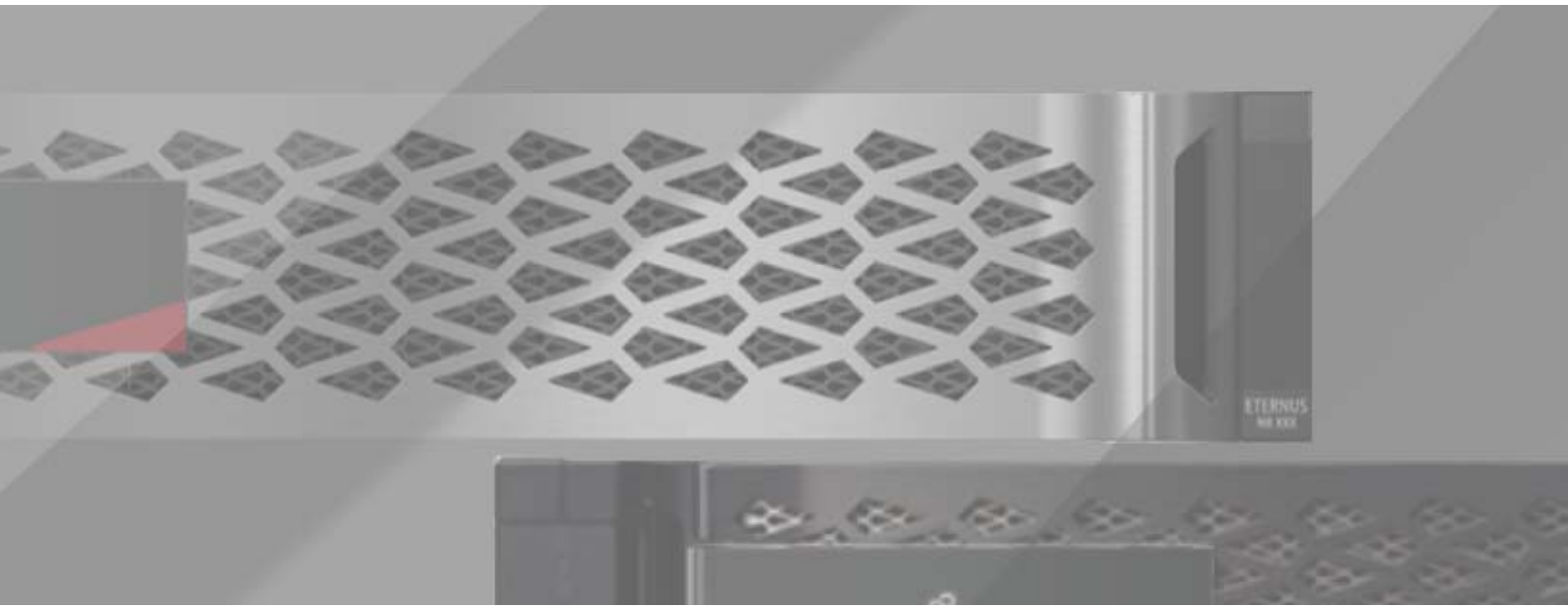
P3AG-5652-01ENZ0

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

Technical Overview of ONTAP FlexGroup Volumes



FUJITSU

Table of Contents

List of Figures

List of Tables

Preface

This document is an overview of ONTAP FlexGroup volumes, a feature of ONTAP. FlexGroup is an evolution of scale-out NAS containers that blends near-infinite capacity with predictable, low-latency performance in meta-data-heavy workloads.

First Edition
November 2020

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.

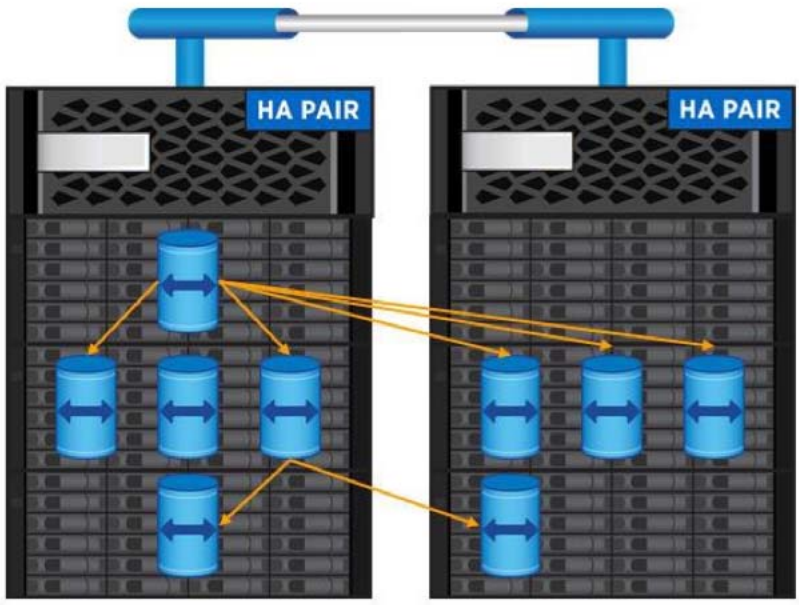
Volumes: A Tried-and-True Solution

The flexible volume, FlexVol software, was introduced in Data ONTAP technology in 2005 as part of the Data ONTAP 7.0 (Data ONTAP operating in 7-Mode) release. The concept was 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.

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. But the use case for large buckets of storage in a single namespace required petabytes of storage.
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.

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



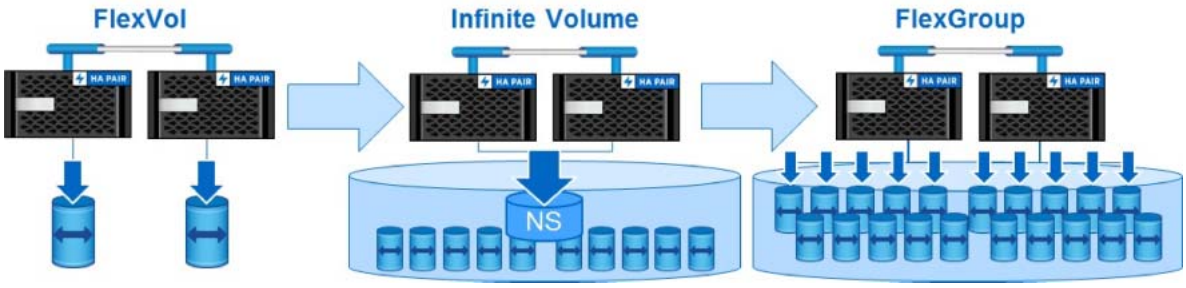
Although this architecture worked for many environments, it was awkward to manage and did not give a "single-bucket" approach to the namespace, where the FlexVol volume's capacity and file count constraints are limiting factors.

FlexGroup: An Evolution of NAS

ONTAP 9.1 brought innovation to scale-out NAS file systems: the ONTAP FlexGroup volume.

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 (Figure 2).

Figure 2 Evolution of NAS file systems in ONTAP



2. Advantages of ONTAP FlexGroup

Massive Capacity and Predictable Low Latency for High-Metadata Workloads

ONTAP FlexGroup offers 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, FlexGroup volumes can be managed 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.

3. 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 (AIR) 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
  See "What Are Large Files?" (page 11) for details.
- 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.
  If there are 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 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?

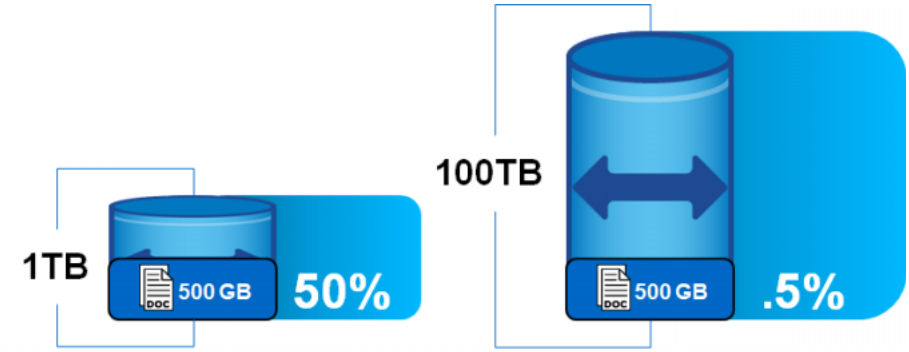
This document uses the term "large file" liberally. Therefore, it's important to define exactly what a large file is in the context of FlexGroup.

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 the section "Use Cases"). One type of workload that can create problems, 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 take up only 0.5% of the volume capacity. 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 (Figure 3).

Starting in ONTAP 9.6, elastic sizing helps mitigate concerns with larger files: ONTAP borrows space from other member volumes to allow large files to complete their writes. ONTAP 9.7 also introduces ingest algorithm changes to help balance large files and/or datasets with mixed file sizes. Both of these features make FlexGroup volumes a realistic landing place for most workloads. (See "Elastic Sizing" (page 25).)

Figure 3 What Are Large Files?



4. Supported Features with FlexGroup

Table 1 shows the current list of supported ONTAP features for FlexGroup. Contact Fujitsu Support for questions about supported features not listed.

Table 1 General ONTAP feature support

Supported Feature	Version of ONTAP First Supported
Snapshot 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 support	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
ETERNUS AX	ONTAP 9.1
Quota reporting	ONTAP 9.1
SnapMirror technology	ONTAP 9.1
User and group quota reporting (no enforcement)	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/unified replication	ONTAP 9.3
Antivirus scanning for SMB	ONTAP 9.3
Volume Autosize (Autogrow/Autoshrink)	ONTAP 9.3
QoS maximums/ceilings	ONTAP 9.3
FlexGroup expansion without SnapMirror rebaseline	ONTAP 9.3
Inode count factored into ingest	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 (ETERNUS AX only)	ONTAP 9.4
Relaxed SnapMirror limits	ONTAP 9.4
SMB 3.x Multichannel	ONTAP 9.4
FabricPool	ONTAP 9.5
Quota Enforcement Example	ONTAP 9.5
Qtree statistics	ONTAP 9.5
Inherited SMB watches and change notifications	ONTAP 9.5
SMB copy offload (offloaded data transfer (ODX))	ONTAP 9.5

Supported Feature	Version of ONTAP First Supported
Storage-Level Access Guard	ONTAP 9.5
FlexCache (cache only; FlexGroup as origin supported in ONTAP 9.7)	ONTAP 9.5
Elastic Sizing	ONTAP 9.6
SMB Continuously Available Shares (SQL/Hyper-V only)	ONTAP 9.6
MetroCluster	ONTAP 9.6
Volume rename	ONTAP 9.6
Volume shrink	ONTAP 9.6
Aggregate Encryption (AE)	ONTAP 9.6
Cloud Volumes ONTAP	ONTAP 9.6
FlexCache	ONTAP 9.7
NDMP	ONTAP 9.7
vStorage APIs for Array Integration (VAAI)	ONTAP 9.7
NFSv4.0 and NFSv4.1 (including parallel NFS, or pNFS)	ONTAP 9.7
FlexVol to FlexGroup In-Place Conversion	ONTAP 9.7
FlexGroup volumes as FlexCache origin volumes	ONTAP 9.7

Table 2 General NAS protocol version support

Supported NAS Protocol Version	Version of ONTAP First Supported
NFSv3	ONTAP 9.0
SMB 2.1, SMB 3.x	ONTAP 9.1 RC2
NFSv4.x	ONTAP 9.7

Table 3 Unsupported SMB 2.x and 3.x features

Unsupported SMB 2.x Features	Unsupported SMB 3.x
SMB Remote Volume Shadow Copy Service (VSS)	• SMB transparent failover • SMB scale-out • SMB Remote VSS • SMB directory leasing • SMB direct or remote direct memory access (RDMA) **Caution** SMB 3.0 encryption is supported with FlexGroup volumes.

Caution

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.

