volume names are longer than 197 characters.
- The volume is associated with an application.
 - ONTAP 9.7 only; ONTAP 9.8 removes this limitation.
- ONTAP processes are running (mirrors, jobs, waffiron, NDMP backup, inode conversion in process, and so on).
- Storage Virtual Machine (SVM) root volume.
- Volume is too full.

You can check for upgrade issues with the following commands:

```
cluster::*> upgrade-revert show
cluster::*> system node image show-update-progress -node *
```

You can check for transitioned volumes with the following commands:

```
cluster::*> volume show -is-transitioned true
There are no entries matching your query.
```

You can check for Snapshot copies with Snap IDs greater than 255 with the following command:

```
cluster::*> volume snapshot show -vserver DEMO -volume testvol -logical-snap-id >255 -fields
logical-snap-id
```

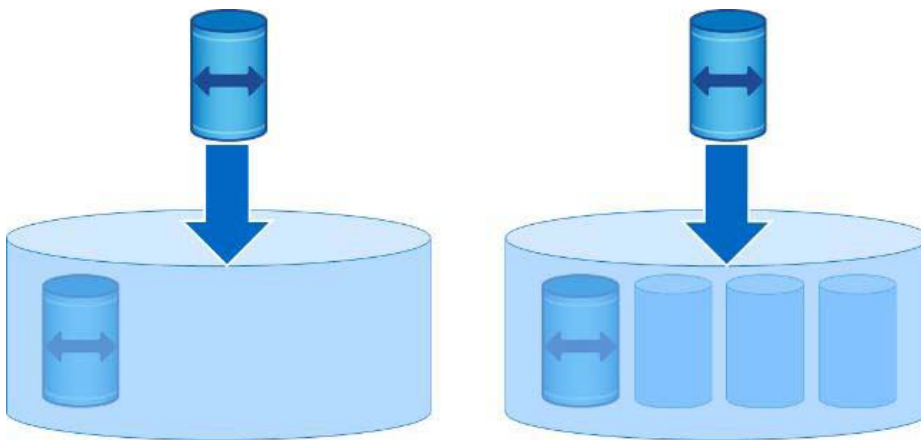
How it works

To convert a FlexVol volume to a FlexGroup volume in ONTAP 9.7 and later, you run a single, simple command at the advanced privilege level:

```
cluster::*> volume conversion start ?  
-vserver <vserver name> *Vserver Name  
[-volume] <volume name> *Volume Name  
[ -check-only [true] ] *Validate the Conversion Only  
[ -foreground [true] ] *Foreground Process (default: true)
```

When you run this command, ONTAP converts a single FlexVol volume into a FlexGroup volume with one member. You can even run a validation of the conversion before you do the real thing.

Figure 83 Converting a FlexVol volume to a FlexGroup and adding member volumes



The process is 1:1, so you cannot currently convert multiple FlexVol volumes into a single FlexGroup volume. When the conversion is done, you have a single-member FlexGroup volume to which you can then add more member volumes of the same size to increase capacity and performance.

Other considerations and caveats

Although the actual conversion process is simple, there are some things to consider before converting. Most of these considerations will go away with future ONTAP releases as support is added for features, but it is still prudent to identify them here.

After the initial conversion is performed, ONTAP unmounts the volume internally and remounts it to get the new FlexGroup information into the appropriate places. Clients do not have to remount or reconnect, but they will see a disruption that last less than 1 minute while this takes place. Refer to ["Sample FlexVol to FlexGroup conversion"](#) for an example. Data does not change at all; file handles all stay the same.

- FabricPool does not need anything. It just works. No need to rehydrate data on premises.
- Snapshot copies remain available for clients to access data from, but you are not able to use them to restore the volume through `snaprestore` commands. Those Snapshot copies are marked as pre-conversion.
- SnapMirror relationships pick up where they left off without rebaselining, provided the source and destination volumes have both been converted. But there are no SnapMirror restores of the volume—just file retrieval from clients. SnapMirror destinations need to be converted first.
- FlexClone volumes need to be deleted or split from the volume to be converted.
- Storage efficiencies need to be disabled during the conversion, but your space savings are preserved after the conversion.
- FlexCache instances with an origin volume being converted must be deleted.

- Space guarantees can affect how large a FlexGroup volume can become if they are volume guarantees. New member volumes must be the same size as the existing members, so you need adequate space to honor them.
- Quotas are supported in FlexGroup volumes but are done a bit differently than in FlexVol volumes. So, while the conversion is being performed, quotas must be disabled (`quota off`) and then reenabled later (`quota on`).

Conversion to FlexGroup volumes is a one-way street after you expand it, so be sure you're ready to make the jump. If anything goes wrong during the conversion process, there is a rescue method that Fujitsu Support can help you use so that your data is safe even if you run into an issue.

When you expand the FlexGroup volume to add new member volumes, they are the same size as the converted member volume, so be sure there is adequate space available. Additionally, the existing data that resides in the original volume remains in that member volume. Data does not redistribute. Instead, the FlexGroup volume favors newly added member volumes for new files.

Are you nervous about converting?

If you do not feel comfortable about converting your production FlexVol volume to a FlexGroup volume right away, you have options.

First, ONTAP allows you to run a check on the conversion command with `-check-only true` that tells you what prerequisites you might be missing.

For example:

```
cluster::*> volume conversion start -vserver DEMO -volume flexvol -foreground true -check-only true
Error: command failed: Cannot convert volume "flexvol" in Vserver "DEMO" to a FlexGroup.
Correct the following issues and retry the command:
* The volume has Snapshot copies with IDs greater than 255. Use the (privilege: advanced)
"volume snapshot show -vserver DEMO -volume flexvol -logical-snap-id >255 -fields logical-snap-id"
command to list the Snapshot copies with IDs greater than 255 then delete them using the
"snapshot delete -vserver DEMO -volume flexvol" command.
* Quotas are enabled. Use the 'volume quota off -vserver DEMO -volume flexvol' command to
disable quotas.
* Cannot convert because the source "flexvol" of a SnapMirror relationship is source to more
than one SnapMirror relationship. Delete other Snapmirror relationships, and then try the
conversion of the source "flexvol" volume.
* Only volumes with logical space reporting disabled can be converted. Use the 'volume modify -
vserver DEMO -volume flexvol -is-space-reporting-logical false' command to disable logical space
reporting.
```

For an example of a FlexVol to FlexGroup conversion, refer to ["Sample FlexVol to FlexGroup conversion"](#).

Creating a conversion sandbox – migrating data

ONTAP can create multiple SVMs, which can be fenced off from network access. You can use this approach to test things such as volume conversion. The only trick is getting a copy of that data over—but it is really not that tricky.

■ Option 1: SnapMirror

You can use SnapMirror to replicate your to-be-converted volume to the same SVM or a new SVM. Then, break the mirror and delete the relationship. Now you have a sandbox copy of your volume, complete with Snapshot copies, to test out conversion, expansion, and performance.

■ Option 2: FlexClone and volume rehost

If you do not have SnapMirror or you want to try a method that is less taxing on your network, you can use a combination of FlexClone (an instant copy of your volume backed by a Snapshot copy) and `volume rehost` (an instant move of the volume from one SVM to another). Keep in mind that FlexClone copies cannot be rehosted, but you can split the clone and then rehost.

Essentially, the process is as follows:

Procedure ▶▶▶ —————

- 1 Use `flexclone create`.
- 2 Use `FLEXCLONE SPLIT`.
- 3 Issue `volume rehost` to the new SVM (or convert on the existing SVM).



Note

Alternately, you can create FlexClone volumes from a source SVM to a destination SVM and then split the FlexClone as covered in "[FlexClone to different storage virtual machine \(SVM\)](#)".

Converting a FlexVol volume in a SnapMirror relationship

You can also convert FlexVol volumes that are part of existing SnapMirror relationships without disruption.

The basic steps are:

Procedure ▶▶▶ —————

- 1 Break the SnapMirror.
- 2 Convert the SnapMirror destination FlexVol volume to a FlexGroup volume.
- 3 Convert the source SnapMirror FlexVol volume to a FlexGroup volume.
- 4 Resync the SnapMirror.



If you expand the newly converted FlexGroup volume to add more member volumes, ONTAP automatically expands the destination volume without needing to rebaseline the SnapMirror.

For an example of this process, refer to "[Converting FlexVols in existing SnapMirror relationships – example](#)".

Does a high file count affect the conversion process?

Short answer: No!

In the sample conversion shown in "[Sample FlexVol to FlexGroup conversion](#)", a volume with 300,000 files was converted. However, 300,000 files in a volume is not a true high file count. For an example of converting a FlexVol volume with 500 million files, refer to "[Converting FlexVols in existing SnapMirror relationships – example](#)".

Note

For a video example, see [Statistics show-periodic during FlexVol - FlexGroup convert](#).

Migrating from third-party storage to FlexGroup

When migrating from non-Fujitsu storage (SAN or NAS), the migration path is a file-based copy. Various methods are available to perform this migration; some are free and some are paid through third-party vendors.

For NFSv3-only data, Fujitsu strongly recommends the [XCP Migration Tool](#). XCP is a free, license-based tool that can vastly improve the speed of data migration of high-file-count environments. XCP also offers robust reporting capabilities. XCP 1.5 and later versions also offer NFSv4.x and NFSv4.x ACL support, as well as being officially supported by NetApp.

Note

XCP is supported only for migration to a Fujitsu storage system.

For CIFS/SMB data, XCP for SMB is available. Robocopy is a free tool, but the speed of transfer depends on using its [multithreaded capabilities](#). Third-party providers can also perform this type of data transfer.

Migrating from Data ONTAP operating in 7-Mode

Migrate data from Data ONTAP operating in 7-Mode to FlexGroup in one of two ways:

- Full migration of 7-Mode systems to ONTAP systems by using the copy-based or copy-free transition methodology. When using copy-free transition, the process is followed by copy-based migration of data in FlexVol volumes to FlexGroup volumes.
- Copy-based transition from a FlexVol or host-based copy from a LUN by using the previously mentioned tools for migrating from non-Fujitsu storage to FlexGroup.

If you wish to migrate to FlexGroup volumes from FlexVol, you can use [FlexVol to FlexGroup conversion](#). In ONTAP 9.8 and later, this works on volumes transitioned from 7-Mode systems.

Migrating from SAN LUNs in ONTAP

When migrating from existing ONTAP objects such as SAN-based LUNs, the current migration path is copy-based. The previously mentioned tools for migrating from non-Fujitsu storage to FlexGroup can also be used for migrating from ONTAP objects.

XCP Migration Tool

The XCP Migration Tool is free and was designed specifically for scoping, migration, and management of large sets of unstructured NAS data. The initial version was NFSv3 only, but a CIFS version is now available. To use the tool, download it and request a free license (for software tracking purposes only).

XCP addresses the challenges that high-file-count environments have with metadata operation and data migration performance by using a multicore, multichannel I/O streaming engine that can process many requests in parallel.

These requests include the following:

- Data migration
- File or directory listings (a high-performance, flexible alternative to `ls`)
- Space reporting (a high-performance, flexible alternative to `du`)

XCP has sometimes reduced the length of data migration by 20 to 30 times for high-file-count environments. In addition, XCP has reduced the file list time for 165 million files from 9 days on a competitor's system to 30 minutes on NetApp technology—a performance improvement of 400 times. As of XCP 1.5, the tool is officially supported by NetApp support.

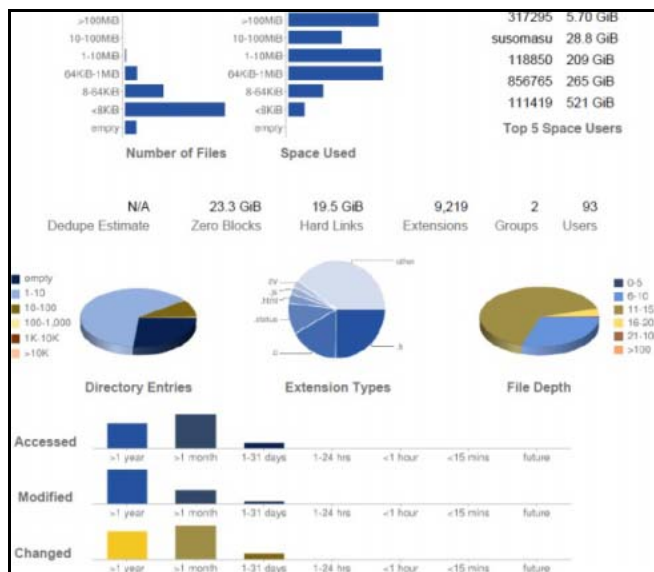
XCP 1.6 also adds File Systems Analytics functionality. This is similar to the [functionality added to ONTAP 9.8](#), but is able to scan systems that are not running ONTAP as well.

Note

For best results, use the latest XCP release available.

XCP also gives some handy reporting graphs, as shown in [Figure 84](#).

Figure 84 XCP reporting graphs



For more information, see the official XCP website at <http://xcp.netapp.com>.

Using XCP to scan files before migration

When deploying a FlexGroup volume, evaluate the file system and structure to help you determine initial sizing considerations and the best way to lay out member volumes. In high-file-count environments, this can be time consuming and tedious. XCP allows you to scan files and export to the CSV or XML format to easily review your file system.

The following example shows a FlexGroup volume with over a million files. Ideally, we don't want to spend much time analyzing these files.

```
Cluster::> vol show -vserver DEMO -fields files,files-used -volume flexgroup_16
vserver volume      files      files-used
-----
DEMO      flexgroup_16 318766960 1103355
```

To streamline this process, you can use xcp scan to get file information. Here's a sample command:

```
C:\> xcp scan -stats \\demo\flexgroup > C:\destination.csv
```

When you do this, the client scans the files and adds information to a comma-separated values (CSV) document. This document shows information such as the following:

- Maximum and average values for size, depth of directory, and dirsiz

```
== Maximum Values ==
Size Depth Namelen Dirsize
340MiB 9 86 500

== Average Values ==
Size Depth Namelen Dirsize
1.61KiB 4 6 11
```

- Top file extensions

```
== Top File Extensions ==
.docx .png .pptx .pdf .css other
1000038 260 175 128 91 33 219
```

- Number of files, broken down by size ranges

```
== Number of files ==
empty <8KiB 8-64KiB 64KiB-1MiB 1-10MiB 10-100MiB >100MiB
8 1000215 156 288 265 10 2
```

- Space used by size range

```
== Space used ==
empty <8KiB 8-64KiB 64KiB-1MiB 1-10MiB 10-100MiB >100MiB
0 28.7MiB 3.94MiB 124MiB 695MiB 272MiB 453MiB
```

- Directory entries, broken down by file counts

```
== Directory entries ==
empty 1-10 10-100 100-1K 1K-10K >10k
7 100118 30 200
```

- Directory depth ranges

```
== Depth ==
0-5 6-10 11-15 16-20 21-100 >100
1100966 333
```

- Modified and created date ranges

```
== Modified ==
>1 year >1 month 1-31 days 1-24 hrs <1 hour <15 mins future
579 1100559 11 150

== Created ==
>1 year >1 month 1-31 days 1-24 hrs <1 hour <15 mins future
1100210 1089
```

16. Migrating to ONTAP FlexGroup
Using XCP to scan files before migration

- A summary of the file structure, including total file count, total directories, symlinks, junctions, and total space used

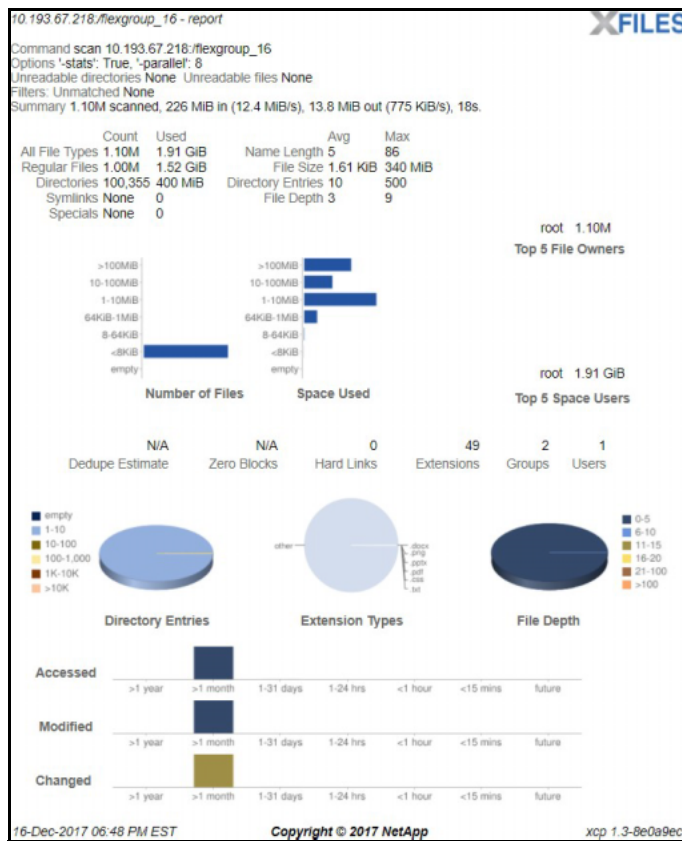
```
Total count: 1101299
Directories: 100355
Regular files: 1000944
Symbolic links:
Junctions:
Special files:
Total space for regular files: 1.54GiB
Total space for directories: 0
Total space used: 1.54GiB
1,101,299 scanned, 0 errors, 26m34s
```

You can also use XCP over NFS to scan CIFS volumes and get more robust reporting and the ability to export to HTML, which presents the data in graphical format.