5. Use Cases

The ONTAP FlexGroup design is most beneficial in specific use cases, which are considered to be ideal. Other use cases for a FlexGroup volume are possible, but they generally depend on feature support. In most instances, the use case is limited to the supported feature set. For example, virtualization workloads can work on FlexGroup volumes, but they currently lack support for SIS cloning and offer no Virtual Storage Console integration.

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
- Artificial intelligence and machine learning log file repositories
- Software build and test environments (such as Git)
- Seismic, oil, and gas
- Media asset or HIPAA archives
- File streaming workflows
- Unstructured NAS data (such as home directories)
- Big data and data lakes (Hadoop with the NFS connector)

Non-Ideal Cases

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

- Virtualized workloads
- Workloads that require 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 that require specific features and functionality that are not currently available with FlexGroup volumes

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

6. Performance

Although ONTAP FlexGroup technology is positioned as a capacity play, it's a performance play as well. With a FlexGroup volume, there are no trade-offs; you can have massive capacity and predictable low-latency and high-throughput performance with the same storage container. A FlexGroup volume can accomplish this goal by adding concurrency to workloads and presenting multiple volume affinities to a single storage container without the clients or storage administrators needing to manage anything. In metadata-intensive workloads with high file counts, being able to present multiple volumes and cluster nodes quickly and easily enables ONTAP to use multiple hardware assets and CPU cores to perform at a higher performance threshold.
Note that the ETERNUS AX/HX has the same performance tendency.

FlexVol Versus FlexGroup: Software Build

In a simple workload benchmark using a software build tool (Git), a Linux kernel was compiled on a single FAS8080 node running ONTAP 9.1 with two aggregates of SAS drives and eight FlexVol member constituents in a FlexGroup, versus a single FlexVol on the same hardware. The metric being measured was a simple time-to-completion test. In this benchmark, the FlexGroup volume outperformed the FlexVol volume by two to six times across multiple Git operations. In addition, the same Git test was run with a gcc compile on the ETERNUS AX (Figure 4 and Figure 5).

Caution

The GCC compile works with a higher file count, thus the differences in completion times.

Figure 4 Git benchmark: Linux compile in FlexGroup versus FlexVol

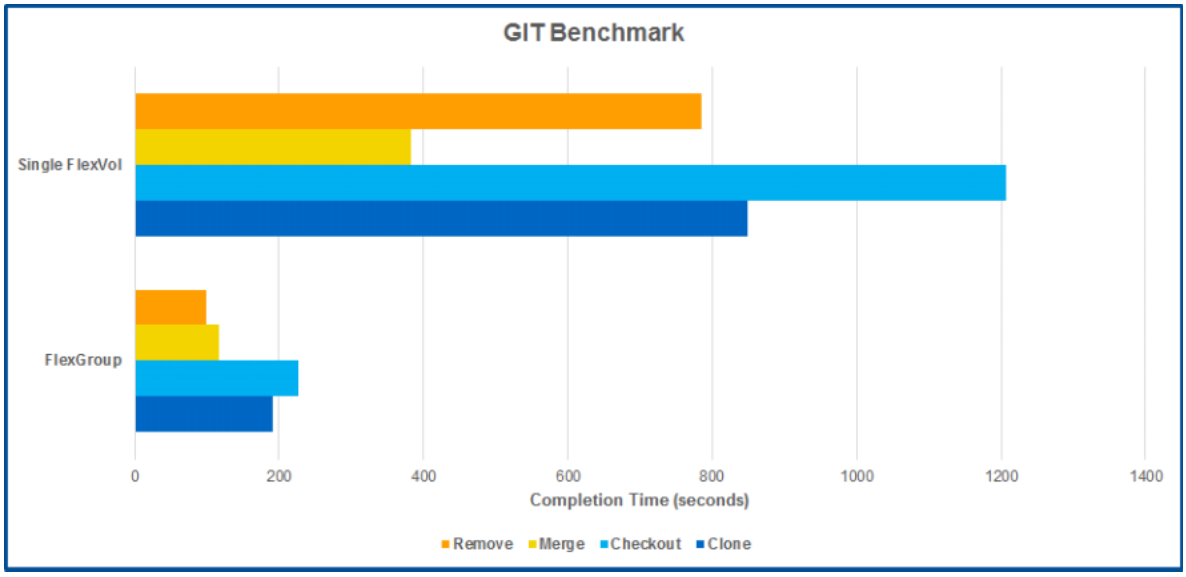
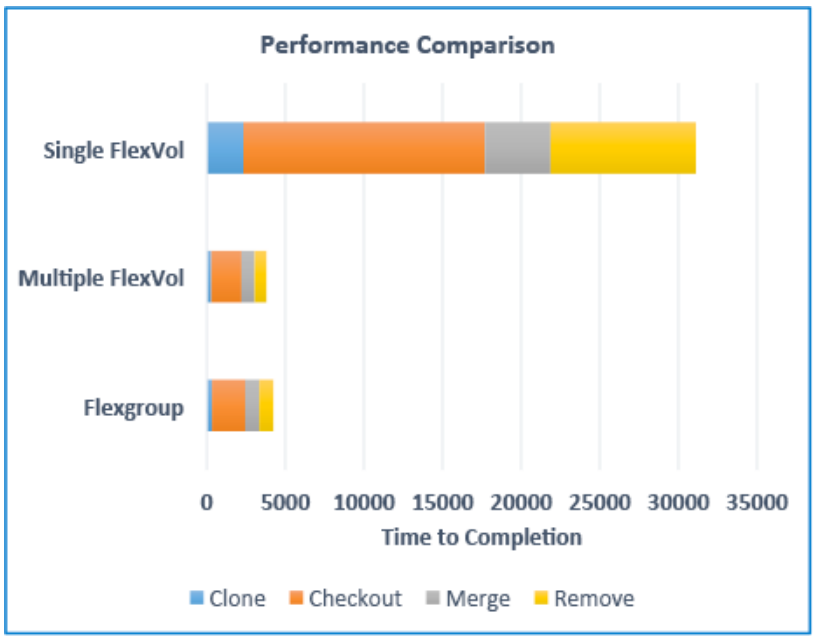


Figure 5 Git benchmark: GCC compile in FlexGroup versus FlexVol

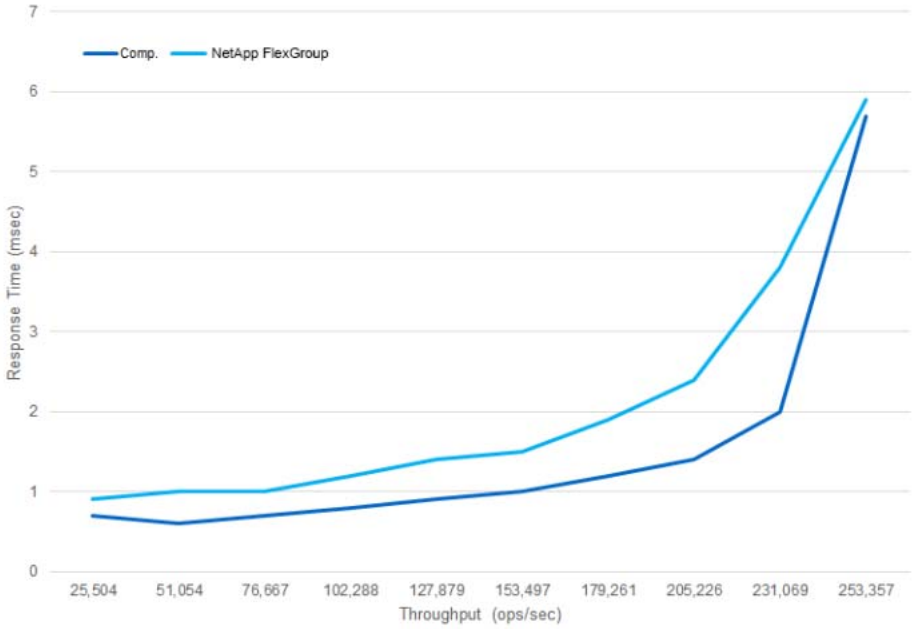


FlexGroup Versus Scale-Out NAS Competitor: Do More with Less

In another benchmark, we compared a FlexGroup volume on a two-node FAS8080 cluster running ONTAP 9.1 using SAS drives against a competitor system using 14 nodes. The competitor system also used some solid-state drives (SSDs) for metadata caching. This test used a standard NAS workload generation tool to simulate workloads.

In the test, we saw that a single FlexGroup volume with eight member constituents was able to ingest nearly the same number of operations per second at essentially the same latency curve as the competitor's 14-node cluster (Figure 6).

Figure 6 FlexGroup (two-node cluster) versus competitor (14-node cluster): standard NAS workload



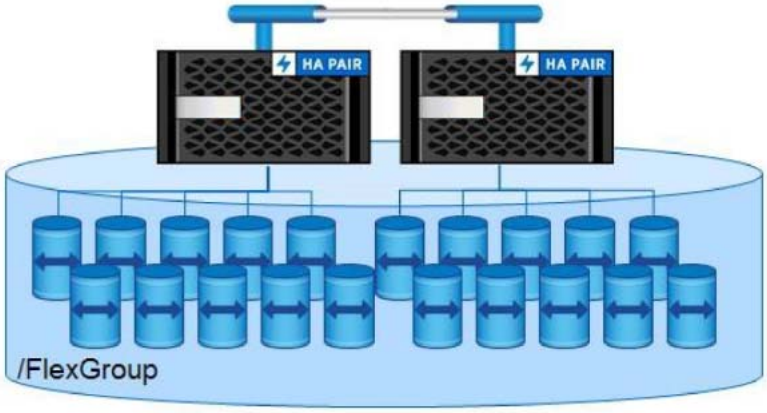
7. FlexGroup Technical Overview

ONTAP FlexGroup technology has taken the concept of the FlexVol volume in ONTAP and applied a WAFL subsystem known as the remote access layer (RAL). RAL directs the ingestion of new files and folders and tracks existing files and folders for fast redirection on reads. This capability provides automatic load balancing of incoming writes across participating member volumes.

Overview of a FlexGroup Volume

At a high level, a FlexGroup volume is simply a collection of FlexVol volumes acting as a single entity. NAS clients access the FlexGroup volume just as they would any normal FlexVol volume: from an export or a CIFS/SMB share (Figure 7). ONTAP redirects incoming NAS requests using symlinks within ONTAP that are transparent to clients.

Figure 7 FlexGroup volume



Although the underlying construct of a FlexGroup volume is a FlexVol volume, there are several benefits a FlexGroup volume offers that a normal FlexVol volume cannot. See "2. Advantages of ONTAP FlexGroup" (page 9) for details.

A FlexGroup volume creates files on a per FlexVol basis–there is no file striping. The throughput and performance gains for a FlexGroup volume are seen by way of concurrency of operations across multiple FlexVol volumes, aggregates, and nodes. A series of operations can occur in parallel across all hardware on which the FlexGroup volume resides. FlexGroup volumes are an ideal complement to the clustered ONTAP scale-out architecture.

File Creation and Automatic Load Balancing

When a file is created in a FlexGroup volume, that file is directed to the "best available" FlexVol member in the FlexGroup volume. "Best available" in this case means "most free space available," "most free inodes available," and the recent load on a FlexVol member. ONTAP makes these decisions without the need of administrator intervention. The goal of the FlexGroup volume is to keep member volumes as evenly allocated with capacity as possible, and also keep the ingest workload of a FlexGroup volume as evenly distributed across members as possible, with the fewest number of remote hard links possible.
This approach is known as "automatic load balancing" and is transparent to clients and storage administrators. This concept adds to the overall simplicity of the FlexGroup story: Storage administrators provision the storage in seconds and usually don't have to think about the design or layout.

Keep in mind the following features in various versions of ONTAP:

- ONTAP 9.7 and later versions help mitigate "out of space" (ENOSPC) issues by supporting volume autogrow.
- ONTAP 9.6 and later versions use Elastic Sizing to add an extra layer of protection for failed writes to large files when a member fills.
- ONTAP 9.7 introduces ingest changes to better handle streaming I/O workload placement.