For example, the following command creates the report shown in [Figure 85](#):

```
xcp scan -stats -html demo:/flexgroup_16 > /flexgroup.html
```

Figure 85 XCP report



Using XCP to scan file systems provides average file size information, largest file size, capacity, and file count measurements for the top five file owners, and much more. These statistics are available only in the NFS version of XCP, but you can still run NFS scans on datasets that only do SMB traffic by setting up a virtual machine that can use NFS.

Using XCP to run disk usage (du) scans

One common complaint is that, in high-file-count environments, running commands like `du` can take an exceedingly long time. For example, this `du` command ran on a FlexGroup volume with 1,101,002 files and folders and took 21 minutes and 22.600 seconds.

With XCP, this command scanned the same dataset in 22.852 seconds with the same client:

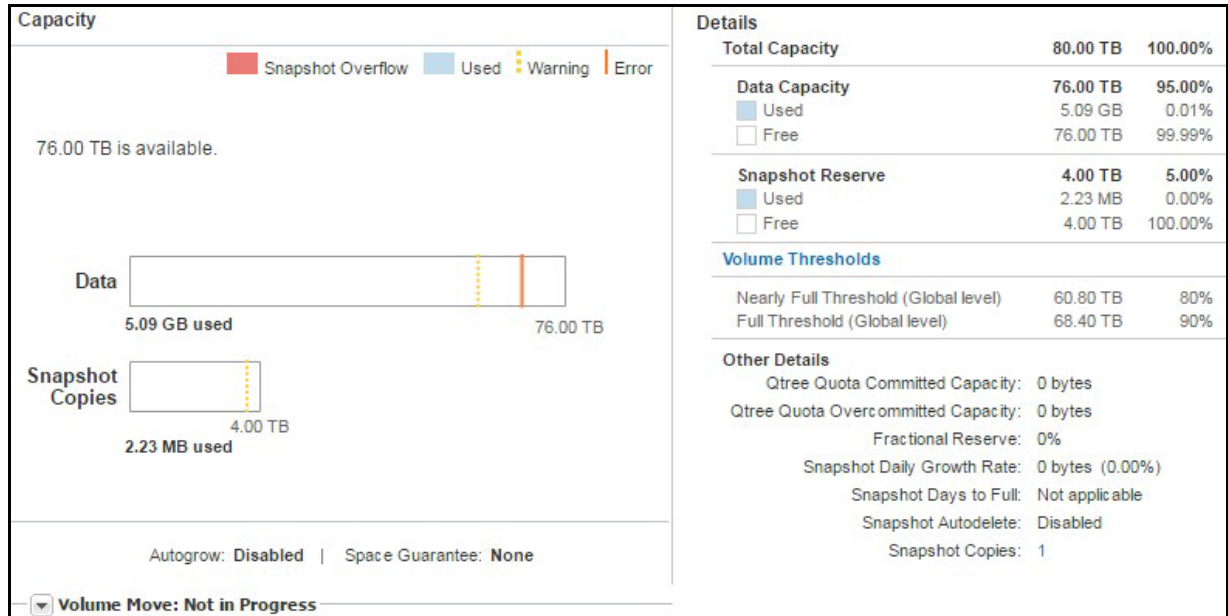
```
[root@centos7 ~]# xcp -duk DEMO:/FGlocal 2>/dev/null | egrep -v '.*?/.*?/'
```

17. Examples

Thin provisioning example

The following image shows that the FlexGroup volume has a total capacity of 80TB and 5GB used. Also, 4TB have been reserved for Snapshot copies (5%). The available space is 76TB.

Figure 86 FlexGroup capacity breakdown–Active IQ Unified Manager



However, in the following CLI output, a few anomalies stand out:

- In Active IQ Unified Manager, the FlexGroup volume shows as having 76TB available, but, in the CLI, only 11.64TB is available.
- The FlexGroup volume shows as having 11.64TB available, but the member FlexVol volumes all show roughly 5.8TB available.
- The percentage used for the FlexGroup volume shows as 85%, even though we have used only 5GB, which is a negligible amount of space compared with 80TB (5GB of 81920GB is less than 1%).
- The FlexGroup volume shows as 85% used, but the member FlexVol volumes all show as 41% used, despite each having a different amount of space per FlexVol member.

Example:

```
cluster::> volume show -is-constituent true -fields size,used,percent-used,available
-vserver SVM -volume
flexgroup* -sort-by volume
vserver volume size available used percent-used
-----
SVM flexgroup 80TB 11.64TB 5.08GB 85%
SVM flexgroup__0001 10TB 5.81TB 147.5MB 41%
SVM flexgroup__0002 10TB 5.83TB 145.2MB 41%
SVM flexgroup__0003 10TB 5.81TB 144.9MB 41%
SVM flexgroup__0004 10TB 5.83TB 148.0MB 41%
SVM flexgroup__0005 10TB 5.81TB 4.08GB 41%
SVM flexgroup__0006 10TB 5.83TB 147.6MB 41%
SVM flexgroup__0007 10TB 5.81TB 145.3MB 41%
SVM flexgroup__0008 10TB 5.83TB 146.5MB 41%
9 entries were displayed.
```

The anomalies are due to ONTAP calculating against the aggregate's available space. The FlexVol member volumes show equivalent available values depending on the aggregates where they are located.

```
cluster::> volume show -is-constituent true -fields available,aggregate -vserver SVM
-volume
flexgroup* -sort-by aggregate
vserver volume                aggregate    available
-----
SVM    flexgroup__0001            aggr1_node1 5.81TB
SVM    flexgroup__0003            aggr1_node1 5.81TB
SVM    flexgroup__0005            aggr1_node1 5.81TB
SVM    flexgroup__0007            aggr1_node1 5.81TB
SVM    flexgroup__0002            aggr1_node2 5.83TB
SVM    flexgroup__0004            aggr1_node2 5.83TB
SVM    flexgroup__0006            aggr1_node2 5.83TB
SVM    flexgroup__0008            aggr1_node2 5.83TB

cluster::> storage aggregate show -aggregate aggr1* -fields availsize
aggregate    availsize
-----
aggr1_node1 5.81TB
aggr1_node2 5.83TB
2 entries were displayed.
```

Using thin provisioning means that you must consider the aggregate capacity and the volume footprint when monitoring space.

Volume Autosize example

In the following example, we'll attempt to show how volume autosize works when a member volume reaches a capacity threshold.

In this case, the member volumes are all 1GB in size (not recommended). This was done to make filling the volume with a single file easier.

```
cluster::*> vol show -vserver DEMO -volume fgautogrow* -sort-by used -fields
available
vserver volume                size
-----
DEMO    fgautogrow__0001            1GB
DEMO    fgautogrow__0002            1GB
DEMO    fgautogrow__0003            1GB
DEMO    fgautogrow__0004            1GB
DEMO    fgautogrow__0005            1GB
DEMO    fgautogrow__0006            1GB
DEMO    fgautogrow__0007            1GB
DEMO    fgautogrow__0008            1GB
```

Note

1GB is not a recommended size for member volumes; the minimum member volume size should be no less than 100GB. ONTAP programmatically prevents creation of FlexGroup volumes that have member volumes smaller than 100GB with REST APIs and in ONTAP System Manager and warns you in the CLI.

With volume autosize, the write succeeds because the member volume in which the write lands grows to the appropriate size to honor the write. In this case, the file was written to member volume fgautogrow 0003.

```
cluster::*> vol show -vserver DEMO -volume fgautogrow* -sort-by used -fields
available,size
vserver volume          size  available used
-----
DEMO    fgautogrow__0004 1GB   915.6MB  57.23MB
DEMO    fgautogrow__0005 1GB   915.6MB  57.23MB
DEMO    fgautogrow__0006 1GB   915.6MB  57.23MB
DEMO    fgautogrow__0007 1GB   915.6MB  57.23MB
DEMO    fgautogrow__0008 1GB   915.6MB  57.23MB
DEMO    fgautogrow__0002 1GB   915.5MB  57.26MB
DEMO    fgautogrow__0001 1GB   915.5MB  57.27MB
DEMO    fgautogrow__0003 1.60GB 498.7MB 1.03GB
```

When this happens, an event is triggered in the event management system and can be seen with event log show.

```
INFORMATIONAL wafl.vol.autoSize.done: Volume Autosize: Automatic grow of volume
'fgautogrow__0003@vserver:7e3cc08e-d9b3-11e6-85e2-00a0986b1210' by 611MB complete.
```

This event can also be monitored with SNMP, by sending alerts through event destinations, or with Active IQ Unified Manager.

```
cluster:::> event route show -message-name wafl.vol.autoSize.done -instance

                                Message Name: wafl.vol.autoSize.done
                                Severity: INFORMATIONAL
                                Corrective Action: (NONE)
                                Description: This message occurs on successful autosize of volume.
                                Supports SNMP trap: true
                                Destinations: -
                                Number of Drops Between Transmissions: 0
                                Dropping Interval (Seconds) Between Transmissions: 0
```

Snapshot spill example

For example, if snap reserve is set to 5% on a 400GB FlexGroup volume, then that is a total of 20GB of snapshot reserve. If there are four member volumes, then the snapshot reserve per member volume is ~5GB.

```
cluster::*> vol show -vserver DEMO -volume FG_SM_400G* -fields size,used,size-used-
by-snapshots,snapshot-reserve-available
vserver volume          size  used  size-used-by-snapshots snapshot-reserve-available
-----
DEMO    FG_SM_400G 420.9GB 2.01GB 3.16MB
DEMO    FG_SM_400G__0001
          105.2GB 513.7MB
          860KB
          5.26GB
DEMO    FG_SM_400G__0002
          105.2GB 513.8MB
          432KB
          5.26GB
DEMO    FG_SM_400G__0003
          105.2GB 513.8MB
          828KB
          5.26GB
DEMO    FG_SM_400G__0004
          105.2GB 513.7MB
          1.09MB
          5.26GB
```

17. Examples
Snapshot spill example

If I write a 4GB file to that volume, nothing gets used in the snapshot. That's because no blocks have been overwritten:

```
cluster::*> vol show -vserver DEMO -volume FG_SM_400G* -fields size,used,size-used-by-snapshots,snapshot-reserve-available
```

| vserver | volume | size | used | size-used-by-snapshots | snapshot-reserve-available |
|---------|------------------|---------|---------|------------------------|----------------------------|
| DEMO | FG_SM_400G | 420.9GB | 6.70GB | 3.17MB | 21.04GB |
| DEMO | FG_SM_400G__0001 | 105.2GB | 5.19GB | 868KB | 5.26GB |
| DEMO | FG_SM_400G__0002 | 105.2GB | 513.9MB | 432KB | 5.26GB |
| DEMO | FG_SM_400G__0003 | 105.2GB | 513.9MB | 828KB | 5.26GB |
| DEMO | FG_SM_400G__0004 | 105.2GB | 513.8MB | 1.09MB | 5.26GB |

If I take a snapshot now, the existing blocks are locked into place in case an overwrite occurs later. But again, no space is used by the snapshot yet.

```
cluster::*> snapshot create -vserver DEMO -volume FG_SM_400G -snapshot file1
```

```
cluster::*> vol show -vserver DEMO -volume FG_SM_400G* -fields size,used,size-used-by-snapshots,snapshot-reserve-available
```

| vserver | volume | size | used | size-used-by-snapshots | snapshot-reserve-available |
|---------|------------------|---------|---------|------------------------|----------------------------|
| DEMO | FG_SM_400G | 420.9GB | 6.69GB | 3.78MB | 21.04GB |
| DEMO | FG_SM_400G__0001 | 105.2GB | 5.18GB | 1.02MB | 5.26GB |
| DEMO | FG_SM_400G__0002 | 105.2GB | 514.0MB | 580KB | 5.26GB |
| DEMO | FG_SM_400G__0003 | 105.2GB | 514.0MB | 976KB | 5.26GB |
| DEMO | FG_SM_400G__0004 | 105.2GB | 513.9MB | 1.25MB | 5.26GB |

Space is used if we overwrite data. This can happen on a delete or if I copy over the file, or I were simply to change the file data. But notice that the space change only happens on the member volume where the file lives. This is reflected by the FlexGroup itself, but it doesn't incorporate the actual snapshot used space, because it's below the reserved capacity. In the example below, deleting the file uses a full 4.4GB of space in the snapshot and reduces the space used in the volume by the same amount:

```
cluster::*> vol show -vserver DEMO -volume FG_SM_400G* -fields size,used,size-used-by-snapshots,snapshot-reserve-available
```

| vserver | volume | size | used | size-used-by-snapshots | snapshot-reserve-available |
|---------|------------------|---------|---------|------------------------|----------------------------|
| DEMO | FG_SM_400G | 420.9GB | 772.7MB | 4.41GB | 16.63GB |
| DEMO | FG_SM_400G__0001 | 105.2GB | 596.7MB | 4.41GB | 874.4MB |
| DEMO | FG_SM_400G__0002 | 105.2GB | 58.65MB | 624KB | 5.26GB |
| DEMO | FG_SM_400G__0003 | 105.2GB | 58.69MB | 1012KB | 5.26GB |
| DEMO | FG_SM_400G__0004 | 105.2GB | 58.71MB | 1.29MB | 5.26GB |

17. Examples
Snapshot spill example

If we start to overrun the snapshot reserve available, used snapshot space will start to spill into the AFS. For example, if I reduce the snap reserve to 1%, we can see that happen with our existing volume.

```
cluster::*> vol show -vserver DEMO -volume FG_SM_400G* -fields size,used,size-used-by-snapshots,snapshot-reserve-available
-----
vserver volume          size    used    size-used-by-snapshots snapshot-reserve-available
-----
DEMO    FG_SM_400G    420.9GB 3.86GB 4.64GB                                3.15GB
DEMO    FG_SM_400G__0001
        105.2GB 3.69GB 4.64GB                                0B
DEMO    FG_SM_400G__0002
        105.2GB 58.72MB
        624KB                                1.05GB
DEMO    FG_SM_400G__0003
        105.2GB 58.84MB
        1012KB                               1.05GB
DEMO    FG_SM_400G__0004
        105.2GB 58.71MB
        1.29MB                               1.05GB
```

Now we see nearly all of the snapshot space reflected in used and we have 0 snapshot reserve available. This also affects the snap reserve used percentage we see in the df output:

```
cluster::*> df -g FG_SM_400G
Filesystem          total used avail capacity Mounted on      Vserver
/vol/FG_SM_400G/    416GB 3GB 412GB    0%    /FG_SM_400G    DEMO
/vol/FG_SM_400G/.snapshot 4GB    4GB    0GB    110% /FG_SM_400G/.snapshot DEMO
```

Now we are using 110% of our snapshot reserve. That space has to go somewhere, so it goes into the AFS and now we're using 3GB, when we were only using ~597MB before the snapshot reserve was adjusted. This is known as snapshot spill, and it can negatively affect capacity reporting and potentially cause a FlexGroup volume to run out of space, even if we are currently reporting free space available.

You can see the snapshot spill amount with the `volume show-space` command:

```
cluster:::> volume show-space -vserver DEMO -volume FG_SM_400G* -fields snapshot-spill
-----
vserver volume          snapshot-spill
-----
DEMO    FG_SM_400G__0001 3.58GB
DEMO    FG_SM_400G__0002 -
DEMO    FG_SM_400G__0003 -
DEMO    FG_SM_400G__0004 -
```


Capacity monitoring and alerting examples in Active IQ Unified Manager

Active IQ Unified Manager provides methods to monitor and alert on various storage system functionalities, including used and free capacities. On the main Health page, Active IQ OnCommand displays active warnings and errors about capacity.



Also, Active IQ Unified Manager has a more detailed view of capacity-related events.



When you click one of the events, a full report of the issue is shown.



17. Examples
Capacity monitoring and alerting examples in Active IQ Unified Manager

In this detailed view, you can also configure alerts. To do so, click the Add link next to Alert Settings.

Summary

Severity: Error
 State: New
 Impact Level: Risk
 Impact Area: Capacity
 Source: SVM1:/nfs

Source Annotations:
 Source Groups:
 Source Type: Volume

Acknowledged By:
 Resolved By:
 Assigned To:

Triggered Time: 14 Hours 53 Mins Ago
 Trigger Condition: The full threshold set at 90% is breached. 4.14 GB (99.07%) of 4.18 GB is used.

Alert Settings: Add

You can also view volume capacities from the Volume screen. When you click Storage > Volumes and select a volume, a screen like the following appears:

Volume: nfs (Online) ★ Actions View Volumes ?

! Error - Volume Space Full (15 Hours 6 Mins Ago)
 Days to Full (current usage statistics): Not applicable | Daily Growth Rate: -3.72 %

Capacity Efficiency Configuration Protection

Capacity

Only 35.66 MB is available.

Legend: ■ Snapshot Overflow ■ Used ! Warning ! Error

Data
 305.85 MB used (Warning) / 4.18 GB total

Snapshot Copies
 225.28 MB used (Warning) / 3.85 GB of data space consumed

Autogrow: Disabled | Space Guarantee: None

Volume Move: Not in Progress

Details

| | | |
|---------------------------------------|--------------------|---------|
| Total Capacity | | |
| Total Capacity | 4.40 GB | 100.00% |
| Data Capacity | | |
| Used | 4.18 GB | 95.00% |
| Free | 305.85 MB | 7.15% |
| Free | 35.66 MB | 0.83% |
| Snapshot Overflow | 3.85 GB | 92.02% |
| Snapshot Reserve | | |
| Used | 225.28 MB | 5.00% |
| Free | 225.28 MB | 100.00% |
| Free | 0 bytes | 0.00% |
| Total Snapshot Used Capacity: 4.07 GB | | |
| Volume Thresholds | | |
| Nearly Full Threshold (Global level) | 3.34 GB | 80% |
| Full Threshold (Global level) | 3.76 GB | 90% |
| Other Details | | |
| Qtree Quota Committed Capacity | 0 bytes | |
| Qtree Quota Overcommitted Capacity | 0 bytes | |
| Fractional Reserve | 100% | |
| Snapshot Daily Growth Rate | 170.72 MB (75.78%) | |
| Snapshot Days to Full | 0 | |
| Snapshot Autodelete | Disabled | |
| Snapshot Copies | 11 | |

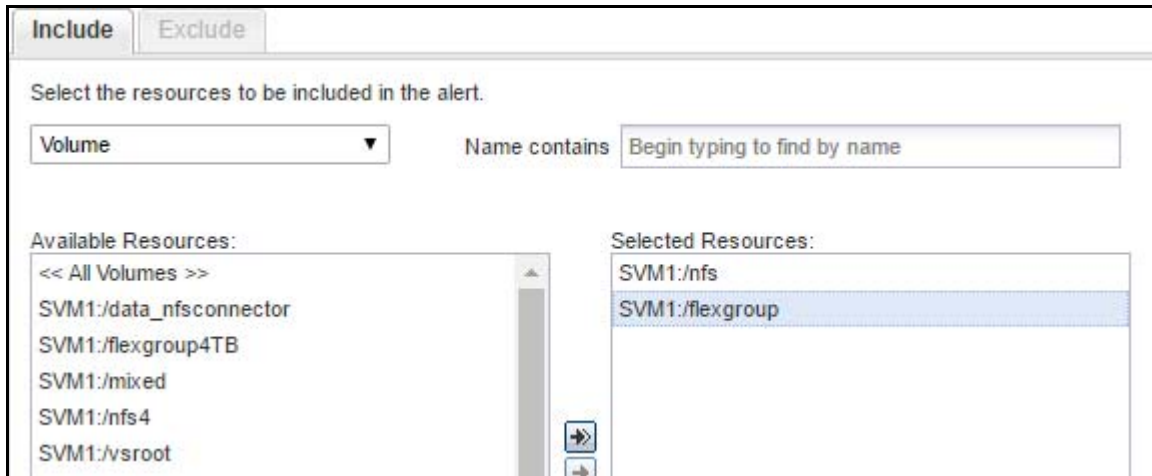
Click the Actions button to create alerts that are specific to the volume.

★ **Actions** View Volumes ?

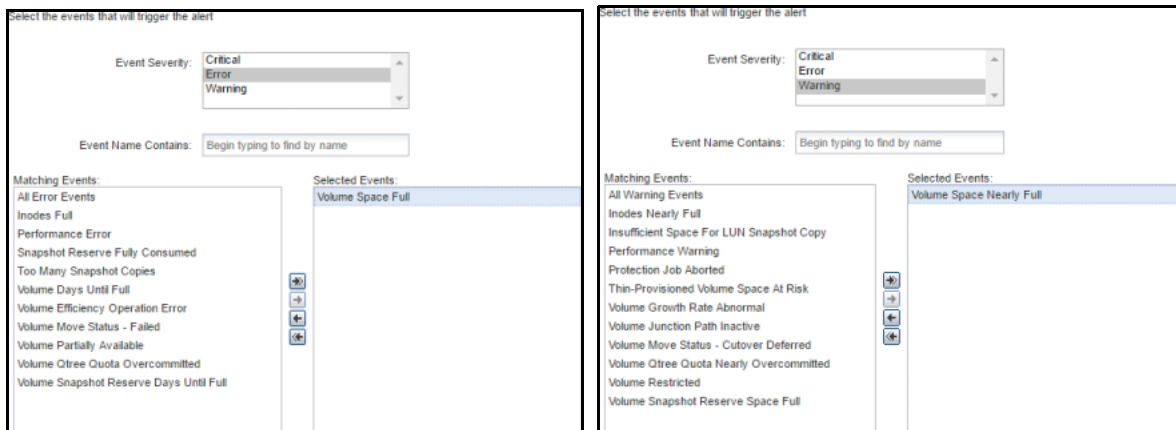
- + Add Alert
- ✎ Edit Thresholds
- ↶ Restore
- 🏷️ Annotate

17. Examples
Capacity monitoring and alerting examples in Active IQ Unified Manager

With the alert, you can add one or many volumes to various events (or exclude them).

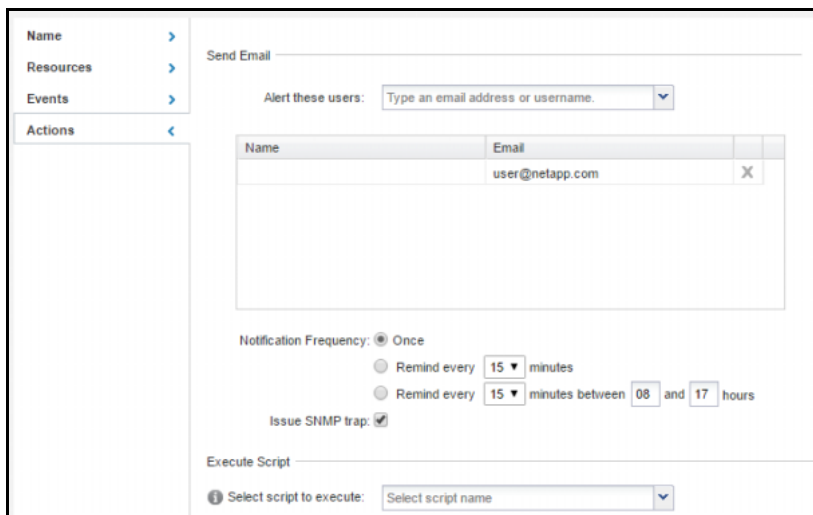


Events are organized by severity and include the Critical, Error, and Warning levels. Volume Space Full is included under the Error level, and Volume Space Nearly Full is under the Warning level.



When an event is triggered, the alert mechanism in Active IQ Unified Manager can do the following:

- Send an email to a user, a list of users, and a distribution list
- Trigger an SNMP trap
- Send reminders
- Execute scripts (such as an automated volume grow script)



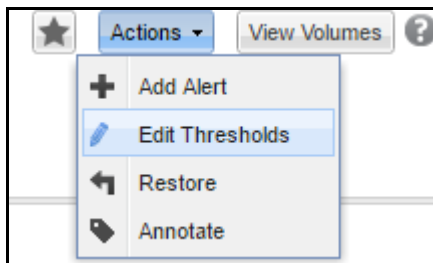
Editing volume thresholds in Active IQ Unified Manager

Thresholds for Volume Nearly Full and Volume Full control when an event management system event is triggered by the cluster. This control helps storage administrators stay on top of the volume capacities to prevent volumes from running out of space. In FlexGroup, this approach also involves remote allocation of files and folders, because ingest remoteness increases as a volume gets closer to full. As mentioned earlier, the Volume Nearly Full and Volume Full thresholds should be modified for a FlexGroup volume so that storage administrators are notified about potential capacity issues earlier than the defaults provide.

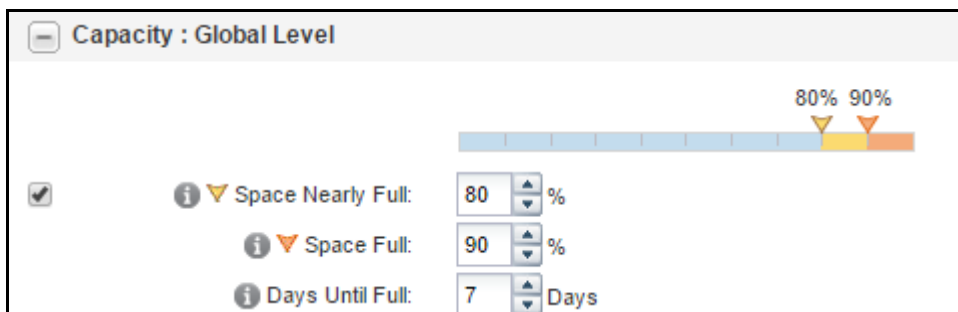
For more information, refer to [Best Practice 24: Volume Space Threshold Recommendations for FlexGroup](#).

The command line provides a method to modify the thresholds, as does ONTAP System Manager. Under the Actions button of the volume detail, select Edit Thresholds to modify the volume threshold on a per-volume basis. With a FlexGroup volume, the setting is applied to the whole FlexGroup volume, and thresholds are set on each member volume individually.

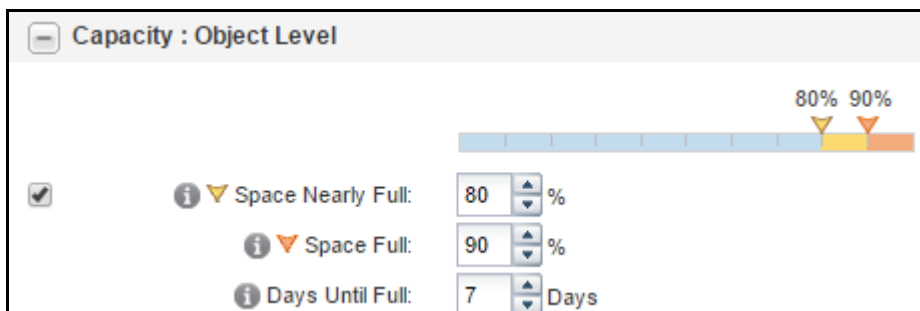
Figure 87 Editing volume thresholds



When you initially select the checkbox under Capacity: Global Level, the defaults are as shown. These defaults are unaffiliated with the ONTAP event management system volume thresholds. Rather, they are specific to Active IQ Unified Manager.



Changing the values modifies the threshold to be an Object Level.



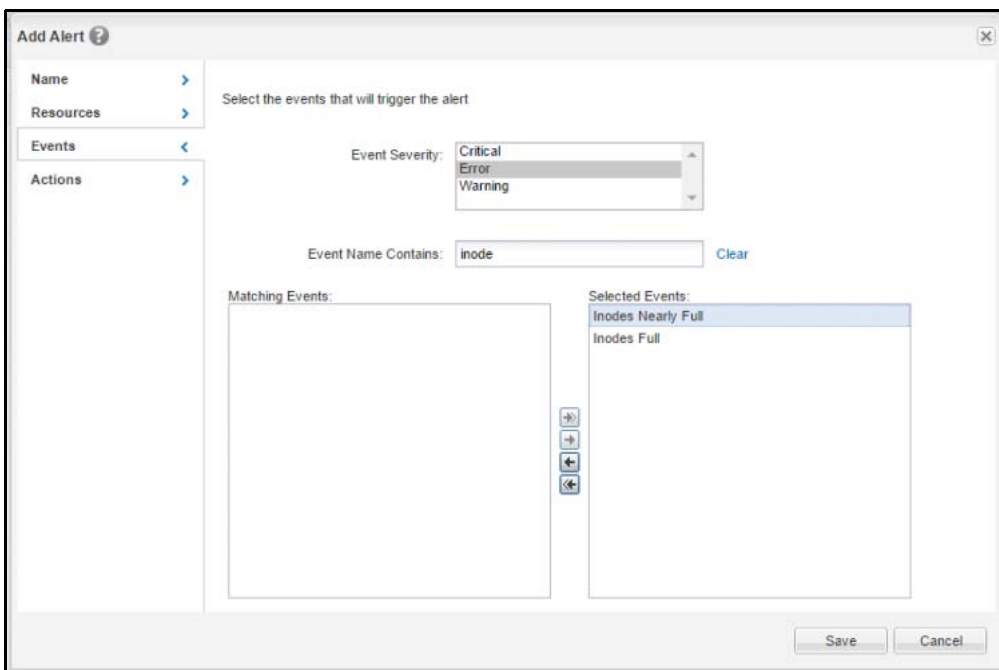
On the cluster, the volume-level threshold options are unchanged.

```
cluster::*> vol show -fields space-nearly-full-threshold-percent,space-full-
threshold-percent -sort-by space-nearly-full-threshold-percent -volume flexgroup
vserver volume      space-nearly-full-threshold-percent  space-full-threshold-percent
-----
SVM      flexgroup 95%                                           98%
```

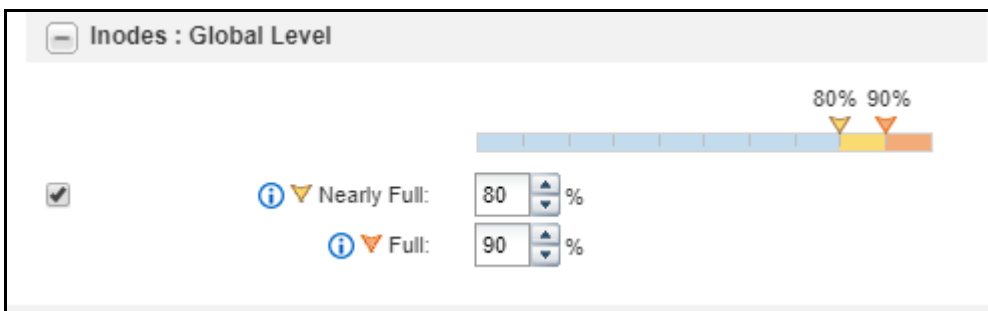
You can use Active IQ Unified Manager alerting along with the cluster's event management system alerting and event destination logic, or independently of this logic.

Inode monitoring

Active IQ Unified Manager also enables you to alert on inode count in FlexGroup volumes with the Inodes Nearly Full (Warning) and Inodes Full (Error) events. Alerts for inodes are configured similarly to the alerts for capacity.



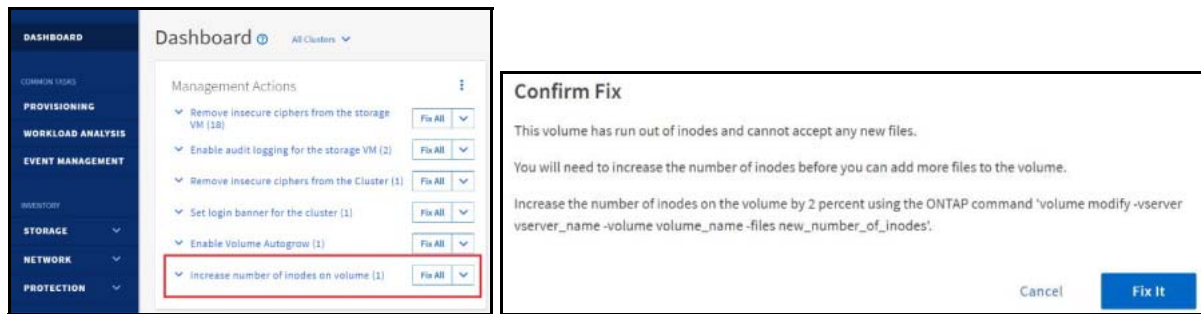
You can also edit inode thresholds from the Edit Thresholds window for more granular control over alerting.



Active IQ "Fix It"

Active IQ Unified Manager 9.8 and later introduces "fix it" functionality that allows storage admins to use a single click to resolve a set list of issues. One of those issues is when a volume reaches an inode threshold. These management actions are present on the dashboard when you log into Active IQ Unified Manager. In the example below, we have a volume that has exceeded the threshold for used inodes.

Figure 88 Active IQ Unified Manager – fix out of inodes



When you click Fix, Active IQ Unified Manager increases the total files by 2%.