Local Versus Remote Placement

Because a FlexGroup volume has multiple constituent volumes and FlexGroup is designed to place data evenly in all constituents, there is a notion of "remote placement" of files. ONTAP can operate in up to 24- node clusters in NAS-only configurations, so there is also a notion of "remote traffic."

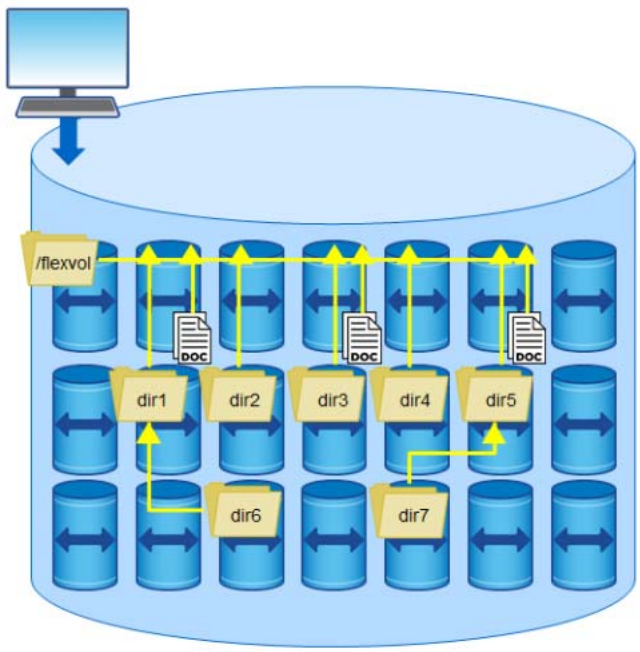
However, these two concepts are not synonymous. Remote traffic is traffic over the cluster interconnect network and is present in all ONTAP configurations, not just in a FlexGroup volume. Remote placement is unique to FlexGroup volumes and occurs when a file or folder is created on a FlexGroup member that does not own the parent directory. Remote placement can occur even on the node that owns the parent volume by way of a remote hard link allocated through the RAL.

Local placement involves creating the object so that it is stored in the same constituent as its parent directory. Local placement provides the highest metadata performance for that file or subdirectory in future operations. Remote placement, in contrast, causes that file or subdirectory to suffer a slight metadata access penalty. However, it allows the FlexGroup volume to place the new content on a different constituent from its parent directory to better distribute the collective dataset and workload. The penalty seen from remote placement of files or folders is more than offset by the performance and throughput gains of having multiple volume affinities for workloads (Figure 8).

To accomplish an evenly balanced workload, ONTAP monitors the status and member volume attributes every second for optimal placement of data. This process seeks to accomplish a balance of multiple goals:

- Space usage remains balanced.
- Overutilized member constituents are avoided by analyzing data creation patterns.
- Files are placed locally to ensure that latency is as low as possible.

Figure 8 Remote placement of files through remote hard links



Example of File and Folder Ingest

On GitHub, we added a `dd` script to do parallel `dd` operations on a client. This script provides a better throughput test than a single-threaded `dd` command. We used the script to illustrate how ONTAP places files across a FlexGroup on ingest. We then used diag-level commands to find the location of the files according to their inodes and mapped out some of their locations in the following scripts.

The configuration was as follows:

- ONTAP 9.7
- A single-node FlexGroup with eight members
- A single client running the `dd` script

This is the FlexGroup volume size output before we ran the script:

```
cluster::*> vol show -vserver DEMO -volume flexgroup_local* -fields used,percent-
used,size
vserver volume                size   used    percent-used
------- --------------------- ------ ------- ------------
DEMO    flexgroup_local__0003 2.50TB 57.28MB 5%
DEMO    flexgroup_local__0004 2.50TB 57.28MB 5%
DEMO    flexgroup_local__0005 2.50TB 57.28MB 5%
DEMO    flexgroup_local__0007 2.50TB 57.28MB 5%
DEMO    flexgroup_local__0002 2.50TB 57.28MB 5%
DEMO    flexgroup_local__0001 2.50TB 57.29MB 5%
DEMO    flexgroup_local__0006 2.50TB 57.29MB 5%
DEMO    flexgroup_local__0008 2.50TB 57.29MB 5%
DEMO    flexgroup_local       20TB   458.2MB 61%
9 entries were displayed.
```

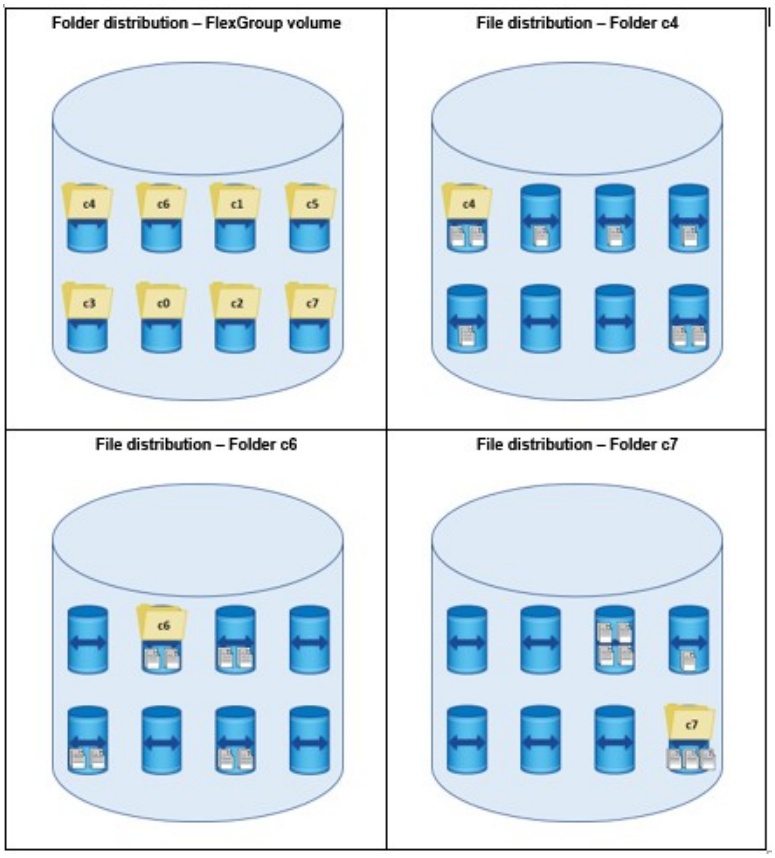
The script creates folders at the top level specified in the script, and then populates each folder with the specified number of files at a specific size. In this test, we chose eight folders and eight files to mirror the FlexGroup member volume count. We chose a file size of 1GB.

After we ran the script, the FlexGroup members looked like this (deduplication and compression kept the file sizes below 1GB):

```
cluster::*> vol show -vserver DEMO -volume flexgroup_local* -fields used,percent-
used,size,files,files-used
vserver volume                size   used    percent-used files     files-used
------- --------------------- ------ ------- ------------ --------- ----------
DEMO    flexgroup_local__0004 2.50TB 86.88MB 5%           21251126  106
DEMO    flexgroup_local__0007 2.50TB 90.85MB 5%           21251126  106
DEMO    flexgroup_local__0005 2.50TB 91.20MB 5%           21251126  107
DEMO    flexgroup_local__0006 2.50TB 91.23MB 5%           21251126  108
DEMO    flexgroup_local__0008 2.50TB 91.58MB 5%           21251126  107
DEMO    flexgroup_local__0002 2.50TB 92.48MB 5%           21251126  107
DEMO    flexgroup_local__0001 2.50TB 92.93MB 5%           21251126  114
DEMO    flexgroup_local__0003 2.50TB 96.44MB 5%           21251126  108
DEMO    flexgroup_local         20TB 733.6MB 61%          170009008 863
9 entries were displayed
```

The files were placed relatively evenly across all member volumes, as shown in the output. A closer look at the layout shows that the folders were allocated evenly across the member volumes. However, on a per-folder basis, the files were not allocated evenly. This is because the files were 1GB in size and were more likely to go remote from their parent folder due to their size (Figure 9).

Figure 9 File and folder distribution in a FlexGroup volume: dd script



Although the files were not evenly distributed on a per-folder basis, the result was that the files across all folders were evenly allocated over time. This result is shown in the space distribution output (Figure 10).

```
cluster::*> vol show -vserver DEMO -volume flexgroup_local* -fields used,percent-
used,size
vserver volume                size   used    percent-used
------- --------------------- ------ ------- ------------
DEMO    flexgroup_local__0002 2.50TB 85.77MB 5%
DEMO    flexgroup_local__0008 2.50TB 85.92MB 5%
DEMO    flexgroup_local__0003 2.50TB 86.04MB 5%
DEMO    flexgroup_local__0006 2.50TB 89.72MB 5%
DEMO    flexgroup_local__0004 2.50TB 90.34MB 5%
DEMO    flexgroup_local__0007 2.50TB 94.10MB 5%
DEMO    flexgroup_local__0005 2.50TB 94.18MB 5%
DEMO    flexgroup_local__0001 2.50TB 94.23MB 5%
DEMO    flexgroup_local       20TB   720.3MB 61%
9 entries were displayed.
```

Figure 10 Space distribution across member volumes: dd script

FlexGroup Load-Balancing Concepts

The following terms and concepts are central to how FlexGroup volumes ingest data and balance load across member volumes.

- Remote access layer
  The RAL is a new mechanism in ONTAP provided with the FlexGroup feature. The RAL allows a single WAFL message that runs against one member volume to manipulate inodes in that member volume and other member volumes.
- Ingest heuristics
  Ingest heuristics allow ONTAP to base intelligent decisions for file ingest on a series of decision points. Member volumes participating in a FlexGroup volume refresh every second to provide an up-to-date view of the current state of the volumes. Each ONTAP release features ingest heuristic improvements, so if you're using a FlexGroup volume, be sure to run the latest patched release.
- Remote allocation
  A FlexGroup volume allocates workloads based on the following:
  - Amount of data a member constituent holds (% used)
  - Amount of available free space in members
  - Last N seconds' worth of new content allocation requests
  - Last N seconds' worth of inode allocations and where they were drawn from
  - Number of free inodes/files in a member volume (ONTAP 9.3 and later)

  A FlexGroup volume generally favors local placement when possible, but sometimes, remote allocation of data is more likely. These scenarios include the following:
  - Creating subdirectories near the top-level junction of the FlexGroup volume
  - Creating files or folders in directories that already have many files or folders
  - Situations in which member constituent space allocation has a high discrepancy in capacity utilization
  - Situations in which member constituent volumes approach 90% capacity or inode utilization
  - Situations in which there is an unbalanced load (one member is getting more traffic than others)

  Sometimes, the member is a different FlexVol volume than the parent directory, but the FlexGroup ingest heuristics tend to favor local traffic for files over remote traffic. For directories, a FlexGroup volume tends to favor remote creation over local creation in FlexVol members. As traffic normalizes on a FlexGroup volume and the volume begins to allocate files, the allocation favors local traffic more, and remote traffic hovers around 5% to 10%.
  In ONTAP 9.7 and later, the ingest algorithms attempt to detect workload types to make more intelligent placement decisions based on the kinds of operations dominating them. This change helps the volume make more intelligent placement decisions for streaming and large-file workloads, and improves the workload balance across member volumes as data is ingested.

- Urgency
  Urgency in a FlexGroup volume is how full a member is (or, as of ONTAP 9.3, how close to the max file count the volume is) versus how likely it is to be used for ingest of data. Each node maintains two global variables used in determining how likely it is that new data will hit a member volume in a FlexGroup volume:
  - Free-warning
    This variable is set to 50GB by default. This number represents the amount of free space available in a FlexGroup member and is used to calculate the probability of how urgent it is to place content in remote members.
  - Free-danger
    This variable is set to 10GB by default. This number represents the threshold at which the member volume's urgency will be set to 100%. All ingest traffic will avoid the member volume until sufficient free space is added to the member or data is deleted.
- Tolerance
  Tolerance in a FlexGroup volume is a measure of usage disparity between members or the percentage of disparity of used space between members that a FlexGroup volume can tolerate before generating more remote allocation decisions.
  - Max-tolerance
    This value is set to a default of 10%. This means that a member volume can tolerate up to 10% of the working-set value (100GB) of used space before it has to send traffic remote a higher percentage of the time. For example, if one member is >10% more full than another member, traffic will be diverted elsewhere. Empty member volumes always use the max- tolerance value.
  - Min-tolerance
    This value is set to a default of 0%. When a member volume is full, the min- tolerance value is enforced and traffic is sent remote 100% of the time in an attempt to even up the space distribution.
  - Working-set
    This variable defines the free space level that the max- and min-tolerance percentages use for their calculations. The default of this value is 100GB.

Caveats

In cases in which a member volume starts to become "more full" than other member volumes, performance can deteriorate on the FlexGroup volume because the workload is creating more remote hard links.

In cases in which member volumes fill up completely in a FlexGroup volume, the entire FlexGroup volume reports ENOSPC errors to the client. Remediation steps must be taken to correct the issue (such as growing the FlexGroup members or deleting data). This also applies to member volumes running out of inodes. ONTAP 9.3 improved the ingest calculations to take member volume inode counts into consideration when allocating files.

ONTAP 9.6 and later versions include the elastic sizing feature described later in this document, which can help avoid scenarios where a member volume running out of space can fail a write operation.

Local Versus Remote Test

A simple file and directory creation test was performed to measure the local versus the remote placement for files and directories. The following setup was used:

- Two FAS8040 nodes
- Two SSD aggregates (non-ETERNUS AX personality)
- Four FlexVol member constituents per aggregate; eight total members
- 100,000 directories
- 10,000,000,000 files (100,000 per directory x 100,000 directories)
- Red Hat 7.x client

- Simple `mkdir` and `truncate` commands for loops:

```
for x in `seq 1 100000`; do mkdir dir$x; done
for x in dir{1..100000}; do truncate -s 1k /mnt/$x/file{1..100000}; done
```

In the above scenario, the remote allocation of directories was at 90%:

```
remote_dirs                                                    90
```

The remote allocation of files was only 10%:

```
remote_files                                                   10
```

The statistics above were pulled from the command `statistics show -object flexgroup` at the advanced privilege level. See the appendix for information about how to collect and view FlexGroup statistics.

Elastic Sizing

Files written to a FlexGroup volume live in individual member volumes. They do not stripe across member volumes, so if a file is written and grows over time, or a large file is written to a FlexGroup volume, that write might fail because of lack of space in a member volume.

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

- If you 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 can eventually fill 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, if your volume is composed of 4K files but then you zip some up and create a giant single file).

One solution is to grow volumes or delete data. However, administrators often don't see the issue until it's too late and "out of space" errors have occurred.

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 20TB/[number of member volumes in a FlexGroup volume].

In a two-node cluster, a FlexGroup that spans both nodes is likely to have 16 member volumes. That means if 20TB are available in a FlexGroup volume, the member volumes have 1.25TB available. Before ONTAP 9.6, any single file that exceeds 1.25TB in size could not write to a FlexGroup volume without volume autogrow enabled.

Starting in ONTAP 9.6, 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: an Airbag for Your Data

One of our FlexGroup volume developers refers to elastic sizing as an "airbag" in that it's 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.

Here's how it works:

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, member volume performance (new in ONTAP 9.6), and so on.
3. When a file is placed, since ONTAP doesn't know how large a file will get, 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 will query 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 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 will continue.
5. During the time ONTAP is looking for space to borrow, that file write is paused. This will appear to the client as a performance issue. But the overall goal isn't to finish the write fast–it's to allow the write to finish at all. Usually, a member volume will be 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 will need to query to borrow space more often.
6. The capacity borrowing will maintain the overall size of the FlexGroup–for example, if your FlexGroup volume is 40TB in size, it will remain 40TB.

Figure 11 File write behavior before elastic sizing

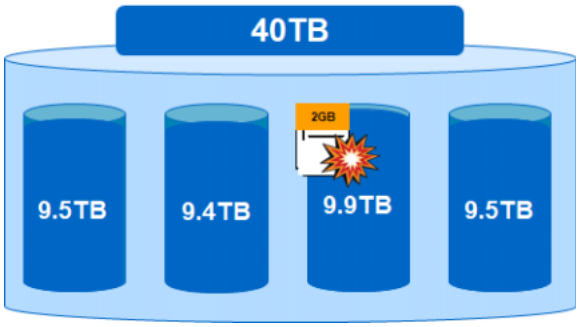
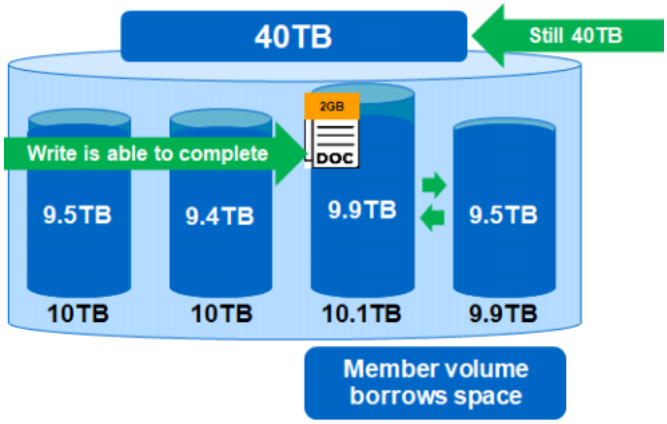


Figure 12 File write behavior after elastic sizing



After files are deleted or volumes are grown and space is available in that member volume again, ONTAP will re-adjust the member volumes back to their original sizes to maintain an evenness in space.

Ultimately, elastic sizing helps remove the administrator overhead of managing space and worrying so much about the initial sizing and deployment of a FlexGroup volume. You can spend less time thinking about how many member volumes you need, what size they should be, and so on.

When you combine elastic sizing in ONTAP 9.6 with features like autogrow/shrink, ONTAP can manage your capacity and help avoid emergency space issues.

FlexVol to FlexGroup In-Place Conversion

In ONTAP 9.7, it is now possible to convert a single FlexVol volume to a FlexGroup volume with a single member volume, in place. The conversion takes less than 40 seconds of disruption, regardless of how much data capacity or how many 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 expand the capacity.

Reasons to 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 2 billion files in a single volume
- Ability to scale out capacity or performance nondisruptively
- Multithreaded 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 still want to provide more capacity. Or perhaps a workload isn't performing well on a FlexVol volume, so you want to provide better performance handling with FlexGroup. Converting can help here.
For more information about FlexVol to FlexGroup conversion, see this blog or this podcast.

Volume Autosize (Autogrow/Autoshrink)

In ONTAP 9.3, support for volume autogrow was added for FlexGroup volumes. 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. This ability is especially useful in a FlexGroup volume, because volume autogrow can help prevent member volumes from filling prematurely and causing premature "out of space" scenarios in the entire FlexGroup volume. Applying volume autogrow to a FlexGroup volume is done in the same way as with a FlexVol volume. See the appendix for an example of how to apply volume autogrow.

Autosize Interaction with Elastic Sizing

Starting in ONTAP 9.6, 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.

When autosize is enabled for a volume, elastic sizing no longer takes effect for that volume. If you want to use elastic sizing for a volume, disable autosize.

64-Bit File Identifiers

By default, NFS in ONTAP uses 32-bit file IDs. 32-bit file IDs are limited to 2,147,483,647 maximum unsigned integers. With the 2 billion inode limit in FlexVol, this value fits nicely into the architecture.

However, because FlexGroup volumes can officially support up to 400 billion files in a single container (and theoretically, many more), the implementation of 64-bit file IDs was needed. 64-bit file IDs support up to 9,223,372,036,854,775,807 unsigned integers.

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. Check with your application and OS vendor for their support for 64-bit file IDs before enabling them.
Alternatively, 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 currently be enabled only with the advanced privilege level on the command line:

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

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. For more information about how enabling or disabling FSID change options can affect SVMs in high-file-count environments, see "How FSIDs Operate with Snapshot Copies" (page 30).

If a FlexGroup volume does not 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
```

Caution

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

NFSv3 Versus NFSv4.x

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/disabling 64-bit file IDs.

When you use both NFSv3 and NFSv4.x in an SVM you want the 64-bit ID option 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 2 billion files, whereas NFSv4.x would throw an error.

The options are:

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

Using Quota Enforcement to Limit File Count

Starting with ONTAP 9.5, it's possible to set up a quota policy that prevents a FlexGroup volume from exceeding 2 billion files if 32-bit file handles are still being used by way of quota enforcement.

Because quota policies don't apply to files created below the parent volume, create a qtree inside the FlexGroup volume. Then create a default quota rule with 2 billion files as the limit to help reduce the risk of users overrunning the 32-bit file ID limitations.

```
cluster::*> qtree create -vserver DEMO -volume FG4 -qtree twobillionfiles -secu-
rity-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 vol-
ume "FG4" in Vserver "DEMO".
cluster::*> quota report -vserver DEMO -volume FG4
Vserver: DEMO
                                     ----Disk---- ----Files-----  Quota
Volume   Tree      Type    ID       Used  Limit   Used   Limit   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 to limit access, preventing users from creating files at the volume level. Apply SMB shares to the qtree rather than the volume, and 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

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

System Manager Support for the 64-Bit File ID Option

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 enable or disable the NFS server option in System Manager is to use the CLI.

Impact of File ID Collision

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

Effects of File System ID (FSID) 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 with NFSv3, be sure to enable 64-Bit File Identifiers in ONTAP 9, but keep in mind that this option could affect older legacy applications that require 32-bit file IDs.

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 will cause the 32-bit or 64-bit file identifiers to apply to the SVM itself, meaning that the 2 billion file limit with 32-bit would apply to all volumes.
Therefore, the SVM would be limited to 2 billion files, rather than the FlexVol or FlexGroup volume. Leaving the FSID change option enabled allows volumes to operate as independent file systems with their own dedicated file counts.

To prevent file ID collisions, leaving the FSID change option enabled with FlexGroup volumes is recommended.

How FSIDs Operate with Snapshot Copies

When a Snapshot copy of a volume is taken, 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 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 option `-v3-fsid-change`. The NFS RFC spec does not require FSIDs for a file to be identical across file versions.

Caution

The `-v4-fsid-change` option does not apply to FlexGroup volumes, because NFSv4 is currently unsupported with FlexGroup volumes.

Directory Size Considerations

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. That means the memory allocation for the directory size can reach a maximum of only 320MB before a directory can no longer grow larger.

Number of Files That Fit into a Single Directory with the Default Maximum Size

To determine how many files can fit into a single directory with the default `maxdirsize` setting, use this formula:

- Memory in KB * 53 * 25%

Since the `maxdirsize` value is set to 320MB by default on larger systems, the maximum number of files in a single directory would be 4,341,760 for SMB and NFS. Keeping the `maxdirsize` value as low as possible is strongly recommended, but no higher than 80% of the 320MB limit (256MB, or around 3.4 million files).