Sample FlexVol to FlexGroup conversion

In this sample conversion, before we converted a volume, we added around 300,000 files to help determine how long the process might take with many files present.

```
cluster::*> df -i lotsafiles
Filesystem  iused  ifree  %iused Mounted on
/vol/lotsafiles/ 330197 20920929 1% /lotsafiles DEMO

cluster::*> volume show lotsa*
Vserver   Volume      Aggregate   State   Type Size      Available  Used%
-----
DEMO     lotsafiles  aggr1_nod1 online   RW      10TB    7.33TB    0%
```

First, let's try out the validation.

```
cluster::*> volume conversion start -vserver DEMO -volume lotsafiles -foreground true -check-only true
Error: command failed: Cannot convert volume "lotsafiles" in Vserver "DEMO" to a FlexGroup.
Correct the following issues and retry the command:
* SMB1 is enabled on Vserver "DEMO". Use the 'vserver cifs options modify -smb1-enabled false -vserver DEMO' command to disable SMB1.
* The volume contains LUNs. Use the "lun delete -vserver DEMO -volume lotsafiles -lun *" command to remove the LUNs, or use the "lun move start" command to relocate the LUNs to other FlexVols.
* NFSv3 MS-DOS client support is enabled on Vserver "DEMO". Use the "vserver nfs modify -vserver DEMO -v3-ms-dos-client disabled" command to disable NFSv3 MS-DOS client support on the Vserver.
Note that disabling this support will disable access for all NFSv3 MS-DOS clients connected to Vserver "DEMO".
```

17. Examples
Sample FlexVol to FlexGroup conversion

As you can see, there are some blockers, such as SMB1 and the LUN we created (to intentionally break conversion). So, we clear them with the recommendations and run the validation again. We see some caveats:

```
cluster::*> volume conversion start -vserver DEMO -volume lotsafiles -foreground true -check-only true
Conversion of volume "lotsafiles" in Vserver "DEMO" to a FlexGroup can proceed with the following warnings:
* After the volume is converted to a FlexGroup, it will not be possible to change it back to a flexible volume.
* Converting flexible volume "lotsafiles" in Vserver "DEMO" to a FlexGroup will cause the state of all Snapshot copies from the volume to be set to "pre-conversion". Pre-conversion Snapshot copies cannot be restored.
```

Now, let's convert. First, we start a script that takes a while to complete, while also using Active IQ Performance Manager to monitor performance during the conversion.

The conversion of the volume takes less than 1 minute, and the only disruption is a slight drop in IOPS.

```
cluster::*> volume conversion start -vserver DEMO -volume lotsafiles -foreground true

Warning: After the volume is converted to a FlexGroup, it will not be possible to change it back to a flexible volume.
Do you want to continue? {y|n}: y
Warning: Converting flexible volume "lotsafiles" in Vserver "DEMO" to a FlexGroup will cause the state of all Snapshot copies from the volume to be set to "pre-conversion". Pre-conversion Snapshot copies cannot be restored.
Do you want to continue? {y|n}: y [Job 23671] Job succeeded: success cluster::*> statistics show-periodic
cpu cpu total fcache total data data cluster cluster cluster disk disk pkts pkts avg busy ops nfs-ops cifs-ops ops spin-ops recv sent busy recv sent busy recv sent read write recv sent
-----
34% 44% 14978 14968 10 0 14978 14.7MB 15.4MB 0% 3.21MB 3.84MB 0% 11.5MB 11.6MB 4.43MB 1.50MB
49208 55026
40% 45% 14929 14929 0 0 14929 15.2MB 15.7MB 0% 3.21MB 3.84MB 0%
12.0MB 11.9MB?3.93MB 641KB 49983
55712
36% 44% 15020 15020 0 0 15019 14.8MB 15.4MB 0% 3.24MB 3.87MB 0%
11.5MB 11.5MB?3.91MB 23.9KB 49838
55806
30% 39% 15704 15694 10 0 15704 15.0MB 15.7MB 0% 3.29MB 3.95MB 0% 11.8MB 11.8MB 2.12MB 4.99MB
50936 57112
32% 43% 14352 14352 0 0 14352 14.7MB 15.3MB 0% 3.33MB 3.97MB 0% 11.3MB 11.3MB 4.19MB 27.3MB 49736
55707
37% 44% 14807 14797 10 0 14807 14.5MB 15.0MB 0% 3.09MB 3.68MB 0% 11.4MB 11.4MB 4.34MB 2.79MB
48352 53616
39% 43% 15075 15075 0 0 15076 14.9MB 15.6MB 0% 3.24MB 3.86MB 0%
11.7MB 11.7MB 3.48MB 696KB 50124
55971
32% 42% 14998 14998 0 0 14997 15.1MB 15.8MB 0% 3.23MB 3.87MB 0%
11.9MB 11.9MB 3.68MB 815KB 49606
55692
38% 43% 15038 15025 13 0 15036 14.7MB 15.2MB 0% 3.27MB 3.92MB 0% 11.4MB 11.3MB 3.46MB 15.8KB
50256 56150
43% 44% 15132 15132 0 0 15133 15.0MB 15.7MB 0% 3.22MB 3.87MB 0% 11.8MB 11.8MB 1.93MB 15.9KB 50030
55938
34% 42% 15828 15817 10 0 15827 15.8MB 16.5MB 0% 3.39MB 4.10MB 0% 12.4MB 12.3MB 4.02MB 21.6MB
52142 58771
28% 39% 11807 11807 0 0 11807 12.3MB 13.1MB 0% 2.55MB 3.07MB 0%
9.80MB 9.99MB 6.76MB 27.9MB 38752
43748
33% 42% 15108 15108 0 0 15107 15.1MB 15.5MB 0% 3.32MB 3.91MB 0%
11.7MB 11.6MB 3.50MB 1.17MB 50903
56143
```

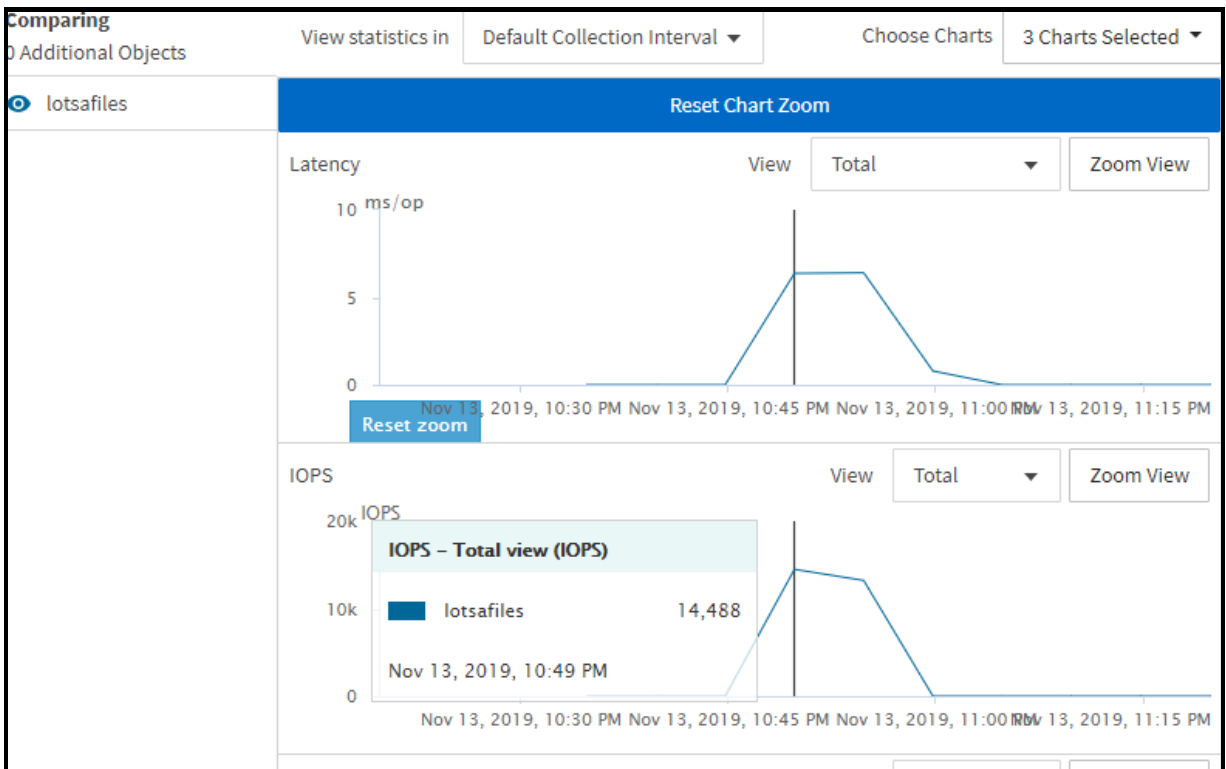
17. Examples
Sample FlexVol to FlexGroup conversion

```

32% 42% 16143 16133 10 0 16143 15.1MB 15.8MB 0% 3.28MB 3.95MB 0% 11.8MB 11.8MB 3.78MB
9.00MB
50922 57403
24% 34% 8843 8843 0 0 8861 14.2MB 14.9MB 0% 3.70MB 4.44MB 0% 10.5MB 10.5MB
8.46MB 10.7MB 46174
53157
27% 37% 10949 10949 0 0 11177 9.91MB 10.2MB 0% 2.45MB 2.84MB 0% 7.46MB 7.40MB
5.55MB 1.67MB 31764
35032
28% 38% 12580 12567 13 0 12579 13.3MB 13.8MB 0% 2.76MB 3.26MB 0% 10.5MB 10.6MB 3.92MB 19.9KB
44119 48488
30% 40% 14300 14300 0 0 14298 14.2MB 14.7MB 0% 3.09MB 3.68MB 0% 11.1MB 11.1MB 2.66MB 600KB
47282
52789
31% 41% 14514 14503 10 0 14514 14.3MB 14.9MB 0% 3.15MB 3.75MB 0% 11.2MB 11.2MB 3.65MB 728KB
48093
53532
31% 42% 14626 14626 0 0 14626 14.3MB 14.9MB 0% 3.16MB 3.77MB 0% 11.1MB 11.1MB 4.84MB 1.14MB
47936
53645
cluster: cluster.cluster: 11/13/2019 22:44:39
cpu cpu total fcache total data data cluster cluster cluster disk disk pkts pkts avg
busy ops nfs-ops cifs-ops ops spin-ops rcv sent busy rcv sent busy rcv sent read write rcv
sent
-----
30% 39% 15356 15349 7 0 15370 15.3MB 15.8MB 0% 3.29MB 3.94MB 0% 12.0MB 11.8MB 3.18MB 6.90MB 50493
56425
32% 42% 14156 14146 10 0 14156 14.6MB 15.3MB 0% 3.09MB 3.68MB 0% 11.5MB 11.7MB 5.49MB 16.3MB
48159 53678

```

This is what the performance looked like from Active IQ:



And now we have a single member FlexGroup volume.

```
cluster::*> volume show lots*
Vserver   Volume           Aggregate      State    Type  Size    Available  Used%
-----
DEMO      lotsafiles       -              online   RW    10TB    7.33TB    0%
DEMO      lotsafiles__0001 aggr1_node1   online   RW    10TB    7.33TB    0%
2 entries were displayed.
```

And our Snapshot copies are still there but are marked as pre-conversion.

```
cluster::> set diag
cluster::*> snapshot show -vserver DEMO -volume lotsafiles -fields is-convert-
recovery,state
vserver volume      snapshot                                     state          is-convert-recovery
-----
DEMO    lotsafiles base                               pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_1705          pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_1805          pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_1905          pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_2005          pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_2105          pre-conversion false
DEMO    lotsafiles hourly.2019-11-13_2205          pre-conversion false
DEMO    lotsafiles clone_clone.2019-11-13_223144.0 pre-conversion false
DEMO    lotsafiles convert.2019-11-13_224411       pre-conversion true
9 entries were displayed.
```

When a Snapshot copy is in pre-conversion state, using it for SnapRestore operation fails.

```
cluster::*> snapshot restore -vserver DEMO -volume lotsafiles -snapshot convert.2019-11-
13_224411

Error: command failed: Promoting a pre-conversion Snapshot copy is not supported.
```

However, we can still obtain files from the client using the Snapshot copies.

```
[root@centos7 scripts]# cd /lotsafiles/.snapshot/convert.2019-11-13_224411/pre-convert/
[root@centos7 pre-convert]# ls
topdir_0 topdir_14 topdir_2 topdir_25 topdir_30 topdir_36 topdir_41 topdir_47 topdir_52
topdir_58 topdir_63 topdir_69 topdir_74 topdir_8 topdir_85 topdir_90 topdir_96
topdir_1 topdir_15 topdir_20 topdir_26 topdir_31 topdir_37 topdir_42 topdir_48 topdir_53
topdir_59 topdir_64 topdir_7 topdir_75 topdir_80 topdir_86 topdir_91 topdir_97
topdir_10 topdir_16 topdir_21 topdir_27 topdir_32 topdir_38 topdir_43 topdir_49 topdir_54
topdir_6 topdir_65 topdir_70 topdir_76 topdir_81 topdir_87 topdir_92 topdir_98
topdir_11 topdir_17 topdir_22 topdir_28 topdir_33 topdir_39 topdir_44 topdir_5 topdir_55
topdir_60 topdir_66 topdir_71 topdir_77 topdir_82 topdir_88 topdir_93 topdir_99 topdir_12
topdir_18 topdir_23 topdir_29 topdir_34 topdir_4 topdir_45 topdir_50 topdir_56 topdir_61
topdir_67 topdir_72 topdir_78 topdir_83 topdir_89 topdir_94
topdir_13 topdir_19 topdir_24 topdir_3 topdir_35 topdir_40 topdir_46 topdir_51 topdir_57
topdir_62 topdir_68 topdir_73 topdir_79 topdir_84 topdir_9 topdir_95
```

Growing the newly converted FlexGroup volume is simple. We can add more member volumes by using volume expand.

```
cluster::*> volume expand -vserver DEMO -volume lotsafiles -aggr-list
aggr1_node1,aggr1_node2 - aggr-list-multiplier 2

Warning: The following number of constituents of size 10TB will be added to FlexGroup
"lotsafiles": 4. Expanding the FlexGroup will cause the state of all Snapshot copies
to be set to "partial".
Partial Snapshot copies cannot be restored.
Do you want to continue? {y|n}: y

Warning: FlexGroup "lotsafiles" is a converted flexible volume. If this volume is
expanded, it will no longer be able to be converted back to being a flexible volume.
Do you want to continue? {y|n}: y
[Job 23676] Job succeeded: Successful
```

But remember, the data doesn't redistribute. The original member volume keeps the files in place.

```
cluster::*> df -i lots*
Filesystem iused ifree %iused Mounted on Vserver
/vol/lotsafiles/ 3630682 102624948 3% /lotsafiles DEMO
/vol/lotsafiles__0001/ 3630298 17620828 17% /lotsafiles DEMO
/vol/lotsafiles__0002/ 96 21251030 0% --- DEMO
/vol/lotsafiles__0003/ 96 21251030 0% --- DEMO
/vol/lotsafiles__0004/ 96 21251030 0% --- DEMO
/vol/lotsafiles__0005/ 96 21251030 0% --- DEMO
6 entries were displayed.

cluster::*> df -h lots*
Filesystem total used avail capacity Mounted on Vserver
/vol/lotsafiles/ 47TB 2735MB 14TB 0% /lotsafiles DEMO
/vol/lotsafiles/.snapshot
2560GB 49MB 2559GB 0% /lotsafiles/.snapshot DEMO
/vol/lotsafiles__0001/ 9728GB 2505MB 7505GB 0% /lotsafiles DEMO
/vol/lotsafiles__0001/.snapshot
512GB 49MB 511GB 0% /lotsafiles/.snapshot DEMO
/vol/lotsafiles__0002/ 9728GB 57MB 7505GB 0% --- DEMO
/vol/lotsafiles__0002/.snapshot
512GB 0B 512GB 0% --- DEMO
/vol/lotsafiles__0003/ 9728GB 57MB 7766GB 0% --- DEMO
/vol/lotsafiles__0003/.snapshot
512GB 0B 512GB 0% --- DEMO
/vol/lotsafiles__0004/ 9728GB 57MB 7505GB 0% --- DEMO
/vol/lotsafiles__0004/.snapshot
512GB 0B 512GB 0% --- DEMO
/vol/lotsafiles__0005/ 9728GB 57MB 7766GB 0% --- DEMO
/vol/lotsafiles__0005/.snapshot 512GB 0B
512GB 0% --- DEMO
12 entries were displayed.
```

Converting FlexVols in existing SnapMirror relationships – example

The following shows an example of converting a FlexVol that has an existing SnapMirror.