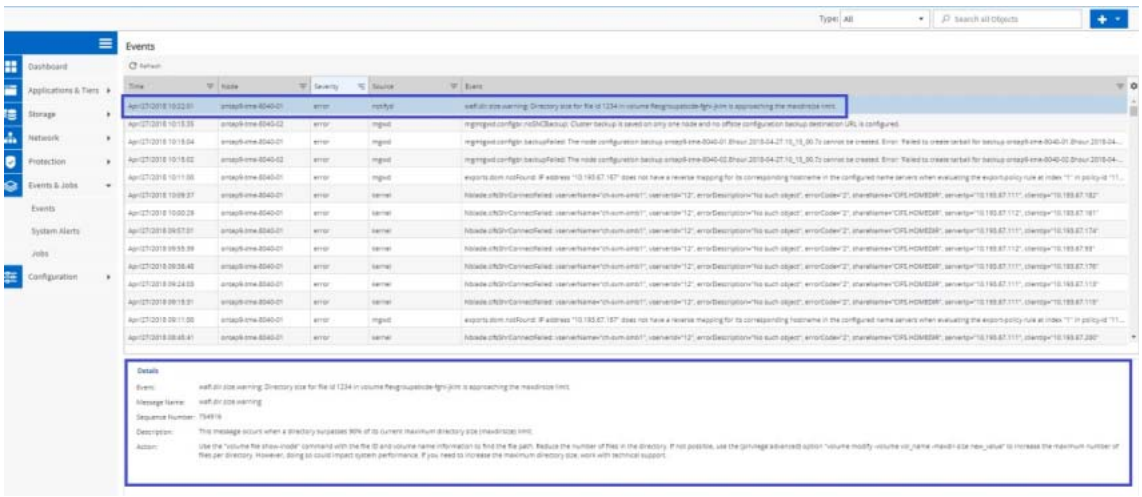
Event Management System Messages Sent When maxdirsize Is Exceeded

The following event management system messages are triggered when the `maxdirsize` value 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 through the ONTAP System Manager event section.
Active IQ Unified Manager can be used to monitor `maxdirsize`, trigger alarms, and notify before the 90% threshold (see Figure 13).

```
                                    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 max-
imum number of files per directory. However, doing so could impact system perfor-
mance. 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 max-
imum number of files per directory. However, doing so could impact system perfor-
mance. If you need to increase the maximum directory size, work with technical
support.
                                    Description: This message occurs when a direc-
tory has reached or surpassed 90% of its current maximum directory size (maxdir-
size) 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 max-
imum number of files per directory. However, doing so could impact system perfor-
mance. If you need to increase the maximum directory size, work with technical
support.
                                    Description: This mesaage occurs when a direc-
tory 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
```

Figure 13 ONTAP System Manager event screen with maxdirsize warning



Impact of Increasing the maxdirsize value

When a single directory contains many files, the lookups (such as in a "find" operation) can consume large amounts of CPU and memory. Starting in ONTAP 9.2, "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.

Do FlexGroup Volumes Avoid maxdirsize Limitations?

In FlexGroup volumes, each member volume has the same `maxdirsize` setting. Even though a directory could potentially span multiple FlexVol member volumes and nodes, the `maxdirsize` performance impact can still come into play, because directory size is the key component, not individual FlexVol volumes. Therefore, the size of a directory will still be an issue. Thus, FlexGroup volumes do not provide relief for environments facing maxdirsize limitations. Although newer platforms offer more memory and CPU, and the ETERNUS AX systems provide performance benefits, the best way to reduce performance impact in directories with large numbers of files is to spread files across more directories in a file system.

Impact of Exceeding the maxdirsize Value

When the `maxdirsize` value is exceeded in ONTAP, an "out of space" error (ENOSPC) is issued to the client and an event management system message is triggered. To remediate this, the storage administrator must increase the `maxdirsize` setting or move files out of the directory.

8. Features of FlexGroup

ONTAP FlexGroup features are grouped into the following categories:

- Simplicity
- Data protection
- Storage efficiencies

In addition, the functionality and advantages of FlexGroup are in and of themselves a feature and can be reviewed in "2. Advantages of ONTAP FlexGroup" (page 9).

Simplicity

One of the key benefits of a FlexGroup volume is the capability to create a massive container for capacity that delivers superior performance with the same ease as a normal FlexVol volume. FlexGroup offers support with ONTAP System Manager, Active IQ Performance Manager, and automated commands in the CLI, such as `volume create -auto-provision-as flexgroup` to enable quick and easy deployment of a FlexGroup volume.

Command Line (CLI)

Although most people think of GUIs when they think of simplicity, the command line is also a place where tasks can be made easier. The FlexGroup CLI offers some ways to improve the overall usability experience.

Volume Create (with -auto-provision-as flexgroup)

In ONTAP 9.2, a volume option was introduced for provisioning FlexGroup volumes at the admin privilege level. When specified, this option defaults to eight member FlexVol volumes per node. If no size is specified, the command creates member FlexVol volumes of 200MB each, so it's important to specify a size with the command. Keep in mind that the formula is (8 * aggregates specified/total specified size).
This is important because a member volume must be at least 100GB and can be no larger than 100TB.

If no aggregates are specified, ONTAP attempts to select all aggregates available to the specified SVM. So although it's possible to run a simplified command, it's best to be as prescriptive as possible for the FlexGroup configuration.

At a minimum, specify:

- Autoprovision as FlexGroup (`-auto-provision-as`)
- Volume name (`-volume`)
- SVM (`-vserver`)
- Volume size (`-size`)
- Export/mount point (`-junction-path`)

Optionally specify:

- Volume security style (`-security-style {unix|ntfs|mixed}`)
- UNIX permissions (`-unix-permissions`, if security style is UNIX)
- Thin provisioning (`-space-guarantee none`)

Volume Create (Advanced)

If customization outside of best practices is needed (such as when fewer/more member volumes are needed), `volume create -auto-provision-as flexgroup` might not be the right command to use to create a FlexGroup volume. If a cluster has more than four nodes, or if more granular control over the design and placement of the FlexGroup constituent members is desired, then the alternative is the command `volume create` without the `-auto-provision-as` option specified. Several options were added that are specific to FlexGroup creation.

Table 4 Volume command options for use with FlexGroup

Volume Option	What It Does
`-aggr-list`	Specifies an array of names of aggregates to be used for FlexGroup constituents. Each entry in the list creates a constituent on the specified aggregate. An aggregate can be specified multiple times to have multiple constituents created on it. This parameter applies only to FlexGroup.
`-aggr-list-multiplier`	Specifies the number of times to iterate over the aggregates listed with the `-aggr-list` parameter when creating a FlexGroup volume. The aggregate list will be repeated the specified number of times.
`-max-constituent-size`	Optionally specifies the maximum size of a FlexGroup constituent. The default value is determined by checking the maximum FlexVol size setting on all nodes used by the FlexGroup volume. The smallest value found is selected as the default for the max-constituent size for the FlexGroup constituent. This parameter applies only to FlexGroup volumes.

Volume Modify

After a FlexGroup volume is created, you must carry out changes to the volume options or size by using the CLI command `volume modify` or the ONTAP System Manager GUI.

Volume Expand

Another command for management of a FlexGroup volume is `volume expand`. This command enables storage administrators to add member volumes to an existing FlexGroup volume by using the `-aggr- list` and `-aggr-list-multiplier` options. Simply specify the aggregates to add members to and the number of desired members per aggregate. ONTAP does the rest.

Volume Expand on Volumes in SnapMirror Relationships
The `volume expand` command does not work natively with FlexGroup volumes participating in SnapMirror relationships earlier than ONTAP 9.3, because those required a rebaseline of the SnapMirror relationship. ONTAP 9.3 introduced the enhancement to allow volume expansion on FlexGroup volumes participating in a SnapMirror relationship without the need to rebaseline. As of ONTAP 9.3, ONTAP adjusts the FlexGroup member volume count on the next SnapMirror update.

Caution

Fujitsu recommends upgrading to ONTAP 9.3 or later when you use SnapMirror with FlexGroup volumes.

Expanding FlexGroup Volumes in SnapMirror Relationships Prior to ONTAP 9.3
To expand a volume (to add more members) in a SnapMirror relationship prior to ONTAP 9.3, perform the following steps:

Procedure ▶▶▶

1. Perform `snapmirror delete` on the existing relationship on the destination.
2. Perform `snapmirror release` on the source.

3. Perform `volume delete` on the destination FlexGroup data protection volume.
4. Perform `volume expand` on the source FlexGroup volume.
5. Use `volume create` to create a new destination FlexGroup data protection volume with the same size and constituent count as the source FlexGroup volume.
6. Perform `snapmirror initialize` on the new relationship (rebaseline).

Growing the member volumes without needing to rebaseline the relationship is supported with SnapMirror and FlexGroup.

ONTAP System Manager

ONTAP 9.1 offered ONTAP System Manager support for FlexGroup right out of the gate. A FlexGroup tab was added to the GUI under the Volumes page. On this page, storage administrators can manage an existing FlexGroup volume or create a FlexGroup volume with two clicks. ONTAP 9.4 raised the stakes with an even more robust GUI to support FlexGroup volumes, with the ability to perform virtually all of the same tasks as a FlexVol volume.

Creating a FlexGroup Volume

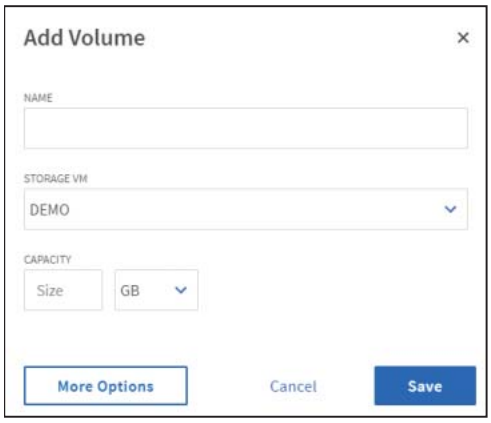
ONTAP 9.7 introduced a redesign of ONTAP System Manager to help simplify configuration operations. However, this redesign also means that fewer advanced configuration options are available. Use the CLI or REST API to do advanced configuration.

Creating a FlexGroup Volume (ONTAP 9.7 and Later)

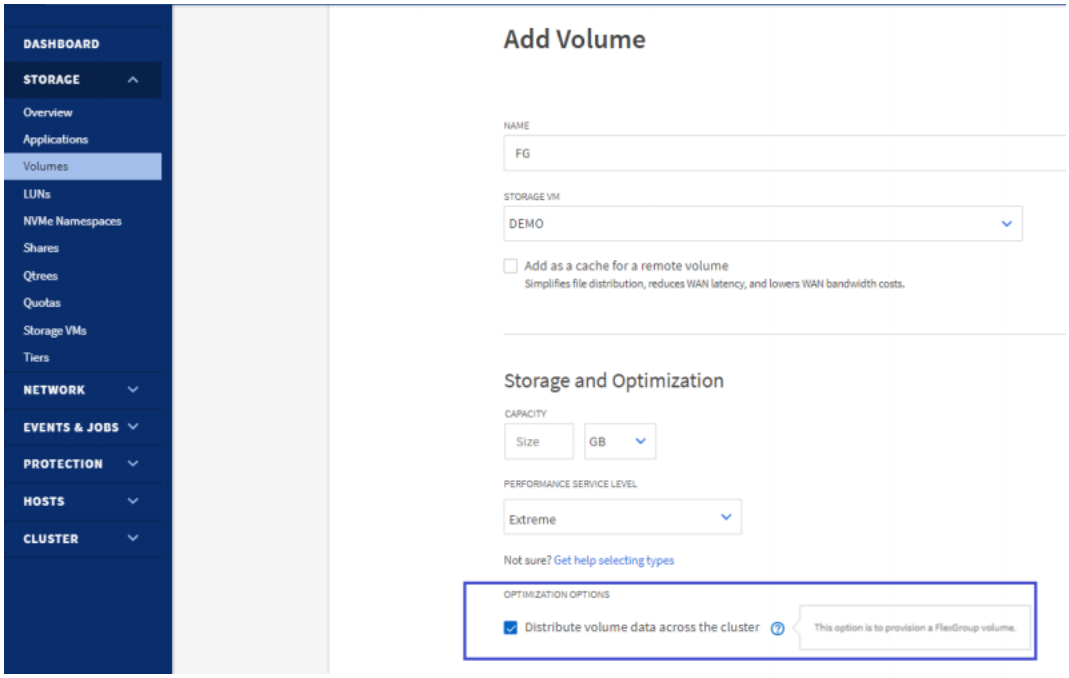
ONTAP 9.7 introduced a redesign of ONTAP System Manager to help simplify configuration operations. This section covers how to create a FlexGroup volume in ONTAP System Manager 9.7 and later.

Procedure

1. Volumes can be created from multiple locations in ONTAP System Manager. In earlier versions, you specified FlexVol or FlexGroup when starting the wizard. Now, there is only one wizard for both volume types. The basic wizard is as simple as name/SVM/size, but it creates a FlexVol volume by default. To create a FlexGroup volume, click the More Options button.



2 After you select More Options, configure the volume using the available options, such as size, sharing, export policies, and data protection. To make the volume a FlexGroup volume, use the Distribute Volume Data Across the Cluster checkbox under Optimization Options.

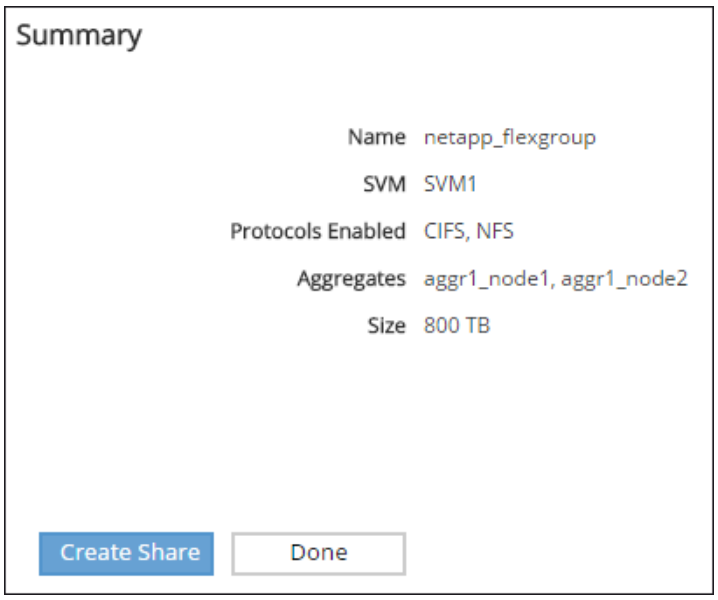


3 Click Save.

Managing a FlexGroup Volume

After the FlexGroup volume finishes creating, a dialog box opens with the option to create a CIFS share or to click Done to finish the process (Figure 14).

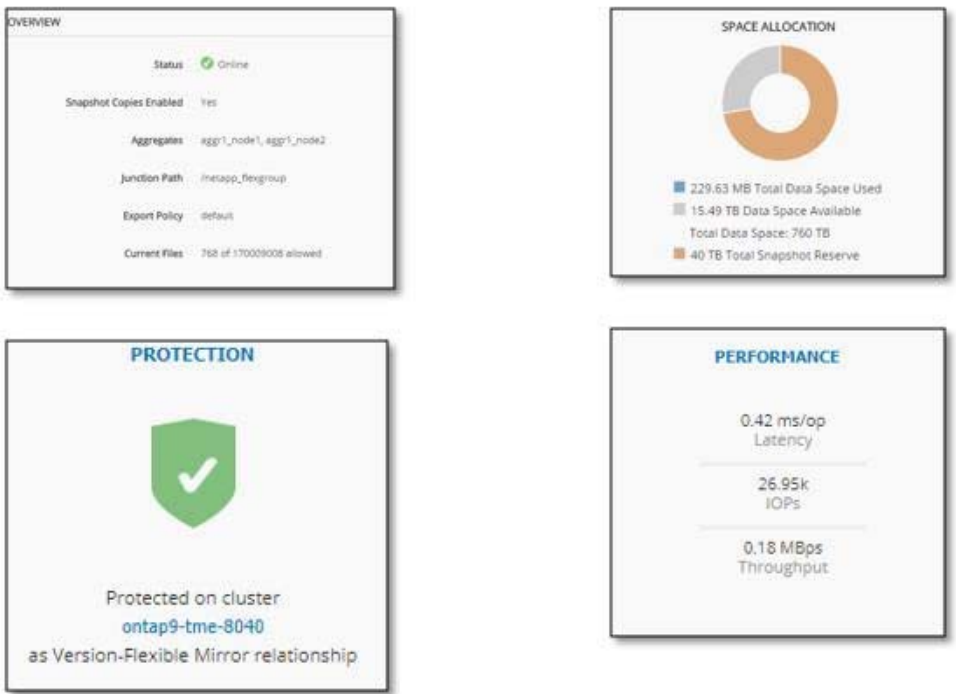
Figure 14 Create shares to a FlexGroup volume in ONTAP System Manager



From there, the administrator can manage the FlexGroup volume from the FlexGroup tab. ONTAP System Manager provides an overview of the FlexGroup volume, including (Figure 15):

- Volume overview
- Space allocation
- Data protection status (SnapMirror)
- Current performance statistics

Figure 15 FlexGroup overview in System Manager

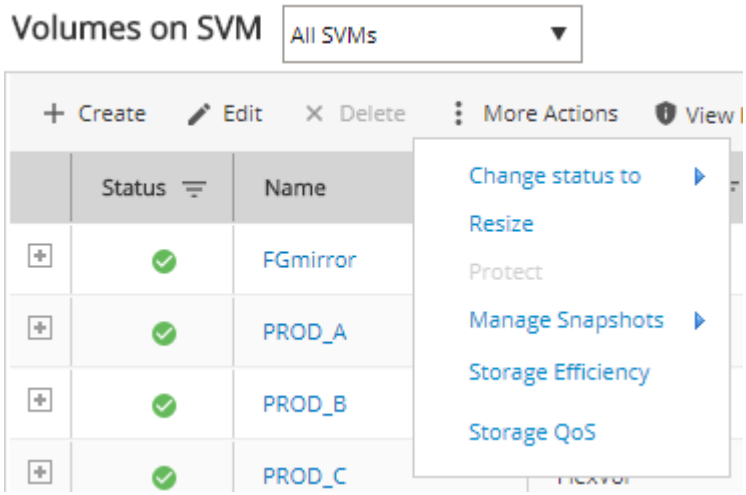


For more detailed information, you can click the hyperlinked volume name or Show More Details. This view gives information such as:

- Volume overview
- Snapshot copies
- Data protection details
- Storage efficiency details
- Performance details (real-time only)

In addition, the FlexGroup volume can be managed from the System Manager GUI via the Edit and More Actions buttons (Figure 16).

Figure 16 Managing an existing FlexGroup volume

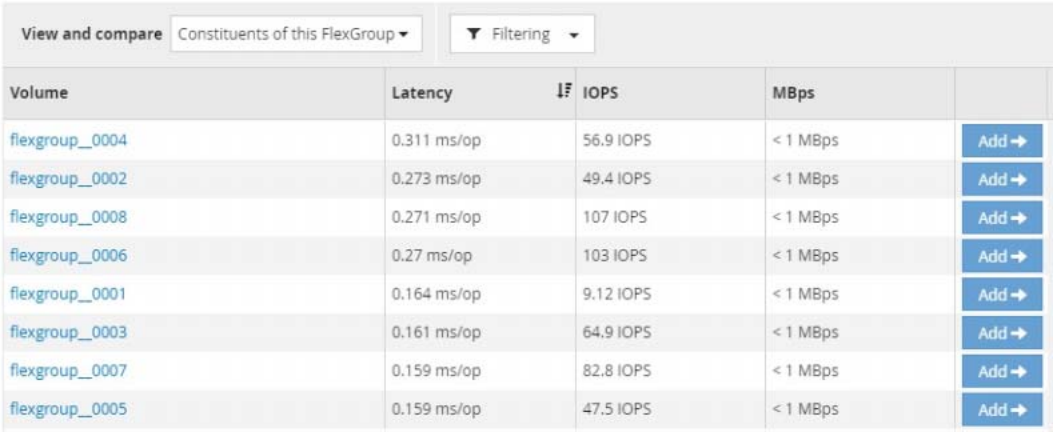


Active IQ Performance Manager

In addition to ONTAP System Manager support, you can use Active IQ Performance Manager to monitor a FlexGroup volume and its members at a granular level.

In Performance Manager, a FlexGroup volume can be found with the other volumes in an SVM. When you click the desired object, you see a screen with a member volume summary. You can also add these members to a graphical view over a specified range of time (Figure 17).

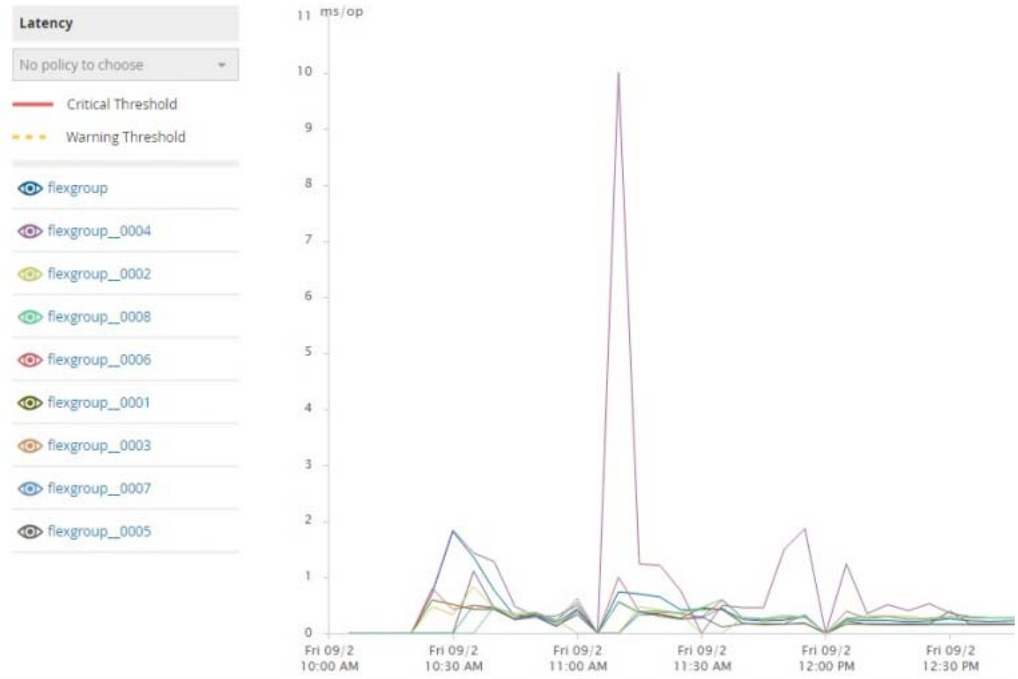
Figure 17 FlexGroup member volumes in Performance Manager



As member volumes are added to the Performance Manager graphical view, they are assigned different graph line colors to differentiate them in the charts. In this way, any member that deviates from the expected performance output can be investigated and remediated (Figure 18 and Figure 19).

Figure 18 Graphical representation of FlexGroup member volumes in Performance Manager

Figure 19 Graphical representation of FlexGroup member volumes in Performance Manager–zoomed



REST APIs

REST API support was introduced in ONTAP 9.6. Rather than navigating a proprietary interface (such as Manageability SDK), REST APIs enable you to use a universal standard for accessing and interacting with a cluster.

REST API documentation can be found at `http://[your_cluster_IP_or_name]/docs/api` and offers examples and an interactive "try it out" feature that allows you to generate your own REST APIs.

Cloud Volumes ONTAP

ONTAP 9.6 introduced official support for Cloud Volumes ONTAP—an ONTAP solution running in the cloud. This means that you can now deploy a FlexGroup volume in Cloud Volumes ONTAP.

FlexGroup volumes running in Cloud Volumes ONTAP are able to 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:

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

Although FlexGroup volumes are able to 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 instance. Also, creating a FlexGroup currently requires use of System Manager or the CLI. There is currently no way to create a FlexGroup volume in Cloud Manager.

Single Transparent Namespace

FlexGroup offers the advantage of massive capacity that exceeds that of a normal FlexVol volume without needing to implement a complicated architecture. The entire volume can be mounted as a single export or share and does not require more changes on the application side, even when more storage is added. This benefit reduces the management complexity associated with managing numerous containers and numerous mount points or shares from the client side.

Qtrees

ONTAP 9.3 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 are useful for home directory workloads, because folders can be named to reflect the usernames of users accessing data, and dynamic shares can be created to provide access based on a username. Qtrees are distributed across a FlexGroup volume in much the same way as a normal folder. Quota monitoring can be applied at the qtree level, and in ONTAP 9.5 and later, quota enforcement policies can be applied. Qtrees are created and managed the same way as a FlexVol qtree is managed. A maximum of 4,995 qtrees is supported per FlexGroup volume.