Procedure ▶▶▶

1 Here is a volume in a SnapMirror relationship.

```
cluster::*> snapmirror show -destination-path data_dst -fields state
source-path destination-path state
-----
DEMO:data    DEMO:data_dst    Snapmirrored
```

2 If you try to convert the source, you get an error.

```
cluster::*> vol conversion start -vserver DEMO -volume data -check-only true

Error: command failed: Cannot convert volume "data" in Vserver "DEMO" to a
FlexGroup. Correct the following issues and retry the command:
  * Cannot convert source volume "data" because destination volume "data_dst"
of the SnapMirror relationship with "data" as the source is not converted. First
check if the source can be converted to a FlexGroup volume using "vol conversion
start -volume data -convert-to flexgroup -check-only true". If the conversion of
the source can proceed then first convert the destination and then convert the
source.
```

3 So, you need to convert the destination first. To do that, you must quiesce the SnapMirror relationship.

```
cluster::*> vol conversion start -vserver DEMO -volume data_dst -check-only true

Error: command failed: Cannot convert volume "data_dst" in Vserver "DEMO" to a
FlexGroup. Correct the following issues and retry the command:
  * The relationship was not quiesced. Quiesce SnapMirror relationship using
"snapmirror quiesce -destination-path data_dst" and then try the conversion.
```

4 Now you can convert the volume.

```
cluster::*> snapmirror quiesce -destination-path DEMO:data_dst
Operation succeeded: snapmirror quiesce for destination "DEMO:data_dst".

cluster::*> vol conversion start -vserver DEMO -volume data_dst -check-only true
Conversion of volume "data_dst" in Vserver "DEMO" to a FlexGroup can proceed
with the following warnings:
  * After the volume is converted to a FlexGroup, it will not be possible to change
it back to a flexible volume.
  * Converting flexible volume "data_dst" in Vserver "DEMO" to a FlexGroup will
cause the state of all Snapshot copies from the volume to be set to "pre-
conversion". Pre-conversion Snapshot copies cannot be restored.
```

5 When you convert the volume, the system lets you know your next steps.

```
cluster::*> vol conversion start -vserver DEMO -volume data_dst

Warning: After the volume is converted to a FlexGroup, it will not be possible to
change it back to a flexible volume.
Do you want to continue? {y|n}: y
Warning: Converting flexible volume "data_dst" in Vserver "DEMO" to a FlexGroup
will cause the state of all Snapshot copies from the volume to be set to "pre-
conversion". Pre-conversion Snapshot copies cannot be restored.
Do you want to continue? {y|n}: y
[Job 23710] Job succeeded: SnapMirror destination volume "data_dst" has been
successfully converted to a FlexGroup volume. You must now convert the
relationship's source volume, "DEMO:data", to a FlexGroup. Then, re-establish
the SnapMirror relationship using the "snapmirror resync" command.
```

6 Now convert the source volume.

```
cluster::*> vol conversion start -vserver DEMO -volume data

Warning: After the volume is converted to a FlexGroup, it will not be possible to
change it back to a flexible volume.
Do you want to continue? {y|n}: y
Warning: Converting flexible volume "data" in Vserver "DEMO" to a FlexGroup will
cause the state of all Snapshot copies from the volume to be set to "pre-
conversion". Pre-conversion Snapshot copies cannot be restored.
Do you want to continue? {y|n}: y
[Job 23712] Job succeeded: success
```

7 Resync the mirror.

```
cluster::*> snapmirror resync -destination-path DEMO:data_dst
Operation is queued: snapmirror resync to destination "DEMO:data_dst".

cluster::*> snapmirror show -destination-path DEMO:data_dst -fields state
source-path destination-path state
-----
DEMO:data DEMO:data_dst Snapmirrored
```

Conversion works fine, but the most important part of a SnapMirror relationship is the restore. So you must see if you can access files from the destination volume's Snapshot copy.

8 First, mount the source and destination and compare ls output.

```
# mount -o nfsvers=3 DEMO:/data_dst /dst
# mount -o nfsvers=3 DEMO:/data /data
```

This is what's in the source volume.

```
# ls -lah /data
total 14G
drwxrwxrwx 6 root root 4.0K Nov 14 11:57 .
dr-xr-xr-x. 54 root root 4.0K Nov 15 10:08 ..
drwxrwxrwx 2 root root 4.0K Sep 14 2018 cifslink
drwxr-xr-x 12 root root 4.0K Nov 16 2018 nas
-rwxrwxrwx 1 prof1 ProfGroup 0 Oct 3 14:32 newfile
drwxrwxrwx 5 root root 4.0K Nov 15 10:06 .snapshot
lrwxrwxrwx 1 root root 23 Sep 14 2018 symlink -> /shared/unix/linkedfile
drwxrwxrwx 2 root bin 4.0K Jan 31 2019 test
drwxrwxrwx 3 root root 4.0K Sep 14 2018 unix
-rwxrwxrwx 1 newuser1 ProfGroup 0 Jan 14 2019 userfile
-rwxrwxrwx 1 root root 6.7G Nov 14 11:58 Windows2.iso
-rwxrwxrwx 1 root root 6.7G Nov 14 11:37 Windows.iso
```

The destination volume matches exactly, as it should.

```
# ls -lah /dst
total 14G
drwxrwxrwx 6 root root 4.0K Nov 14 11:57 .
dr-xr-xr-x. 54 root root 4.0K Nov 15 10:08 ..
drwxrwxrwx 2 root root 4.0K Sep 14 2018 cifslink
dr-xr-xr-x 2 root root 0 Nov 15 2018 nas
-rwxrwxrwx 1 prof1 ProfGroup 0 Oct 3 14:32 newfile
drwxrwxrwx 4 root root 4.0K Nov 15 10:05 .snapshot
lrwxrwxrwx 1 root root 23 Sep 14 2018 symlink -> /shared/unix/linkedfile
drwxrwxrwx 2 root bin 4.0K Jan 31 2019 test
drwxrwxrwx 3 root root 4.0K Sep 14 2018 unix
-rwxrwxrwx 1 newuser1 ProfGroup 0 Jan 14 2019 userfile
-rwxrwxrwx 1 root root 6.7G Nov 14 11:58 Windows2.iso
-rwxrwxrwx 1 root root 6.7G Nov 14 11:37 Windows.iso
```

9 If you ls to the Snapshot copy in the destination volume, you see the expected files.

```
# ls -lah /dst/.snapshot/snapmirror.7e3cc08e-d9b3-11e6-85e2-
00a0986b1210_2163227795.2019-11- 15_100555/
total 14G
drwxrwxrwx 6 root root 4.0K Nov 14 11:57 .
drwxrwxrwx 4 root root 4.0K Nov 15 10:05 ..
drwxrwxrwx 2 root root 4.0K Sep 14 2018 cifslink
dr-xr-xr-x 2 root root 0 Nov 15 2018 nas
-rwxrwxrwx 1 prof1 ProfGroup 0 Oct 3 14:32 newfile
lrwxrwxrwx 1 root root 23 Sep 14 2018 symlink -> /shared/unix/linkedfile
drwxrwxrwx 2 root bin 4.0K Jan 31 2019 test
drwxrwxrwx 3 root root 4.0K Sep 14 2018 unix
-rwxrwxrwx 1 newuser1 ProfGroup 0 Jan 14 2019 userfile
-rwxrwxrwx 1 root root 6.7G Nov 14 11:58 Windows2.iso
-rwxrwxrwx 1 root root 6.7G Nov 14 11:37 Windows.iso
```

10 Now expand the FlexGroup source to provide more capacity.

```
cluster::*> volume expand -vserver DEMO -volume data -aggr-list
aggr1_node1,aggr1_node2 -aggr- list-multiplier

Warning: The following number of constituents of size 30TB will be added to
FlexGroup "data": 4. Expanding the FlexGroup will cause the state of all
Snapshot copies to be set to "partial".
Partial Snapshot copies cannot be restored.
Do you want to continue? {y|n}: y [Job 23720] Job succeeded: Successful
```

The source volume now has five member volumes. The destination volume has only one.

```
cluster::*> vol show -vserver DEMO -volume data*
Vserver Volume Aggregate State Type Size Available Used%
-----
DEMO data - online RW 150TB 14.89TB 0%
DEMO data__0001 aggr1_node2 online RW 30TB 7.57TB 0%
DEMO data__0002 aggr1_node1 online RW 30TB 7.32TB 0%
DEMO data__0003 aggr1_node2 online RW 30TB 7.57TB 0%
DEMO data__0004 aggr1_node1 online RW 30TB 7.32TB 0%
DEMO data__0005 aggr1_node2 online RW 30TB 7.57TB 0%
DEMO data_dst - online DP 30TB 7.32TB 0%
DEMO data_dst__0001
aggr1_node1 online DP 30TB 7.32TB 0%
8 entries were displayed.
```

11 Update the mirror, and ONTAP fixes it for you.

```
cluster::*> snapmirror update -destination-path DEMO:data_dst  
Operation is queued: snapmirror update of destination "DEMO:data_dst".
```

The update initially fails with the following error message:

```
Last Transfer Error: A SnapMirror transfer for the relationship with destination  
FlexGroup "DEMO:data_dst" was aborted because the source FlexGroup was expanded.  
A SnapMirror AutoExpand job with id "23727" was created to expand the  
destination FlexGroup and to trigger a SnapMirror transfer for the SnapMirror  
relationship. After the SnapMirror transfer is successful, the "healthy" field  
of the SnapMirror relationship will be set to "true". The job can be monitored  
using either the "job show -id 23727" or "job history show -id 23727" commands.
```

12 The job expands the volume, and then you can update again.

```
cluster::*> job show -id 23727  
Owning  
Job ID Name Vserver Node State  
-----  
23727 Snapmirror Expand cluster  
node1 Success  
Description: SnapMirror FG Expand data_dst  
  
cluster::*> snapmirror show -destination-path DEMO:data_dst -fields state  
source-path destination-path state  
-----  
DEMO:data DEMO:data_dst Snapmirrored
```

Now both FlexGroup volumes have the same number of member volumes.

```
cluster::*> vol show -vserver DEMO -volume data*  
Vserver Volume Aggregate State Type Size Available Used%  
-----  
DEMO data - online RW 150TB 14.88TB 0%  
DEMO data__0001 aggr1_node2 online RW 30TB 7.57TB 0%  
DEMO data__0002 aggr1_node1 online RW 30TB 7.32TB 0%  
DEMO data__0003 aggr1_node2 online RW 30TB 7.57TB 0%  
DEMO data__0004 aggr1_node1 online RW 30TB 7.32TB 0%  
DEMO data__0005 aggr1_node2 online RW 30TB 7.57TB 0%  
DEMO data_dst - online DP 150TB 14.88TB 0%  
DEMO data_dst__0001  
aggr1_node1 online DP 30TB 7.32TB 0%  
DEMO data_dst__0002  
aggr1_node1 online DP 30TB 7.32TB 0%  
DEMO data_dst__0003  
aggr1_node2 online DP 30TB 7.57TB 0%  
DEMO data_dst__0004  
aggr1_node1 online DP 30TB 7.32TB 0%  
DEMO data_dst__0005  
aggr1_node2 online DP 30TB 7.57TB 0%  
12 entries were displayed.
```



Sample FlexVol to FlexGroup conversion – 500 million files

In this example, we convert a FlexVol volume with 500 million files to a FlexGroup volume.

```
cluster::*> vol show -vserver DEMO -volume fvconvert -fields files,files-used,is-
flexgroup
vserver volume    files      files-used is-flexgroup
-----
DEMO      fvconvert 2040109451 502631608 false
```

Because it took so long to create that many files, we created a [FlexClone volume](#) of it and split it. This approach lets us keep the origin volume intact and test without risk.

In this example, the cloning process took about 30 minutes:

```
cluster::*> vol clone split start -vserver DEMO -flexclone fvconvert -foreground true

Warning: Are you sure you want to split clone volume fvconvert in Vserver DEMO ?
{y|n}: y
[Job 24230] 0% inodes processed.

cluster::*> job history show -id 24230 -fields starttime,endtime
node          record  vserver      endtime      starttime
-----
node1         2832338 cluster      12/09 10:27:08 12/09 09:58:16
```

After the clone split, we ran the check. We had to run `volume clone sharing-by-split undo` to get rid of shared FlexClone blocks, which took some time, but then the check produced this output:

```
cluster::*> volume conversion start -vserver DEMO -volume fvconvert -foreground true
-check-only true
Conversion of volume "fvconvert" in Vserver "DEMO" to a FlexGroup can proceed with the
following warnings:
* After the volume is converted to a FlexGroup, it will not be possible to change it
back to a flexible volume.
```

We then ran the script that we ran earlier to generate load and watched the statistics on the cluster to see if we hit any outage. Again, the conversion took seconds (with 500 million files) and there was just a small, barely noticeable delay.

```
cluster::*> volume conversion start -vserver DEMO -volume fvconvert -foreground
true

Warning: After the volume is converted to a FlexGroup, it will not be possible to
change it back to a flexible volume.
Do you want to continue? {y|n}: y
[Job 24259] Job succeeded: success
```