Quota Enforcement

When quota enforcement is enabled on a qtree or for a user, ONTAP disallows new file creations or writes after a quota is exceeded. In addition, an EMS message is logged at DEBUG severity level to notify storage administrators of the quota violation. You can configure these EMS messages so that the system forwards them as SNMP traps or as syslog messages.

Integrated Data Protection

FlexGroup supports several methods of data protection, including RAID DP software, RAID Triple Erasure Coding (RAID-TEC technology), Snapshot technology, SnapMirror replication technology, and NFS or CIFS-mounted tape backup.

RAID DP and RAID Triple Erasure Coding (RAID-TEC)

RAID DP is known as "dual parity" RAID and can survive two simultaneous disk failures per RAID group. This means that if a drive fails, data is still protected with another parity drive.

RAID Triple Erasure Coding (RAID-TEC) was new in ONTAP 9.0 and provides an extra parity drive for RAID groups using larger-capacity drives. This feature helps protect against drive failures during longer rebuild times for larger-capacity drives. RAID-TEC also provides larger RAID groups in terms of drive numbers.

All RAID protection features are supported with FlexGroup.

Snapshot Technology

Snapshot copies are automatically scheduled point-in-time copies that take up no space and incur no performance overhead when created. Over time, Snapshot copies consume minimal storage space, because only changes to the active file system are written. Individual files and directories can be easily recovered from any Snapshot copy, and the entire volume can be restored back to any Snapshot state in seconds.

Snapshot copies are supported for use with FlexGroup. Each Snapshot copy is made as a consistency group of the FlexVol members in which all members are quiesced and prepared for a Snapshot copy to ensure a consistent point-in-time copy of all members in a FlexGroup volume.

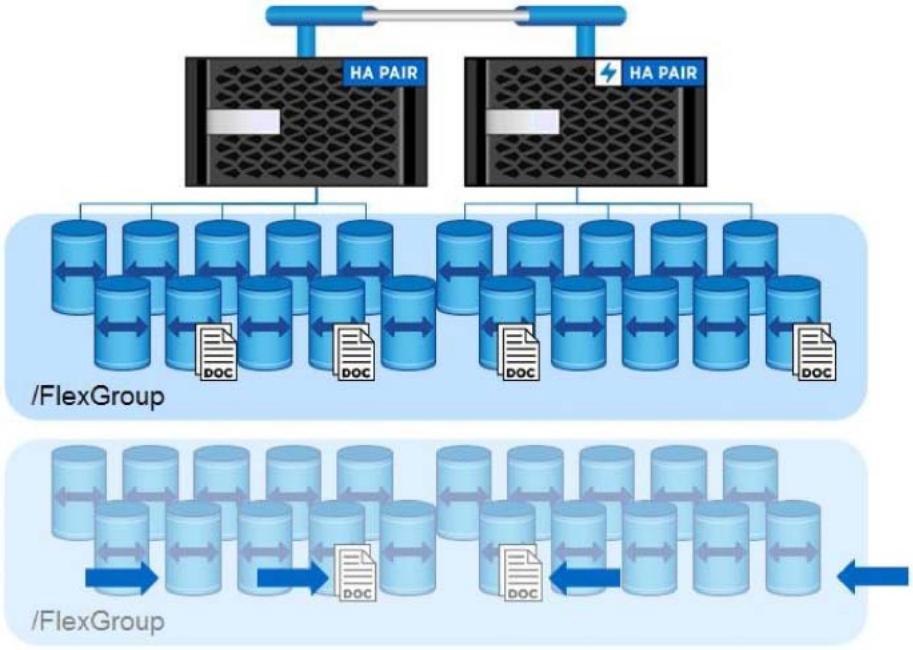
If any member in a FlexGroup volume in unable to perform the Snapshot operation (out of space, offline, too busy to complete), then the entire FlexGroup Snapshot copy is considered partial and to have failed. ONTAP cleans up the remnants of the attempted Snapshot copy and issues an EMS event (Figure 20).

After a Snapshot copy is created, storage administrators can perform the following operations for restores:

- Use SnapRestore technology to restore the entire FlexGroup volume.
- Navigate the Snapshot directories with the .snapshot (NFS) or ~snapshot (CIFS/SMB) folders to restore individual files and folders.

> **Caution**
>
> Single File SnapRestore is not supported, nor is restoring a single FlexGroup member volume. Using SnapRestore for a FlexGroup volume restores the entire volume, not just single member volumes. Currently, full Snapshot restores are available only with diag privileges.

Figure 20 FlexGroup Snapshot copy



SnapMirror and SnapVault

SnapMirror provides asynchronous replication of volumes, independent of protocol for data protection and disaster recovery. SnapVault provides asynchronous Snapshot copy retention for data protection and backup/archive use cases.

SnapMirror support with FlexGroup was provided starting in ONTAP 9.1. SnapVault support for FlexGroup volumes was added in ONTAP 9.3.

SnapMirror and SnapVault function in a similar manner as Snapshot copies. All member volumes need to have a successful Snapshot copy, and all member volumes are concurrently replicated to the DR site. If any component of that operation fails, mirroring with SnapMirror fails as well.

See the FUJITSU Storage ETERNUS AX/HX series FlexGroup Volume Data Protection Best Practices for details of best practices and limits for SnapMirror and SnapVault with FlexGroup volumes.

Tape Backup with CIFS/SMB or NFS

Tape backup for FlexGroup volumes can be performed by using external backup applications such as Commvault Simpana and Symantec NetBackup over CIFS or NFS mounts. ONTAP 9.7 and later versions also offer support for NDMP. Performance for NDMP on FlexGroup volumes is expected to be similar to that of FlexVol volumes. Keep in mind the total file count being backed up to maintain backup service- level objectives (SLOs). If necessary, break up backup jobs into smaller chunks by backing up at the folder level.

See the FUJITSU Storage ETERNUS AX/HX series FlexGroup Volume Data Protection Best Practices for more information about NDMP support with FlexGroup volumes.

MetroCluster

ONTAP 9.6 introduces support for FlexGroup on MetroCluster deployments (Fibre Channel and 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.

See FUJITSU Storage ETERNUS AX/HX series MetroCluster Solution Design and Architecture for more information about MetroCluster.

Storage Efficiencies

FlexGroup also offers support for the following storage efficiency technologies:

- **Inline and postprocess deduplication** removes duplicate data blocks in primary and secondary storage, storing only unique blocks. This action results in storage space and cost savings. Deduplication runs on a customizable schedule.
- **Inline aggregate deduplication (or cross-volume deduplication)** provides inline storage efficiency at the aggregate level. This allows duplicate blocks to be reduced if they exist in multiple FlexVol volumes in the same aggregate. FlexGroup member volumes are an excellent use case for this feature, introduced in ONTAP 9.2. ONTAP 9.3 automated scheduling and scheduled background inline aggregate deduplication.
- **Inline adaptive compression** was introduced for primary workloads such as database and desktop virtualization with ONTAP 8.3.1. Inline compression is on by default in the ETERNUS AX product family starting with 8.3.1.
- **Inline data compaction** was introduced in ONTAP 9.0 and further reduces the physical used space needed to store data. Data compaction is a significant addition to our storage efficiency portfolio and complements deduplication and compression technologies. Data compaction takes I/Os that normally consume a 4K block on physical storage and packs multiple such I/Os into one physical 4K block.
- **Thin provisioning** has been around for years and offers storage administrators the ability to overprovision virtual containers (FlexVol volumes) on physical storage (aggregates). With FlexGroup, thin provisioning can play an important role in how initial deployment of the FlexGroup volume is handled. Thin provisioning also allows member constituents to be much larger than their physical aggregate counterparts, which provides flexibility in the design of the container.

These features are applied by ONTAP at the member volume level individually, but configured by the storage administrator at the FlexGroup level for ease of management. In earlier releases of ONTAP, the features were applied at the FlexVol member level, so Table 5 gives guidance on what ONTAP versions support management of these features at the FlexGroup level, and which ONTAP versions require more granular management of the efficiencies per member volume.

Caution

Fujitsu highly recommends using ONTAP 9.3 or later for maximum storage efficiency with FlexGroup volumes.

Table 5 Storage efficiency guidance for FlexGroup in ONTAP versions

	9.1RC1	9.1RC2 and later	9.2RC1 and later
Thin provisioning	FlexGroup level	FlexGroup level	FlexGroup level
Inline deduplication	FlexVol member	FlexGroup level	FlexGroup level
Postprocess deduplication	FlexVol member	FlexGroup level	FlexGroup level
Inline data compression	FlexVol member	FlexGroup level	FlexGroup level
Inline data compression	FlexVol member	FlexGroup level	FlexGroup level
Postprocess data compression	FlexVol member	FlexGroup level	FlexGroup level
Aggregate inline deduplication	N/A	N/A	FlexGroup level

Applying Storage Efficiencies per FlexGroup Member Volume

If a FlexGroup volume does not currently have support to enable storage efficiencies at the FlexGroup level, use the following command to enable it on every FlexVol member. This should be necessary only in ONTAP 9.1RC1.

```
cluster::*> volume efficiency on -vserver SVM -volume flexgroup4*
Efficiency for volume "flexgroup4TB__0001" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0002" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0003" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0004" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0005" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0006" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0007" of Vserver "SVM" is enabled.
Efficiency for volume "flexgroup4TB__0008" of Vserver "SVM" is enabled.
```

To modify:

```
cluster::*> volume efficiency modify -vserver SVM -volume flexgroup4* -compression
true -data-compaction true -inline-compression true -inline-dedupe true

cluster::*> volume efficiency show -vserver SVM -volume flexgroup4* -fields data-
compaction,compression,inline-compression,inline-dedupe
vserver volume             compression inline-compression inline-dedupe data-com-
paction
------- ------------------ ----------- ------------------ ------------- ----------
-----
SVM     flexgroup4TB__0001 true        true               true          true
SVM     flexgroup4TB__0002 true        true               true          true
SVM     flexgroup4TB__0003 true        true               true          true
SVM     flexgroup4TB__0004 true        true               true          true
SVM     flexgroup4TB__0005 true        true               true          true
SVM     flexgroup4TB__0006 true        true               true          true
SVM     flexgroup4TB__0007 true        true               true          true
SVM     flexgroup4TB__0008 true        true               true          true
```

FabricPool

In ONTAP 9.2, the ability to automatically tier cold data blocks on SSD aggregates to the cloud or on- premises Amazon Simple Storage Service (Amazon S3) object storage was added for FlexVol volumes. This functionality allowed storage administrators to preserve more costly SSDs for active workloads, whereas cold or unused data was moved to more cost-effective cloud tiers. This feature is known as FabricPool. You can learn more about the feature in the FUJITSU Storage ETERNUS AX/HX series FabricPool Best Practices.

ONTAP 9.5 introduced support for FabricPool for FlexGroup volumes. There are no special considerations to make for FlexGroup volumes; the same FlexVol considerations apply.

At-Rest Encryption

ONTAP 9.2 introduced support for Volume Encryption (VE) for FlexGroup volumes. Implementing this feature with FlexGroup volumes follows the same recommendations and best practices as stated for FlexVol volumes. Starting in ONTAP 9.5, you can "rekey" an existing FlexGroup volume. To encrypt existing FlexGroup volumes in versions earlier than ONTAP 9.5, you must create a volume with encryption enabled and then copy the data to the volume at the file level. You can do this with a utility such as the XCP Migration Tool.

Generally speaking, VE requires the following:

- A valid VE license
- A key management server (on-box or off-box as of ONTAP 9.3)
- A cluster-wide passphrase (32 to 256 characters)
- ETERNUS AX/HX hardware that supports AES-NI offloading
- ONTAP 9.5 or later to rekey existing FlexGroup volumes

For information about implementing and managing VE with FlexGroup and FlexVol volumes, see ETERNUS AX/HX series Encryption Power Guide and ETERNUS AX/HX series Scalability and Performance Using FlexGroup Volumes Power Guide on the Fujitsu manual site.