Figure 89 Sample statistics from conversion process

| cpu | | total | fsache | | total | | total data | | data | | cluster | | cluster | | disk | | pkts | | |
|---|------|-------|---------|----------|-------|----------|------------|--------|------|--------|---------|------|---------|--------|--------|--------|-------|-------|--|
| avg | busy | ops | nfs-ops | cifs-ops | ops | spin-ops | recv | sent | busy | recv | sent | busy | recv | sent | read | write | recv | sent | |
| 12% | 21% | 150 | 140 | 9 | 0 | 145 | 2.22MB | 2.48MB | 0% | 50.9KB | 315KB | 0% | 2.17MB | 2.17MB | 16.1MB | 5.62MB | 820 | 810 | |
| 7% | 13% | 46 | 46 | 0 | 0 | 46 | 1.36MB | 1.46MB | 0% | 24.0KB | 123KB | 0% | 1.34MB | 1.34MB | 11.0MB | 19.9KB | 452 | 446 | |
| 14% | 24% | 1721 | 1721 | 0 | 0 | 1721 | 2.25MB | 2.48MB | 0% | 197KB | 438KB | 0% | 2.06MB | 2.06MB | 24.3MB | 9.55MB | 2814 | 2917 | |
| 15% | 25% | 3576 | 3573 | 2 | 0 | 3575 | 5.45MB | 5.77MB | 0% | 985KB | 1.23MB | 0% | 4.53MB | 4.53MB | 18.1MB | 1.22MB | 14847 | 15681 | |
| 16% | 21% | 1211 | 1180 | 30 | 0 | 1209 | 2.41MB | 2.68MB | 0% | 275KB | 559KB | 0% | 2.14MB | 2.14MB | 16.9MB | 2.35MB | 4249 | 4751 | |
| 27% | 34% | 1979 | 1968 | 10 | 0 | 1978 | 39.4MB | 22.2MB | 1% | 19.0MB | 1.69MB | 0% | 20.4MB | 20.5MB | 14.3MB | 1.14MB | 21043 | 9869 | |
| 18% | 23% | 2666 | 2664 | 1 | 0 | 2665 | 3.43MB | 4.54MB | 0% | 583KB | 1.79MB | 0% | 2.86MB | 2.75MB | 14.6MB | 19.9KB | 7686 | 8755 | |
| 23% | 34% | 1917 | 1917 | 0 | 0 | 1917 | 2.88MB | 4.22MB | 0% | 563KB | 1.89MB | 0% | 2.33MB | 2.33MB | 19.7MB | 19.1MB | 7352 | 8323 | |
| 36% | 58% | 2264 | 2260 | 4 | 0 | 2264 | 3.25MB | 4.40MB | 0% | 474KB | 1.61MB | 0% | 2.79MB | 2.79MB | 34.2MB | 19.0MB | 5763 | 6342 | |
| 26% | 45% | 1351 | 1303 | 47 | 0 | 1350 | 7.98MB | 5.64MB | 0% | 2.93MB | 595KB | 0% | 5.05MB | 5.05MB | 34.4MB | 19.7KB | 8267 | 7036 | |
| 28% | 45% | 2032 | 2002 | 29 | 0 | 2031 | 24.9MB | 13.9MB | 0% | 11.7MB | 597KB | 0% | 13.2MB | 13.3MB | 33.5MB | 1.66MB | 15344 | 8798 | |
| 26% | 49% | 1813 | 1745 | 67 | 0 | 1812 | 28.6MB | 16.5MB | 1% | 13.7MB | 728KB | 0% | 14.9MB | 15.8MB | 36.7MB | 19.8KB | 17761 | 9963 | |
| 27% | 50% | 2438 | 2416 | 22 | 0 | 2437 | 18.6MB | 10.3MB | 0% | 8.08MB | 860KB | 0% | 10.5MB | 9.48MB | 37.8MB | 11.9KB | 13884 | 9831 | |
| 31% | 58% | 2043 | 2002 | 40 | 0 | 2043 | 18.8MB | 12.6MB | 0% | 8.38MB | 726KB | 0% | 10.4MB | 11.9MB | 64.3MB | 150MB | 13331 | 9469 | |
| 35% | 67% | 1475 | 1413 | 62 | 0 | 1474 | 22.8MB | 13.1MB | 0% | 10.4MB | 812KB | 0% | 12.3MB | 12.3MB | 86.6MB | 53.7MB | 14723 | 9296 | |
| 35% | 66% | 2028 | 1961 | 66 | 0 | 2022 | 29.3MB | 15.9MB | 1% | 13.5MB | 612KB | 0% | 15.8MB | 15.3MB | 86.3MB | 2.84MB | 16522 | 8716 | |
| 38% | 71% | 2446 | 2413 | 32 | 0 | 2444 | 13.6MB | 9.10MB | 0% | 5.90MB | 911KB | 0% | 7.70MB | 8.21MB | 78.9MB | 19.8KB | 13169 | 10819 | |
| 19% | 34% | 1771 | 1727 | 43 | 0 | 1770 | 34.0MB | 17.9MB | 1% | 15.3MB | 699KB | 0% | 18.7MB | 17.2MB | 11.0MB | 11.9KB | 19605 | 10707 | |
| 17% | 30% | 749 | 696 | 53 | 0 | 748 | 34.3MB | 18.6MB | 1% | 17.5MB | 419KB | 0% | 16.8MB | 18.2MB | 11.3MB | 5.32MB | 18226 | 7898 | |
| 19% | 35% | 1194 | 1137 | 56 | 0 | 1194 | 16.4MB | 8.62MB | 0% | 6.54MB | 261KB | 0% | 9.87MB | 8.37MB | 12.0MB | 55.4MB | 7586 | 3595 | |
| map9-tme-8040: cluster:cluster: l2/10/2019 11:03:06 | | | | | | | | | | | | | | | | | | | |
| cpu | | total | fsache | | total | | total data | | data | | cluster | | cluster | | disk | | pkts | | |
| avg | busy | ops | nfs-ops | cifs-ops | ops | spin-ops | recv | sent | busy | recv | sent | busy | recv | sent | read | write | recv | sent | |
| 25% | 41% | 2954 | 2915 | 38 | 0 | 2953 | 35.2MB | 20.5MB | 1% | 16.5MB | 814KB | 0% | 18.2MB | 19.7MB | 9.84MB | 110MB | 22564 | 13289 | |
| 23% | 41% | 2292 | 2250 | 41 | 0 | 2291 | 17.7MB | 10.3MB | 0% | 6.73MB | 866KB | 0% | 10.5MB | 9.44MB | 11.6MB | 109MB | 15241 | 12193 | |
| 25% | 43% | 2415 | 2369 | 46 | 0 | 2414 | 52.8MB | 19.5MB | 1% | 16.8MB | 909KB | 1% | 36.1MB | 18.6MB | 11.5MB | 26.7MB | 25076 | 13688 | |
| 29% | 40% | 2821 | 2792 | 29 | 0 | 2820 | 49.8MB | 35.5MB | 1% | 22.6MB | 852KB | 1% | 26.2MB | 24.7MB | 11.2MB | 5.45MB | 26766 | 13167 | |
| 26% | 41% | 2584 | 2550 | 33 | 0 | 2582 | 66.3MB | 38.1MB | 2% | 35.1MB | 1.13MB | 1% | 31.2MB | 37.0MB | 9.87MB | 11.7KB | 37901 | 17822 | |
| 34% | 61% | 3438 | 3397 | 40 | 0 | 3437 | 92.6MB | 55.4MB | 4% | 51.7MB | 1.65MB | 2% | 40.5MB | 53.7MB | 9.35MB | 1.04MB | 54703 | 25093 | |
| 25% | 41% | 5686 | 5664 | 22 | 0 | 5684 | 40.8MB | 24.0MB | 1% | 18.9MB | 1.63MB | 0% | 21.9MB | 22.4MB | 11.0MB | 15.5MB | 34334 | 25357 | |
| 19% | 31% | 4678 | 4650 | 28 | 0 | 4678 | 52.0MB | 28.9MB | 2% | 25.1MB | 1.51MB | 1% | 26.5MB | 27.4MB | 13.8MB | 109MB | 35020 | 21615 | |
| 18% | 29% | 3812 | 3794 | 18 | 0 | 3810 | 32.4MB | 18.4MB | 1% | 14.1MB | 1.13MB | 0% | 18.3MB | 17.3MB | 13.4MB | 89.7MB | 24517 | 17097 | |
| 12% | 20% | 661 | 633 | 27 | 0 | 661 | 42.3MB | 22.9MB | 1% | 21.8MB | 458KB | 0% | 20.5MB | 22.5MB | 19.3MB | 63.7MB | 18672 | 4920 | |
| 19% | 29% | 4944 | 4822 | 21 | 0 | 4942 | 47.7MB | 25.1MB | 1% | 21.1MB | 1.51MB | 1% | 26.6MB | 23.6MB | 6.85MB | 19.9KB | 32420 | 21523 | |
| 21% | 34% | 6205 | 6185 | 19 | 0 | 6205 | 44.2MB | 26.5MB | 1% | 20.5MB | 1.82MB | 1% | 23.7MB | 24.7MB | 8.10MB | 2.64MB | 37075 | 28031 | |
| 20% | 33% | 5652 | 5636 | 16 | 0 | 5651 | 34.5MB | 20.0MB | 1% | 15.0MB | 1.46MB | 0% | 19.5MB | 18.5MB | 5.43MB | 19.8KB | 27591 | 20579 | |
| 20% | 29% | 6400 | 6376 | 23 | 0 | 6399 | 40.7MB | 24.7MB | 1% | 18.1MB | 2.11MB | 0% | 22.6MB | 22.6MB | 5.79MB | 1.10MB | 37445 | 29159 | |
| 26% | 41% | 6493 | 6469 | 24 | 0 | 6492 | 49.6MB | 28.3MB | 1% | 23.1MB | 1.82MB | 1% | 26.5MB | 26.5MB | 9.06MB | 61.3MB | 39040 | 27688 | |
| 27% | 40% | 7860 | 7847 | 12 | 0 | 7860 | 30.3MB | 19.9MB | 1% | 12.6MB | 2.17MB | 0% | 17.7MB | 17.7MB | 8.96MB | 94.1MB | 35235 | 31193 | |
| 22% | 34% | 7093 | 7073 | 20 | 0 | 7092 | 41.0MB | 24.9MB | 1% | 18.1MB | 2.03MB | 0% | 22.9MB | 22.9MB | 7.22MB | 1.10MB | 38476 | 31338 | |

Then, as the script was running, we added new member volumes to the FlexGroup volume. Again, there was no disruption.

```
cluster::*> volume expand -vserver DEMO -volume fvconvert -aggr-list aggr1_node1 -
aggr-list- multiplier 3 -foreground true

Warning: The following number of constituents of size 40TB will be added to FlexGroup
"fvconvert": 3.
Do you want to continue? {y|n}: y [Job 24261] Job succeeded: Successful
```

Then we added four more member volumes:

```
cluster::*> volume expand -vserver DEMO -volume fvconvert -aggr-list aggr1_node2 -
aggr-list- multiplier 4

Warning: The following number of constituents of size 40TB will be added to FlexGroup
"fvconvert": 4.
Do you want to continue? {y|n}: y
[Job 24264] Job succeeded: Successful
```

As an added bonus, we started to see more total IOPS for the workload. The job itself took much less time overall than when we ran it on a FlexVol volume, because the FlexGroup volume's parallel ingest started to help the script run faster.

17. Examples
Sample FlexVol to FlexGroup conversion – 500 million files

Figure 90 Sample statistics during conversion process – Adding member volumes

```

23% 33% 7181 7181 0 0 7181 8.71MB 7.13MB 0% 1.43MB 2.72MB 0% 1.43MB 1.83MB 8.07MB 8.71KB 1.78MB 1.97MB
23% 37% 8121 8121 0 0 8120 8.66MB 10.5MB 0% 1.82MB 3.51MB 0% 6.82MB 6.82MB 8.41MB 999KB 25436 28458
27% 40% 8484 8484 0 0 8484 8.49MB 10.1MB 0% 1.83MB 3.39MB 0% 6.66MB 6.66MB 9.65MB 17.0MB 25528 28668
29% 46% 7959 7959 0 0 7959 9.05MB 10.8MB 0% 1.83MB 3.40MB 0% 7.22MB 7.42MB 9.22MB 3.19MB 26265 29159
Added member volumes around here:
0% 5203 7.35MB 8.64MB 0% 1.36MB 2.84MB 0% 5.99MB 5.80MB 8.62MB 11.9KB 20006 22621
32% 46% 7807 7807 0 0 7807 10.7MB 11.7MB 0% 1.59MB 3.22MB 0% 9.12MB 8.45MB 12.9MB 8.52MB 27012 28484
38% 55% 11506 11506 0 0 11506 13.4MB 14.7MB 0% 2.25MB 3.80MB 0% 11.2MB 10.9MB 14.8MB 9.61MB 34364 37793
24% 30% 7100 7100 0 0 7100 7.89MB 10.5MB 0% 1.73MB 3.40MB 0% 6.16MB 7.07MB 11.6MB 29.7MB 20088 23817
28% 28% 9791 9791 0 0 9791 5.27MB 6.93MB 0% 2.11MB 3.74MB 0% 3.16MB 3.19MB 5.50MB 2.12MB 14271 16913
18% 19% 6086 6086 0 0 6085 3.67MB 4.90MB 0% 1.52MB 2.76MB 0% 2.14MB 2.13MB 9.34MB 11.9KB 8681 11684
16% 19% 7452 7452 0 0 7452 4.46MB 6.08MB 0% 1.85MB 3.47MB 0% 2.61MB 2.60MB 5.60MB 2.44MB 11521 13652
16% 21% 7188 7188 0 0 7187 3.91MB 5.43MB 0% 1.63MB 3.17MB 0% 2.27MB 2.26MB 7.65MB 10.8MB 9702 12233
18% 21% 7786 7786 0 0 7787 4.84MB 6.54MB 0% 1.74MB 3.44MB 0% 3.09MB 3.10MB 12.1MB 11.3MB 10390 13281
18% 22% 7938 7938 0 0 7938 3.97MB 5.25MB 0% 1.62MB 2.91MB 0% 2.34MB 2.34MB 8.13MB 1.63MB 9590 12489
18% 21% 7971 7971 0 0 7971 4.56MB 6.20MB 0% 1.89MB 3.54MB 0% 2.67MB 2.66MB 6.46MB 11.9KB 10817 14407
15% 16% 5583 5573 9 0 5578 4.20MB 5.75MB 0% 1.86MB 3.43MB 0% 2.33MB 2.32MB 5.36MB 904KB 11396 12956
sntap9-tme-8040: cluster:cluster: 12/10/2019 11:06:28
cpu cpu total fcache total data data data cluster cluster
avg busy ops nfs-ops cifs-ops ops spin-ops recv sent busy recv recv recv sent read write pkts pkts
-----
17% 21% 7321 7321 0 0 7321 3.83MB 5.85MB 0% 1.67MB 3.38MB 0% 2.16MB 2.45MB 9.50MB 6.07MB 9707 12093
20% 21% 6746 6746 0 0 6746 3.87MB 5.08MB 0% 1.52MB 3.02MB 0% 2.35MB 2.06MB 16.1MB 14.2MB 8513 11285
18% 20% 6747 6744 2 0 6745 3.70MB 5.07MB 0% 1.57MB 2.94MB 0% 2.13MB 2.13MB 9.23MB 899KB 9568 11563
15% 22% 8140 8140 0 0 8140 4.34MB 6.07MB 0% 1.80MB 3.45MB 0% 2.54MB 2.54MB 8.08MB 11.9KB 10437 13132
26% 30% 7517 7517 0 0 7516 3.83MB 5.55MB 0% 1.69MB 3.40MB 0% 2.14MB 2.14MB 8.47MB 1.01MB 9575 12652
24% 25% 8820 8820 0 0 8819 4.32MB 6.06MB 0% 1.94MB 3.69MB 0% 2.37MB 2.37MB 9.98MB 19.9KB 10781 14381
23% 26% 7112 7105 6 0 7131 4.11MB 5.42MB 0% 1.48MB 2.80MB 0% 2.62MB 2.62MB 20.3MB 21.9MB 8485 11614
23% 23% 8454 8454 0 0 8454 4.37MB 6.09MB 0% 1.89MB 3.62MB 0% 2.47MB 2.47MB 6.75MB 2.02MB 10508 13894
19% 23% 7920 7920 0 0 7920 4.21MB 5.87MB 0% 1.67MB 3.33MB 0% 2.54MB 2.54MB 5.82MB 19.9KB 9329 12523
18% 22% 7042 7042 0 0 7047 3.81MB 5.42MB 0% 1.53MB 3.18MB 0% 2.28MB 2.28MB 6.06MB 625KB 9750 11009
20% 24% 7616 7616 0 0 7616 3.82MB 5.13MB 0% 1.83MB 3.14MB 0% 1.99MB 1.99MB 11.0MB 19.9KB 10289 13747
28% 31% 7375 7375 0 0 7380 4.18MB 5.86MB 0% 1.70MB 3.36MB 0% 2.49MB 2.48MB 15.0MB 23.0MB 10752 12486
23% 28% 8314 8314 0 0 8314 4.47MB 6.25MB 0% 1.83MB 3.61MB 0% 2.64MB 2.64MB 5.85MB 1022KB 10857 14271
20% 25% 8258 8257 0 0 8256 4.34MB 6.15MB 0% 1.76MB 3.57MB 0% 2.58MB 2.58MB 10.44MB 23.8KB 10233 13177
21% 25% 7831 7820 10 0 7831 3.93MB 5.23MB 0% 1.73MB 3.05MB 0% 2.19MB 2.20MB 10.8MB 780KB 9500 13304
18% 23% 7665 7665 0 0 7664 3.66MB 5.32MB 0% 1.65MB 3.31MB 0% 2.01MB 2.01MB 6.79MB 19.9KB 10425 12094
24% 25% 8258 8258 0 0 8258 4.30MB 6.08MB 0% 1.87MB 3.65MB 0% 2.43MB 2.43MB 15.7MB 23.4MB 10744 13966
22% 25% 8178 8178 0 0 8183 4.29MB 6.07MB 0% 1.75MB 3.52MB 0% 2.54MB 2.55MB 5.87MB 2.34MB 9990 12995
21% 25% 7921 7921 0 0 7921 4.35MB 5.74MB 0% 1.77MB 3.16MB 0% 2.58MB 2.58MB 12.1MB 11.9KB 9557 13344
sntap9-tme-8040: cluster:cluster: 12/10/2019 11:07:05

```

You can view a [video of this capture](#).

We also captured the time of completion for each job.

```

This was the job on the FlexVol before it was converted:
# python file-create.py /fvconvert/files
Starting overall work: 2019-12-09 10:32:21.966337
End overall work: 2019-12-09 12:11:15.990707
total time: 5934.024611

```

Converting the FlexVol volume to a FlexGroup volume (with added member volumes) saved some time:

```

# python file-create.py /fvconvert/files2
Starting overall work: 2019-12-10 11:02:28.621532
End overall work: 2019-12-10 12:22:48.523772
total time: 4819.95753193

```

That's savings of about 1100 seconds, or 18 minutes—which saved us around 20% of the total time of completion.

The following output shows the file distribution before the script run. Note that the first member volume has the most files, because it was previously the FlexVol volume.

```

cluster::*> volume show -vserver DEMO -volume fvconvert* -fields files,files-used
vserver volume files files-used
-----
DEMO fvconvert__0001 2040109451 502848737
DEMO fvconvert__0002 2040109451 12747
DEMO fvconvert__0003 2040109451 12749
DEMO fvconvert__0004 2040109451 12751

```

17. Examples
Sample FlexVol to FlexGroup conversion – 500 million files

At the end of the job, we can see that the files spread out fairly evenly.

```
cluster::*> volume show -vserver DEMO -volume fvconvert* -fields files,files-used
vserver volume files files-used
-----
DEMO fvconvert__0001 2040109451 506770209
DEMO fvconvert__0002 2040109451 3345330
DEMO fvconvert__0003 2040109451 3345330
DEMO fvconvert__0004 2040109451 3345319
DEMO fvconvert__0005 2040109451 3331657
DEMO fvconvert__0006 2040109451 3331635
DEMO fvconvert__0007 2040109451 3331657
DEMO fvconvert__0008 2040109451 3331657
```

We ran the script again on the newly converted FlexGroup volume. This time, we wanted to see how much faster the job ran and how the files distributed on the emptier FlexVol member volumes.

Remember, when we started out, the newer member volumes all had less than 1% of files used (3.3 million of two billion possible files). The member volume that was converted from a FlexVol volume was using 25% of the total files (500 million of two billion).

After the job ran, we saw a file count delta of about 3.2 million on the original member volume and of about 3.58 million on all the other members. We're still balancing across all member volumes, but favoring the less full ones for new file and folder creations.

```
cluster::*> volume show -vserver DEMO -volume fvconvert* -fields files,files-used
vserver volume files files-used
-----
DEMO fvconvert__0001 2040109451 509958440
DEMO fvconvert__0002 2040109451 6808792
DEMO fvconvert__0003 2040109451 6809225
DEMO fvconvert__0004 2040109451 6806843
DEMO fvconvert__0005 2040109451 6798959
DEMO fvconvert__0006 2040109451 6800054
DEMO fvconvert__0007 2040109451 6849375
DEMO fvconvert__0008 2040109451 6801600
```

With the new FlexGroup volume, converted from a FlexVol volume, our job time dropped from 5900 seconds to 4656 seconds. We were also able to push two times the amount of IOPS:

```
# python file-create.py /fvconvert/files3
Starting overall work: 2019-12-10 13:14:26.816860
End overall work: 2019-12-10 14:32:03.565705
total time: 4656.76723099
```

Figure 91 Sample statistics of conversion process – two times performance

| cpu | cpu | total | | | fcache | total | total data | data | data cluster | cluster | cluster | disk | disk | pkts | pkts | | | |
|-----|------|-------|---------|---------|--------|----------|------------|--------|--------------|---------|---------|------|--------|--------|--------|--------|-------|-------|
| avg | busy | ops | nfs-ops | dfs-ops | ops | spin-ops | recv | sent | busy | recv | sent | read | write | rcv | sent | | | |
| 26% | 29% | 10403 | 10403 | 0 | 0 | 10403 | 7.55MB | 8.31MB | 0% | 2.25MB | 3.00MB | 0% | 5.30MB | 5.31MB | 4.20MB | 21.6MB | 17915 | 23123 |
| 27% | 31% | 9262 | 9262 | 0 | 0 | 9262 | 6.93MB | 7.82MB | 0% | 2.05MB | 2.56MB | 0% | 4.87MB | 5.26MB | 5.67MB | 35.3MB | 16487 | 20524 |
| 25% | 29% | 8773 | 8773 | 0 | 0 | 8773 | 7.47MB | 7.68MB | 0% | 2.39MB | 3.01MB | 0% | 5.08MB | 4.67MB | 851KB | 7.92KB | 18667 | 22978 |
| 18% | 22% | 6592 | 6592 | 0 | 0 | 6591 | 4.21MB | 4.57MB | 0% | 1021KB | 1.34MB | 0% | 3.21MB | 3.23MB | 1.33MB | 23.9KB | 8963 | 10892 |
| 20% | 21% | 9400 | 9400 | 0 | 0 | 9399 | 6.72MB | 7.32MB | 0% | 2.26MB | 2.87MB | 0% | 4.46MB | 4.45MB | 1.05MB | 8.22MB | 18350 | 21814 |
| 25% | 26% | 12010 | 12010 | 0 | 0 | 12010 | 7.25MB | 8.00MB | 0% | 2.18MB | 2.93MB | 0% | 5.07MB | 5.07MB | 4.67MB | 17.2MB | 17918 | 22028 |
| 22% | 23% | 11266 | 11266 | 0 | 0 | 11266 | 8.23MB | 9.06MB | 0% | 2.49MB | 3.31MB | 0% | 5.73MB | 5.74MB | 5.11MB | 12.1MB | 20029 | 25981 |
| 25% | 26% | 12445 | 12445 | 0 | 0 | 12445 | 11.0MB | 12.0MB | 0% | 3.82MB | 4.84MB | 0% | 7.18MB | 7.12MB | 915KB | 10.3MB | 27291 | 35571 |
| 25% | 26% | 12253 | 12253 | 0 | 0 | 12253 | 8.04MB | 8.77MB | 0% | 2.53MB | 3.26MB | 0% | 5.51MB | 5.52MB | 976KB | 11.7KB | 20328 | 25953 |
| 29% | 34% | 12699 | 12699 | 0 | 0 | 12699 | 8.42MB | 9.29MB | 0% | 2.65MB | 3.52MB | 0% | 5.77MB | 5.77MB | 1.41MB | 3.73MB | 20937 | 27166 |
| 28% | 30% | 12599 | 12599 | 0 | 0 | 12599 | 8.34MB | 9.09MB | 0% | 2.62MB | 3.38MB | 0% | 5.71MB | 5.71MB | 4.20MB | 21.5MB | 20958 | 26748 |
| 30% | 34% | 13929 | 13919 | 9 | 0 | 13924 | 9.41MB | 10.5MB | 0% | 3.00MB | 4.11MB | 0% | 6.41MB | 6.40MB | 3.29MB | 65.7KB | 23395 | 30206 |
| 26% | 28% | 14499 | 14499 | 0 | 0 | 14499 | 9.68MB | 10.6MB | 0% | 3.08MB | 4.00MB | 0% | 6.80MB | 6.60MB | 3.77MB | 25.3MB | 24627 | 31571 |
| 29% | 34% | 13231 | 13231 | 0 | 0 | 13230 | 8.44MB | 9.46MB | 0% | 2.75MB | 3.78MB | 0% | 5.65MB | 5.68MB | 1.77MB | 11.5KB | 21565 | 27726 |
| 26% | 28% | 13505 | 13502 | 2 | 0 | 13503 | 9.10MB | 10.3MB | 0% | 3.01MB | 4.19MB | 0% | 6.05MB | 6.09MB | 2.02MB | 3.45MB | 24130 | 30584 |
| 25% | 29% | 13553 | 13553 | 0 | 0 | 13553 | 8.94MB | 9.64MB | 0% | 2.92MB | 3.73MB | 0% | 6.12MB | 6.13MB | 4.62MB | 23.1MB | 22491 | 28837 |

As you can see, there's an imbalance of files and data in these member volumes (much more in the original FlexVol volume), but performance is still much better than the previous FlexVol performance because work across multiple nodes is more efficient. That's the power of the FlexGroup volume.

Event management system examples

Inode-related EMS examples

Message Name: callhome.no.inodes
Severity: ERROR

Corrective Action: Modify the volume's maxfiles (maximum number of files) to increase the inodes on the affected volume. If you need assistance, contact Fujitsu technical support.

Description: This message occurs when a volume is out of inodes, which refer to individual files, other types of files, and directories. If your system is configured to do so, it generates and transmits an AutoSupport (or 'call home') message to Fujitsu technical support and to the configured destinations. Successful delivery of an AutoSupport message significantly improves problem determination and resolution.

Message Name: fg.inodes.member.nearlyFull
Severity: ALERT

Corrective Action: Adding capacity to the FlexGroup by using the "volume modify -files +X" command is the best way to solve this problem. Alternatively, deleting files from the FlexGroup might work, although it can be difficult to determine which files have landed on which constituent.

Description: This message occurs when a constituent within a FlexGroup is almost out of inodes. This constituent will receive far fewer new create requests than average, which might impact the FlexGroup's overall performance, because those requests are routed to constituents with more inodes.

Message Name: fg.inodes.member.full
Severity: ALERT

Corrective Action: Adding capacity to the FlexGroup by using the "volume modify -files +X" command is the best way to solve this problem. Alternatively, deleting files from the FlexGroup may work, but it is difficult to determine which files have landed on which constituent.

Description: This message occurs when a constituent with a FlexGroup has run out of inodes. New files cannot be created on this constituent. This might lead to an overall imbalanced distribution of content across the FlexGroup.

Message Name: fg.inodes.member.alloK
Severity: NOTICE

Corrective Action: (NONE)

Description: This message occurs when conditions that led to previous "fg.inodes.member.nearlyFull" and "fg.inodes.member.full" events no longer apply for any constituent in this FlexGroup. All constituents within this FlexGroup have sufficient inodes for normal operation.

Example of maxdirsize message

Message Name: wafl.dir.size.max
Severity: ERROR

Corrective Action: Use the "volume file show-inode" command with the file ID and volume name information to find the file path. Reduce the number of files in the directory. If not possible, use the (privilege:advanced) option "volume modify - volume vol_name -maxdir-size new_value" to increase the maximum number of files per directory. However, doing so could impact system performance. If you need to increase the maximum directory size, work with technical support.

Description: This message occurs after a directory has reached its maximum directory size (maxdirsize) limit.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: wafl.dir.size.max.warning
Severity: ERROR

Corrective Action: Use the "volume file show-inode" command with the file ID and volume name information to find the file path. Reduce the number of files in the directory. If not possible, use the (privilege:advanced) option "volume modify - volume vol_name -maxdir-size new_value" to increase the maximum number of files per directory. However, doing so could impact system performance. If you need to increase the maximum directory size, work with technical support.

Description: This message occurs when a directory has reached or surpassed 90% of its current maximum directory size (maxdirsize) limit, and the current maxdirsize is less than the default maxdirsize, which is 1% of total system memory.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: wafl.dir.size.warning
Severity: ERROR

Corrective Action: Use the "volume file show-inode" command with the file ID and volume name information to find the file path. Reduce the number of files in the directory. If not possible, use the (privilege:advanced) option "volume modify - volume vol_name -maxdir-size new_value" to increase the maximum number of files per directory. However, doing so could impact system performance. If you need to increase the maximum directory size, work with technical support.

Description: This message occurs when a directory surpasses 90% of its current maximum directory size (maxdirsize) limit.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Examples of capacity-related event management system messages

Message Name: monitor.volume.full
Severity: DEBUG
Corrective Action: (NONE)

Description: This message occurs when one or more file systems are full, typically indicating at least 98% full. This event is accompanied by global health monitoring messages for the customer. The space usage is computed based on the active file system size and is computed by subtracting the value of the "Snapshot Reserve" field from the value of the "Used" field of the "volume show- space" command. The volume/aggregate can be over 100% full due to space used or reserved by metadata. A value greater than 100% might cause Snapshot(tm) copy space to become unavailable or cause the volume to become logically overallocated. See the "vol.log.overalloc" EMS message for more information.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: monitor.volume.nearlyFull
Severity: ALERT

Corrective Action: Create space by increasing the volume or aggregate sizes, or by deleting data or deleting Snapshot(R) copies. To increase a volume's size, use the "volume size" command. To delete a volume's Snapshot(R) copies, use the "volume snapshot delete" command. To increase an aggregate's size, add disks by using the "storage aggregate add-disks" command. Aggregate Snapshot(R) copies are deleted automatically when the aggregate is full.

Description: This message occurs when one or more file systems are nearly full, typically indicating at least 95% full. This event is accompanied by global health monitoring messages for the customer. The space usage is computed based on the active file system size and is computed by subtracting the value of the "Snapshot Reserve" field from the value of the "Used" field of the "volume show-space" command.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: monitor.volume.ok
Severity: DEBUG
Corrective Action: (UNKNOWN)

Description: The previously-reported volume full condition is fixed. * We log this event, as well as the other monitor.volume events, at LOG_DEBUG level to avoid spamming the messages file with events which are already being reported as part of the global health messages.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

17. Examples
Event management system examples

Message Name: monitor.volumes.one.ok
Severity: DEBUG
Corrective Action: (NONE)

Description: This message occurs when one file system that was nearly full (usually this means $\geq 95\%$ full) is now OK. This event and other "monitor.volume" events are logged at LOG_DEBUG level to avoid spamming the messages file with events that are already being reported as part of the global health messages. The space usage is computed based on the active file system size and is computed by subtracting the value of the "Snapshot Reserve" field from the value of the "Used" field of the "volume show-space" command.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: vol.log.overalloc
Severity: ALERT

Corrective Action: Create space by increasing the volume or aggregate size, deleting data, deleting Snapshot(R) copies, or changing the provisioning from thick to thin. To increase a volume's size, use the "volume size" command. To delete a volume's Snapshot(R) copies, use the "volume snapshot delete" command. To change provisioning in a volume, reserved files can be unreserved by using the "volume file reservation" command. To increase an aggregate's size, add disks by using the "storage aggregate add-disks" command. Aggregate Snapshot(R) copies are deleted automatically when the aggregate is full. To change provisioning of a volume in an aggregate, change the volume guarantee from "volume" to "none" by using the "space-guarantee" field of the "volume modify" command.

Description: This message occurs when the volume or aggregate allocates more space than it can honor by way of reservations, or the aggregate has allocated more space than it can honor by way of guarantees. If the reserved or guaranteed space is consumed, there is insufficient physical space, which can cause the volume or aggregate to be taken offline.

Supports SNMP trap: true
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

Message Name: fg.member.elastic.sizing

Severity: NOTICE
Corrective Action: (NONE)

Description: This message occurs when a FlexGroup constituent undergoes elastic sizing, either to restore balance among constituents or to resize constituents to accommodate space needs.

Supports SNMP trap: false
Destinations: -
Number of Drops Between Transmissions: 0
Dropping Interval (Seconds) Between Transmissions: 0

18. Command Examples

FlexGroup capacity commands

```
cluster::*> aggr show-space -instance -aggregate aggr1_node1

          Aggregate Name: aggr1_node1
          Volume Footprints: 2.05TB
          Volume Footprints Percent: 26%
Total Space for Snapshot Copies in Bytes: 0B
Space Reserved for Snapshot Copies: 0%
          Aggregate Metadata: 15.20MB
Aggregate Metadata Percent: 0%
          Total Used: 2.05TB
          Total Used Percent: 26%
          Size: 7.86TB
          Snapshot Reserve Unusable: -
Snapshot Reserve Unusable Percent: -
          Total Physical Used Size: 143.7GB
          Physical Used Percentage: 2%

          Aggregate Name: aggr1_node2
          Volume Footprints: 2.02TB
          Volume Footprints Percent: 26%
Total Space for Snapshot Copies in Bytes: 0B
Space Reserved for Snapshot Copies: 0%
          Aggregate Metadata: 8.63MB
Aggregate Metadata Percent: 0%
          Total Used: 2.02TB
          Total Used Percent: 26%
          Size: 7.86TB
          Snapshot Reserve Unusable: -
Snapshot Reserve Unusable Percent: -
          Total Physical Used Size: 69.71GB
          Physical Used Percentage: 1%
2 entries were displayed.
```

18. Command Examples
FlexGroup capacity commands

```
cluster::*> volume show-space -vserver SVM -volume flexgroup__*
Vserver : SVM
Volume : flexgroup__0001
Feature                                Used    Used%
-----
User Data                              57.06MB  0%
Filesystem Metadata                     3.51MB  0%
Inodes                                  87.26MB  0%
Snapshot Reserve                        512GB   5%
Deduplication                           12KB    0%
Performance Metadata                    48KB    0%

Total Used                              512.1GB  5%
Total Physical Used                      148.3MB  0%

Vserver : SVM
Volume : flexgroup__0002
Feature                                Used    Used%
-----
User Data                              57.03MB  0%
Filesystem Metadata                     4.66MB  0%
Inodes                                  83.66MB  0%
Snapshot Reserve                        512GB   5%
Deduplication                           20KB    0%
Performance Metadata                    44KB    0%

Total Used                              512.1GB  5%
Total Physical Used                      145.7MB  0%

Vserver : SVM
Volume : flexgroup__0003
Feature                                Used    Used%
-----
User Data                              57.02MB  0%
Filesystem Metadata                     3.66MB  0%
Inodes                                  84.55MB  0%
Snapshot Reserve                        512GB   5%
Deduplication                           12KB    0%
Performance Metadata                    44KB    0%

Total Used                              512.1GB  5%
Total Physical Used                      145.6MB  0%

Vserver : SVM
Volume : flexgroup__0004
Feature                                Used    Used%
-----
User Data                              57.19MB  0%
Filesystem Metadata                     8.93MB  0%
Inodes                                  82.09MB  0%
Snapshot Reserve                        512GB   5%
Deduplication                           12KB    0%
Performance Metadata                    44KB    0%

Total Used                              512.1GB  5%
Total Physical Used                      148.5MB  0%
```


18. Command Examples
FlexGroup capacity commands

| | | |
|--------------------------|---------|-------|
| Vserver : SVM | | |
| Volume : flexgroup__0005 | | |
| Feature | Used | Used% |
| ----- | ----- | ----- |
| User Data | 3.99GB | 0% |
| Filesystem Metadata | 4.88MB | 0% |
| Inodes | 83.54MB | 0% |
| Snapshot Reserve | 512GB | 5% |
| Deduplication | 12KB | 0% |
| Performance Metadata | 52KB | 0% |
| | | |
| Total Used | 516.1GB | 5% |
| Total Physical Used | 4.08GB | 0% |
| | | |
| Vserver : SVM | | |
| Volume : flexgroup__0006 | | |
| Feature | Used | Used% |
| ----- | ----- | ----- |
| User Data | 57.04MB | 0% |
| Filesystem Metadata | 3.50MB | 0% |
| Inodes | 87.26MB | 0% |
| Snapshot Reserve | 512GB | 5% |
| Deduplication | 12KB | 0% |
| Performance Metadata | 44KB | 0% |
| | | |
| Total Used | 512.1GB | 5% |
| Total Physical Used | 148.2MB | 0% |
| | | |
| Vserver : SVM | | |
| Volume : flexgroup__0007 | | |
| Feature | Used | Used% |
| ----- | ----- | ----- |
| User Data | 57.02MB | 0% |
| Filesystem Metadata | 3.50MB | 0% |
| Inodes | 85.03MB | 0% |
| Snapshot Reserve | 512GB | 5% |
| Deduplication | 12KB | 0% |
| Performance Metadata | 44KB | 0% |
| | | |
| Total Used | 512.1GB | 5% |
| Total Physical Used | 145.9MB | 0% |
| | | |
| Vserver : SVM | | |
| Volume : flexgroup__0008 | | |
| Feature | Used | Used% |
| ----- | ----- | ----- |
| User Data | 57.03MB | 0% |
| Filesystem Metadata | 3.52MB | 0% |
| Inodes | 86.12MB | 0% |
| Snapshot Reserve | 512GB | 5% |
| Deduplication | 12KB | 0% |
| Performance Metadata | 44KB | 0% |
| | | |
| Total Used | 512.1GB | 5% |
| Total Physical Used | 147.0MB | 0% |

18. Command Examples
FlexGroup capacity commands

```

cluster::> vol show -is-constituent true -volume flexgroup_*
Vserver  Volume          Aggregate      State   Type  Size      Available  Used%
-----  -
SVM      flexgroup__0001
          aggr1_node1    online        RW      10TB   5.05TB    49%
SVM      flexgroup__0002
          aggr1_node2    online        RW      10TB   5.08TB    49%
SVM      flexgroup__0003
          aggr1_node1    online        RW      10TB   5.05TB    49%
SVM      flexgroup__0004
          aggr1_node2    online        RW      10TB   5.08TB    49%
SVM      flexgroup__0005
          aggr1_node1    online        RW      10TB   5.05TB    49%
SVM      flexgroup__0006
          aggr1_node2    online        RW      10TB   5.08TB    49%
SVM      flexgroup__0007
          aggr1_node1    online        RW      10TB   5.05TB    49%
SVM      flexgroup__0008
          aggr1_node2    online        RW      10TB   5.08TB    49%
8 entries were displayed.

cluster::> storage aggregate show -aggregate aggr1* -fields usedsize,size,percent-
used -sort-by percent-used
aggregate  percent-used  size  usedsize
-----
aggr1_node1 26%          7.86TB 2.05TB
aggr1_node2 26%          7.86TB 2.02TB
2 entries were displayed.