ONAP 9.6 added Aggregate Encryption (AE), which allows you to encrypt at the aggregate level. FlexGroup volumes can use AE, provided all aggregates that contain member volumes are encrypted.

Quality of Service (QoS)

Starting in ONTAP 9.3, you can apply maximum storage QoS policies to a FlexGroup volume to help prevent a FlexGroup volume from acting as a bully workload in ONTAP. Storage QoS can help you manage risks around meeting your performance objectives. You use storage QoS to limit the throughput to workloads and to monitor workload performance. You can reactively limit workloads to address performance problems, and you can proactively limit workloads to prevent performance problems.

How Storage QoS Maximums Work with FlexGroup Volumes

With a FlexGroup volume, storage QoS policies are applied to the entire FlexGroup volume. 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 21 and Figure 22 show how storage QoS gets applied to a FlexGroup volume spanning multiple nodes in a cluster.

Figure 21 Storage QoS on FlexGroup volumes: single-node connection

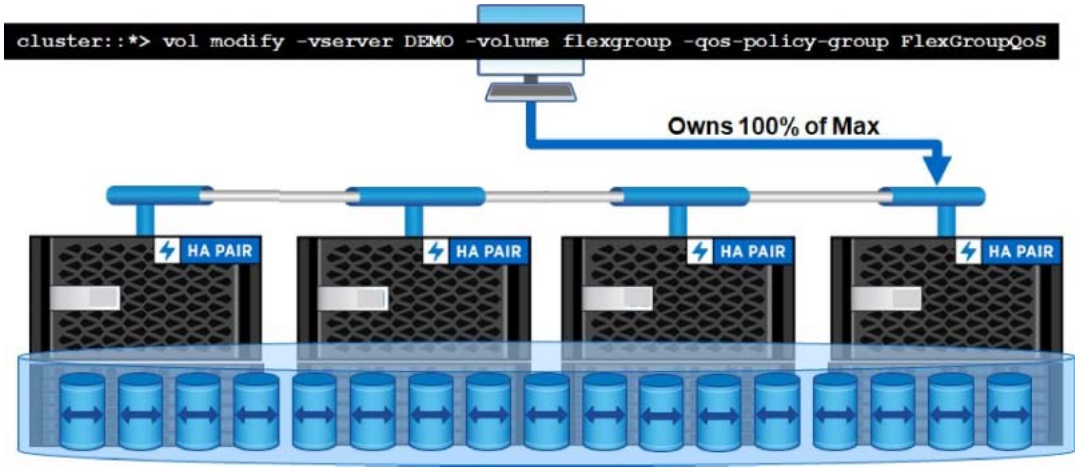
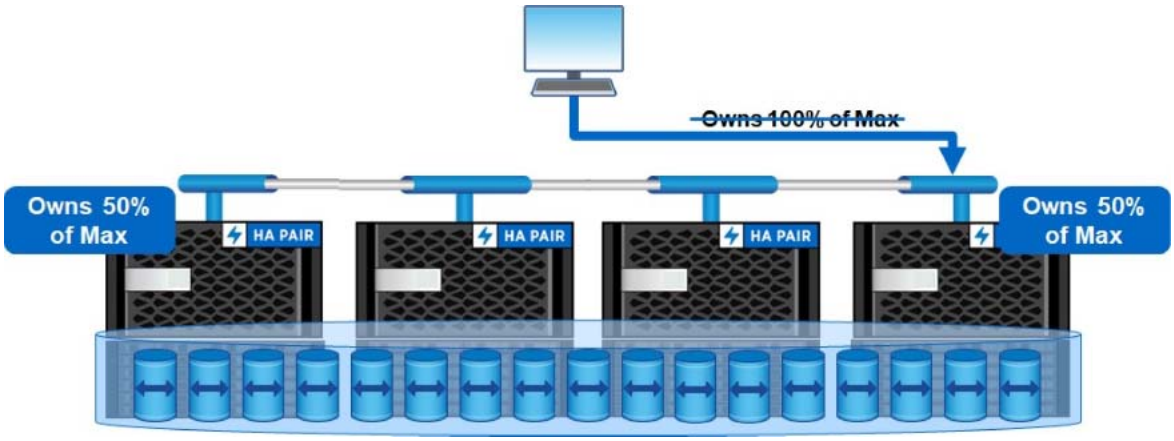


Figure 22 Storage QoS on FlexGroup volumes: multinode connection



Storage QoS Considerations with FlexGroup Volumes

Currently, storage QoS is applied only at the FlexGroup volume level and supports only QoS maximums. QoS minimums, adaptive QoS, file-level QoS, and nested policies are currently not supported with FlexGroup volumes. Policies are currently applied at the command line only. GUI support for FlexGroup volume QoS will be included in future ONTAP releases.

Quality of Service (QoS) Minimums

ONTAP 9.4 added support to 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. This feature is supported for the ETERNUS AX systems only.

Adaptive Quality of Service (QoS)

ONTAP 9.4 also introduced adaptive QoS support for FlexGroup volumes, which allows ONTAP to adjust the IOPS/TB values of a QoS policy as the volume capacity is adjusted.

A. Appendix

The following sections cover FlexGroup information not covered in the previous sections of this document, including:

- Command-line examples of creating and managing ONTAP FlexGroup
- Gathering FlexGroup statistics
- Viewing FlexGroup ingest usage distribution through the CLI
- Sample Python script for generating many files from a client

Command-Line Examples

Using the ONTAP 9.2 `auto-provision-as` option:

```
cluster::> vol create -auto-provision-as flexgroup -vserver SVM -volume flexgroup92
-junction-path /flexgroup92 -size 100t -space-guarantee none -security-style unix
```

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]
```

Resizing the FlexGroup volume:

```
cluster::> volume size -vserver SVM -volume flexgroup -new-size 20PB
```

Adding members to the FlexGroup volume:

```
cluster::> volume expand -vserver SVM -volume flexgroup -aggr-list
aggr1_node1,aggr1_node2 -aggr-list-multiplier 2
```

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.
```

FlexGroup Statistics

In ONTAP 9, a statistic object called `flexgroup` was added. The object is available only with diag privileges. This object gathers the following counters:

```
    cat1_tld_local                cat1_tld_remote
    cat2_hld_local                cat2_hld_remote
    cat3_dir_local                cat3_dir_remote
    cat4_fil_local                cat4_fil_remote
    dsidlist_factory_enomem       groupstate_analyze
    groupstate_create             groupstate_delete
    groupstate_enomem             groupstate_insert
    groupstate_preupdate_fail     groupstate_update
    indextable_factory_enomem     indextableload_factory_enomem
    indextablesave_factory_enomem instance_name
    instance_uuid                 memberstate_create
    memberstate_delete            memberstate_enomem
    memberstate_expired           memberstate_factory_enomem
    memberstate_unhealthy         monitor_receive
    monitor_respond               node_name
    node_uuid                     process_name
    refresh_enomem                refreshclient_create
    refreshclient_delete          refreshserver_create
    refreshserver_delete          remote_dirs
    remote_files                  snapclient_create
    snapclient_delete             snapcoord_create
    snapcoord_delete              snapserver_create
    snapserver_delete             snapserver_fail_fence_down
    snapserver_fail_fence_raise   snapserver_fail_snapid
    snapshot_create               snapshot_enomem
    snapshot_restore              tally_enomem
    vldb_enomem                   vldb_enorecord
    vldbclient_create             vldbclient_delete
    vldbclient_factory_enomem
```

The counters are specific to the FlexGroup volume, measuring remote allocation percentages, number of local versus remote files and directories, refresh counters, and various other objects.

FlexGroup statistics can be captured in the same way that other statistics are captured. You must start a statistics collection with `statistics start`, which creates a sample_id file. After this is done, the statistics can be viewed by using `statistics show`.

If you want to specify multiple objects or counters, use a pipe symbol (|).

Example of `statistics start` for FlexGroup and NFSv3 statistics:

```
cluster::> set diag
cluster::*> statistics start -object nfsv3|flexgroup
Statistics collection is being started for sample-id: sample_2144
```

Example of `statistics show` for FlexGroup counters:

```
cluster::*> statistics show .object flexgroup

Object: flexgroup
Instance: 0
Start-time: 8/9/2016 13:00:22
End-time: 8/9/2016 15:22:29
Elapsed-time: 8527s
Scope: node1
    Counter                                                     Value
    -------------------------------- --------------------------------
    cat4_fil_local                                            3623435
    cat4_fil_remote                                            600298
    groupstate_analyze                                         293448
    groupstate_update                                        59906297
    instance_name                                                   0
    node_name                                                   node1
    process_name                                                    -
    refreshclient_create                                       146724
    refreshclient_delete                                       146724
    refreshserver_create                                       146724
    refreshserver_delete                                       146724
    remote_files                                                   10
```

For more information about statistics command, use the `man statistics start` command in the CLI.

Qtree Statistics

Starting in ONTAP 9.5, qtree statistics were 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

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

Viewing FlexGroup Ingest Distribution

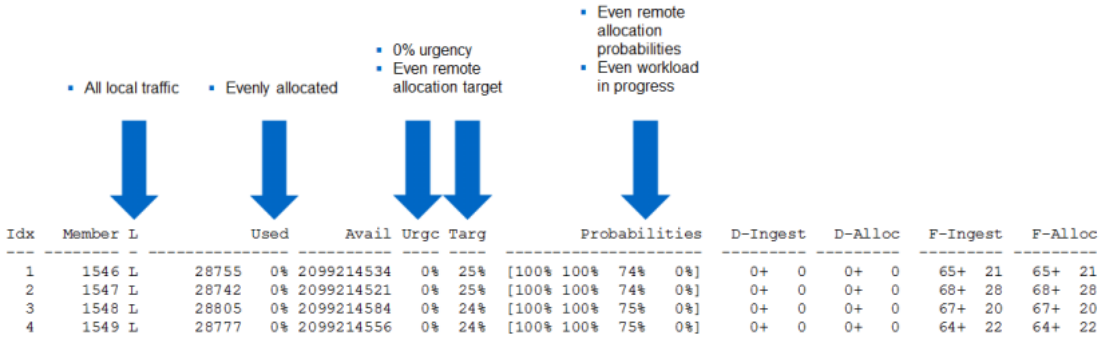
Using the command line, it is possible to get a real-time view of FlexGroup data ingest during workloads to see how evenly allocated the member volumes are with the diag privilege node-level command flexgroup show. In addition, the command provides visibility into the urgency and tolerance percentages as well as calculated probabilities for remote versus local placement of files and folders. For more information, see FlexGroup Load-Balancing Concepts.

This command can be run in the clustershell CLI across multiple nodes.

```
cluster::> set diag
cluster::*> node run * flexgroup show
```

The following graphic shows an "ideal" `flexgroup show` output in which traffic is evenly distributed (Figure 23).

Figure 23 Ideal FlexGroup ingest



Sample Python Script to Generate Files on a FlexGroup Volume

When testing a FlexGroup volume, it is possible to use normal load-generating utilities. In our lab testing, one of the benchmarks used was a basic Git benchmark using a Linux source code compile. Although that type of test is available to everyone, it might be more involved and complicated than most storage administrators want to undertake.

Conversely, it is not ideal to use common file creation utilities such as `dd` or writing bash scripts to create files and folders. These are single-threaded tests and do not fully use the benefit of the client's or storage's CPU and throughput capabilities.

One simple way to create many files and generate sufficient load on a FlexGroup volume to see its benefits is to use a Python script.

The script uses multiprocessor calls to create 1,000 directories, each with 1,000 subdirectories. Below those, the script writes five small text files, for a total of 5 million files. This script can be modified to change the number of files and folders being created.

This script is available on GitHub, but it is not officially supported by Fujitsu Support. This script is not intended to measure load generation or to max out a FlexGroup volume's performance.

FUJITSU Storage
ETERNUS AX series All-Flash Arrays,
ETERNUS HX series Hybrid Arrays
Technical Overview of ONTAP FlexGroup Volumes

P3AG-5652-01ENZ0

Date of issuance: November 2020
Issuance responsibility: FUJITSU LIMITED


FUJITSU