```

Example of statistics show-periodic command for entire cluster

```

cluster::*> statistics show-periodic
cluster: cluster.cluster: 11/30/2016 11:49:46
cpu cpu          total fcache total  total data  data  data
cluster cluster  cluster disk  disk  pkts  pkts
avg busy ops nfs-ops cifs-ops ops spin-ops  recv  sent busy  recv  sent
busy  recv  sent  read  write  recv  sent
-----
5% 5% 0 0 0 0 0 65.3KB 64.4KB 0% 2.22KB 1.13KB
0% 62.7KB 63.2KB 489KB 407KB 91 83
5% 5% 0 0 0 0 62.5KB 61.6KB 0% 1.28KB 767B
0% 61.0KB 60.9KB 23.8KB 23.8KB 64 60
4% 5% 0 0 0 0 62.3KB 61.3KB 0% 1.43KB 708B
0% 60.7KB 60.7KB 15.8KB 15.8KB
cluster: cluster.cluster: 11/30/2016 11:49:53
cpu cpu          total fcache total  total data  data  data
cluster cluster  cluster disk  disk  pkts  pkts
avg busy ops nfs-ops cifs-ops ops spin-ops  recv  sent busy  recv  sent
busy  recv  sent  read  write  recv  sent
-----
Minimums:
4% 5% 0 0 0 0 62.3KB 61.3KB 0% 1.28KB 708B
0% 60.7KB 60.7KB 15.8KB 15.8KB 64 58
4% 5% 0 0 0 0 63.4KB 62.4KB 0% 1.64KB 877B
0% 61.5KB 61.6KB 176KB 149KB 74 67
Maximums:
5% 5% 0 0 0 0 65.3KB 64.4KB 0% 2.22KB 1.13KB
0% 62.7KB 63.2KB 489KB 407KB 91 83
    
```

Real-time SVM-level statistics show-periodic for NFSv3 read and write operations

```

cluster::*> statistics show-periodic -instance SVM -interval 2 -iterations 0 -summary
true -vserver SVM -object nfsv3 -counter nfsv3_ops|nfsv3_read_ops|nfsv3_write_ops
cluster: nfsv3.SVM: 11/30/2016 13:29:57
      nfsv3      nfsv3
nfsv3  read  write  Complete  Number of
ops    ops   ops  Aggregation Constituents
-----
2360   0    697   Yes      16
2245   0    652   Yes      16
2126   0    629   Yes      16
cluster: nfsv3.SVM: 11/30/2016 13:30:04
      nfsv3      nfsv3
nfsv3  read  write  Complete  Number of
ops    ops   ops  Aggregation Constituents
-----
Minimums:
2126   0    629   -        -
Averages for 3 samples:
2243   0    659   -        -
Maximums:
2360   0    697   -        -
    
```

Real-time FlexGroup local and remote statistics

```
cluster::*> statistics show-periodic -instance 0 -interval 2 -iterations 0 -summary
true -object flexgroup -counter
cat1_tld_local|cat1_tld_remote|cat2_hld_local|cat2_hld_remote|cat3_dir_lo-
cal|cat3_dir_remote|cat4
_fil_local|cat4_fil_remote
cluster: flexgroup.0: 11/30/2016 13:34:55
cat1   cat1   cat2   cat2   cat3   cat3   cat4   cat4
tld    tld    hld    hld    dir    dir    fil    fil    Complete   Number of
local  remote local  remote local  remote local  remote Aggregation Constituents
-----
      1      0      17     113      0      0     619      0      n/a        n/a
      0      1      17     114      0      0     654      0      n/a        n/a
      0      2      17     112      0      0     647      0      n/a        n/a
cluster: flexgroup.0: 11/30/2016 13:35:02
cat1   cat1   cat2   cat2   cat3   cat3   cat4   cat4
tld    tld    hld    hld    dir    dir    fil    fil    Complete   Number of
local  remote local  remote local  remote local  remote Aggregation Constituents
-----
Minimums:
      0      0      17     112      0      0     619      0      -          -
Averages for 3 samples:
      0      1      17     113      0      0     640      0      -          -
Maximums:
      1      2      17     114      0      0     654      0      -          -
```

Example of creating a FlexGroup volume and specifying fewer member volumes than the default value

This command creates a 10TB FlexGroup volume with two 5TB member volumes across two nodes.

```
cluster:::> volume create -vserver DEMO -volume flexgroup -aggr-list
aggr1_node1,aggr1_node2 -aggr-list-multiplier 1 -junction-path /flexgroup -size 10t

Warning: The FlexGroup "flexgroup" will be created with the following number of con-
stituents of size 5TB: 2.

Do you want to continue? {y|n}: y
```

Note

The `-aggr-list` flag must be used to make sure that the volume is a FlexGroup volume.

Sample REST API for creating a FlexGroup volume

The following REST API example creates a 2TB, eight-member thin-provisioned FlexGroup volume across a single aggregate.

```
{
  "aggregates": [
    {
      "name": "aggr1_node1"
    }
  ],
  "constituents_per_aggregate": 8,
  "efficiency": {
    "compaction": "inline",
    "compression": "inline",
    "cross_volume_dedupe": "inline",
    "dedupe": "inline"
  },
  "guarantee": {
    "type": "none"
  },
  "name": "RESTAPI_FG",
  "nas": {
    { "export_policy":
      {
        "id": 42949672961,
        "name": "default"
      }
    },
    "gid": 0,
    "path": "/RESTAPI_FG",
    "security_style": "unix",
    "uid": 0,
    "unix_permissions": 755
  },
  "size": "2T",
  "style": "flexgroup",
  "svm": {
    "name": "DEMO",
    "uuid": "7e3cc08e-d9b3-11e6-85e2-00a0986b1210"
  }
}
```

18. Command Examples

Sample REST API for creating a FlexGroup volume

This is what the FlexGroup looks like after it is created:

```
cluster::*> vol show -vserver DEMO -volume REST*
Vserver  Volume          Aggregate      State    Type  Size      Available  Used%
-----  -
DEMO     RESTAPI_FG      -              online   RW    2TB       1.90TB    0%
DEMO     RESTAPI_FG_0001 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0002 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0003 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0004 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0005 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0006 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0007 aggr1_node1    online   RW    256GB     243.1GB   0%
DEMO     RESTAPI_FG_0008 aggr1_node1    online   RW    256GB     243.1GB   0%
9 entries were displayed.
```

To include more than one aggregate in the list, use this REST API as an example:

```
{
  "aggregates": [
    { "name": "aggr1_node1" }, { "name": "aggr1_node2" }
  ],
  "efficiency": {
    "compaction": "inline",
    "compression": "inline",
    "cross_volume_dedupe": "inline",
    "dedupe": "inline"
  },
  "guarantee": {
    "type": "none"
  },
  "name": "RESTAPI_FG3",
  "nas": {
    "export_policy": {
      {
        "id": 42949672961,
        "name": "default"
      }
    },
    "gid": 0,
    "path": "/RESTAPI_FG3",
    "security_style": "unix",
    "uid": 0,
    "unix_permissions": 755
  },
  "size": "2T",
  "style": "flexgroup",
  "svm": {
    "name": "DEMO",
    "uuid": "7e3cc08e-d9b3-11e6-85e2-00a0986b1210"
  }
}
```

18. Command Examples
Sample REST API for creating a FlexGroup volume

This is how it looks:

```
cluster::*> vol show -vserver DEMO -volume *FG3*
```

| Vserver | Volume | Aggregate | State | Type | Size | Available | Used% |
|---------|------------------|-------------|--------|------|-------|-----------|-------|
| DEMO | RESTAPI_FG3 | - | online | RW | 2TB | 1.90TB | 0% |
| DEMO | RESTAPI_FG3_0001 | aggr1_node1 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0002 | aggr1_node2 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0003 | aggr1_node1 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0004 | aggr1_node2 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0005 | aggr1_node1 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0006 | aggr1_node2 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0007 | aggr1_node1 | online | RW | 256GB | 243.1GB | 0% |
| DEMO | RESTAPI_FG3_0008 | aggr1_node2 | online | RW | 256GB | 243.1GB | 0% |

9 entries were displayed.

This REST API creates a four-member FlexGroup volume y using the `style` option and does not specify the `constituents_per_aggregate` option.

```
{
  "aggregates": [
    {
      "name": "aggr1_node1"
    }
  ],
  "efficiency": {
    "compaction": "inline",
    "compression": "inline",
    "cross_volume_dedupe": "inline",
    "dedupe": "inline"
  },
  "guarantee": {
    "type": "none"
  },
  "name": "RESTAPI_FG2",
  "nas": {
    "export_policy": {
      "id": 42949672961,
      "name": "default"
    },
    "gid": 0,
    "path": "/RESTAPI_FG2",
    "security_style": "unix",
    "uid": 0,
    "unix_permissions": 755
  },
  "size": "2T",
  "style": "flexgroup",
  "svm": {
    "name": "DEMO",
    "uuid": "7e3cc08e-d9b3-11e6-85e2-00a0986b1210"
  }
}
```

18. Command Examples
Sample REST API for creating a FlexGroup volume

And this is the resulting FlexGroup:

```
cluster::*> vol show -vserver DEMO -volume RESTAPI_FG2*
```

| Vserver | Volume | Aggregate | State | Type | Size | Available | Used% |
|---------|-------------------|-------------|--------|------|-------|-----------|-------|
| DEMO | RESTAPI_FG2 | - | online | RW | 2TB | 1.90TB | 0% |
| DEMO | RESTAPI_FG2__0001 | aggr1_node1 | online | RW | 512GB | 486.3GB | 0% |
| DEMO | RESTAPI_FG2__0002 | aggr1_node1 | online | RW | 512GB | 486.3GB | 0% |
| DEMO | RESTAPI_FG2__0003 | aggr1_node1 | online | RW | 512GB | 486.3GB | 0% |
| DEMO | RESTAPI_FG2__0004 | aggr1_node1 | online | RW | 512GB | 486.3GB | 0% |

5 entries were displayed.

Example of increasing a FlexGroup volume's size

```

cluster::*> volume show -vserver SVM -volume flexgroup*
SVM      flexgroup      -          online    RW        70.20TB   10.14TB   85%
SVM      flexgroup__0001
          aggr1_node1  online    RW        10TB      5.06TB   49%
SVM      flexgroup__0002
          aggr1_node2  online    RW        10TB      5.08TB   49%
SVM      flexgroup__0003
          aggr1_node1  online    RW        10TB      5.06TB   49%
SVM      flexgroup__0004
          aggr1_node2  online    RW        10TB      5.08TB   49%
SVM      flexgroup__0005
          aggr1_node1  online    RW        10TB      5.06TB   49%
SVM      flexgroup__0006
          aggr1_node2  online    RW        10TB      5.08TB   49%
SVM      flexgroup__0007
          aggr1_node1  online    RW        10TB      5.06TB   49%
SVM      flexgroup__0008
          aggr1_node2  online    RW        10TB      5.08TB   49%

cluster::*> vol size -vserver SVM -volume flexgroup -new-size 100t vol size: Volume
"SVM:flexgroup" size set to 100t.

cluster::*> volume show -vserver SVM -volume flexgroup*
Vserver  Volume          Aggregate    State     Type Size          Available  Used%
-----
SVM      flexgroup      -          online    RW        100TB   10.14TB   89%
SVM      flexgroup__0001
          aggr1_node1  online    RW        12.50TB  5.06TB   59%
SVM      flexgroup__0002
          aggr1_node2  online    RW        12.50TB  5.08TB   59%
SVM      flexgroup__0003
          aggr1_node1  online    RW        12.50TB  5.06TB   59%
SVM      flexgroup__0004
          aggr1_node2  online    RW        12.50TB  5.08TB   59%
SVM      flexgroup__0005
          aggr1_node1  online    RW        12.50TB  5.06TB   59%
SVM      flexgroup__0006
          aggr1_node2  online    RW        12.50TB  5.08TB   59%
SVM      flexgroup__0007
          aggr1_node1  online    RW        12.50TB  5.06TB   59%
SVM      flexgroup__0008
          aggr1_node2  online    RW        12.50TB  5.08TB   59%
  
```

Example of expanding a FlexGroup volume

```

cluster::*> volume show -vserver SVM -volume flexgroup4*
Vserver   Volume           Aggregate      State      Type Size      Available  Used%
-----
SVM       flexgroup4TB -    online      RW        4TB        3.78TB    5%
SVM       flexgroup4TB__0001
          aggr1_node1    online      RW        512GB     485.5GB   5%
SVM       flexgroup4TB__0002
          aggr1_node2    online      RW        512GB     481.2GB   6%
SVM       flexgroup4TB__0003
          aggr1_node1    online      RW        512GB     481.5GB   5%
SVM       flexgroup4TB__0004
          aggr1_node2    online      RW        512GB     485.5GB   5%
SVM       flexgroup4TB__0005
          aggr1_node1    online      RW        512GB     485.5GB   5%
SVM       flexgroup4TB__0006
          aggr1_node2    online      RW        512GB     485.5GB   5%
SVM       flexgroup4TB__0007
          aggr1_node1    online      RW        512GB     485.5GB   5%
SVM       flexgroup4TB__0008
          aggr1_node2    online      RW        512GB     485.5GB   5%

cluster::*> volume expand -vserver SVM -volume flexgroup4TB -aggr-list aggr1_node1,aggr1_node2 -
aggr-list-multiplier 4

cluster::*> volume show -vserver SVM -volume flexgroup4*
Vserver   Volume           Aggregate      State      Type Size      Available  Used%
-----
SVM       flexgroup4TB -    online      RW        8TB        7.78TB    1%
SVM       flexgroup4TB__0001
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0002
          aggr1_node2    online      RW        512GB     481.2GB   1%
SVM       flexgroup4TB__0003
          aggr1_node1    online      RW        512GB     481.5GB   1%
SVM       flexgroup4TB__0004
          aggr1_node2    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0005
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0006
          aggr1_node2    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0007
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0008
          aggr1_node2    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0009
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0010
          aggr1_node2    online      RW        512GB     481.2GB   1%
SVM       flexgroup4TB__0011
          aggr1_node1    online      RW        512GB     481.5GB   1%
SVM       flexgroup4TB__0012
          aggr1_node2    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0013
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0014
          aggr1_node2    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0015
          aggr1_node1    online      RW        512GB     485.5GB   1%
SVM       flexgroup4TB__0016
          aggr1_node2    online      RW        512GB     485.5GB   1%
  
```

Other command-line examples

Creating a FlexGroup volume by using flexgroup deploy

```
cluster::> flexgroup deploy -size 20PB -space-guarantee volume -vserver SVM -volume flexgroup
```

Creating a FlexGroup volume across multiple nodes by using volume create

```
cluster::> volume create -vserver SVM -volume flexgroup -aggr-list aggr1_node1,aggr1_node2 -  
policy default -security-style unix -size 20PB -space-guarantee none -junction-path /flexgroup
```

Modifying the FlexGroup Snapshot policy

```
cluster::> volume modify -vserver SVM -volume flexgroup -snapshot-policy [policyname|none]
```

Applying storage QoS

```
cluster::> volume modify -vserver DEMO -volume flexgroup -qos-policy-group FlexGroupQoS
```

Applying volume autogrow

```
cluster::> volume autosize -vserver DEMO -volume Tech_ONTAP -mode grow -maximum-size 20t  
-grow- threshold-percent 80  
  
cluster::> volume autosize -vserver DEMO -volume Tech_ONTAP Volume autosize is currently ON for  
volume "DEMO:Tech_ONTAP".  
The volume is set to grow to a maximum of 20t when the volume-used space is above 80%.  
Volume autosize for volume 'DEMO:Tech_ONTAP' is currently in mode grow.
```

FUJITSU Storage
ETERNUS AX series All-Flash Arrays,
ETERNUS HX series Hybrid Arrays
ONTAP FlexGroup Volumes
Best Practices and Implementation Guide

P3AG-6202-01ENZO

Date of issuance: September 2021
Issuance responsibility: FUJITSU LIMITED

- The content of this manual is subject to change without notice.
- This manual was prepared with the utmost attention to detail. However, Fujitsu shall assume no responsibility for any operational problems as the result of errors, omissions, or the use of information in this manual.
- Fujitsu assumes no liability for damages to third party copyrights or other rights arising from the use of any information in this manual.
- The content of this manual may not be reproduced or distributed in part or in its entirety without prior permission from Fujitsu.


FUJITSU