



FlexShare™ Design and Implementation Guide

Akshay Bhargava, Network Appliance, Inc., April 2006 | TR-3459

Executive Summary

FlexShare is a Data ONTAP® feature that provides workload prioritization for a storage system. Using FlexShare, storage administrators can confidently host different applications on a single storage system without impacting critical applications – resulting in reduced costs and simplified storage management. This paper provides details on how FlexShare works, FlexShare best practices, and high benefit use cases of FlexShare.

Table of Contents

1. Overview	3
1.1 FlexShare Definition	3
1.2 Benefits.....	3
1.3 Supported Configurations.....	3
1.4 Features	4
2. FlexShare Design	5
2.1 Basic Concepts	5
2.2 How FlexShare Schedules WAFL Operations	7
2.3 How FlexShare Manages System Resources	9
3. FlexShare Administration.....	11
3.1 FlexShare CLI Overview	11
3.2 Expected Behavior with other CLI commands	14
3.3 Manage ONTAP API	14
3.4 Upgrade and Revert.....	14
4. FlexShare Best Practices.....	15
4.1 Set Priority Configuration for All Volumes in an Aggregate	15
4.2 Configure Cluster Configuration Consistently	16
4.3 Set Volume Cache Usage Appropriately.....	17
4.4 Tuning For SnapMirror and Backup Operations	17
5. Understanding FlexShare Behavior and Troubleshooting.....	19
5.1 FlexShare Counters	19
5.2 Troubleshooting.....	21
5.3 Maintaining Priority Configurations	23
6. FlexShare High Benefit Use Cases	24
6.1 Consolidated Environment	24
6.2 Mixed Storage including FC and ATA	25
6.3 Backup/Disaster Recovery Throttling.....	26
6.4 Multiple Application Instances.....	27
7. Summary.....	30
8. Acknowledgements	30
9. Revision History	30

1. Overview

As storage requirements for enterprises continue to grow, storage administrators constantly strive to maximize return on investment (ROI) while scaling the existing infrastructure. Administrators are consistently looking for creative ways to prevent overprovisioning and to maximize the use of the existing resources. FlexShare, a built-in feature of Data ONTAP, allows storage administrators to accomplish these tasks with ease and flexibility.

FlexShare gives administrators the ability to leverage existing infrastructure and increase processing utilization without sacrificing the performance of critical business needs. With the use of FlexShare, administrators can confidently consolidate different applications and data sets on a single storage system. FlexShare gives administrators the control to prioritize applications based on how critical they are to the business.

1.1 FlexShare Definition

FlexShare is a Data ONTAP software feature that provides workload prioritization for a storage system. It prioritizes processing resources for key services when the system is under heavy load. FlexShare does *not* provide guarantees on the availability of resources or how long particular operations will take to complete. FlexShare provides a priority mechanism to give preferential treatment to higher priority tasks.

1.2 Benefits

The use of FlexShare in an environment can result in many benefits. Some of the key benefits are highlighted in the table below:

BENEFIT	FLEXSHARE DETAILS
Simplification of storage management	<ul style="list-style-type: none">▪ Reduces the number of storage systems that need to be managed by enabling consolidation▪ Provides a simple mechanism for managing performance of consolidated environments▪ Easy to administer using the same NetApp CLI and Manage ONTAP™ API
Reduction in costs	<ul style="list-style-type: none">▪ Allows increased capacity and processing utilization per storage system without impact to critical applications▪ No special hardware or software required▪ No additional license required
Flexibility	<ul style="list-style-type: none">▪ Can be easily customized to meet performance requirements of different environment workloads

1.3 Supported Configurations

FlexShare works on NetApp storage systems running Data ONTAP version 7.2 or later.

1.4 Features

FlexShare provides storage systems with the following key features:

- Relative priority of different volumes
- Per-volume user versus system priority
- Per-volume cache policies

These features allow storage administrators to tune how the system should prioritize system resources in the event that the system is overloaded.

Recommendation

The ability to control how system resources will be used under load gives the administrator an exceptional level of control. In order to take advantage of FlexShare features, a storage administrator must take the responsibility to fully understand the impact of different configuration options and optimally configure the storage system.

Before proceeding to configure priority on a storage system, it is essential to understand the different workloads on the storage system, the impact of setting priorities on the storage system, and the FlexShare best practices. Improperly configured priority settings can have undesired effects on application and system performance. The administrator should be well versed in the configuration implications and best practices.

This document is meant to help the storage administrator, providing the fundamental knowledge to configure and tune FlexShare. Understanding the key concepts and following the best practices are essential first steps.

2. FlexShare Design

This section provides an overview of the FlexShare design.

2.1 Basic Concepts

FlexShare provides the ability to assign priorities to different volumes. FlexShare also provides the ability to configure certain *per-volume* attributes, including user versus system priority and cache policies.

WAFL® Operation

A read or write request initiated from any data protocol is translated to individual read or write WAFL operations by the file system. Similarly, a system request is translated into individual WAFL operations.

Data ONTAP classifies each WAFL operation as a user or system operation based on its origin. For example, a client read request is classified as a user operation; a SnapMirror® request is classified as a system operation.

Processing Buckets

FlexShare maintains different processing buckets for each volume that has a configured priority setting. FlexShare populates the processing buckets for each volume with WAFL operations as they are submitted for execution. The processing buckets are only used when the FlexShare service is on; when the FlexShare service is off, all WAFL operations are bypassed from processing buckets and sent directly to WAFL.

Data ONTAP maintains a *Default* processing bucket. When the FlexShare service is on, all WAFL operations associated with volumes that do *not* have a FlexShare priority configuration are populated in the *Default* processing bucket; all WAFL operations for a volume that has a FlexShare priority configuration are populated into a dedicated bucket.

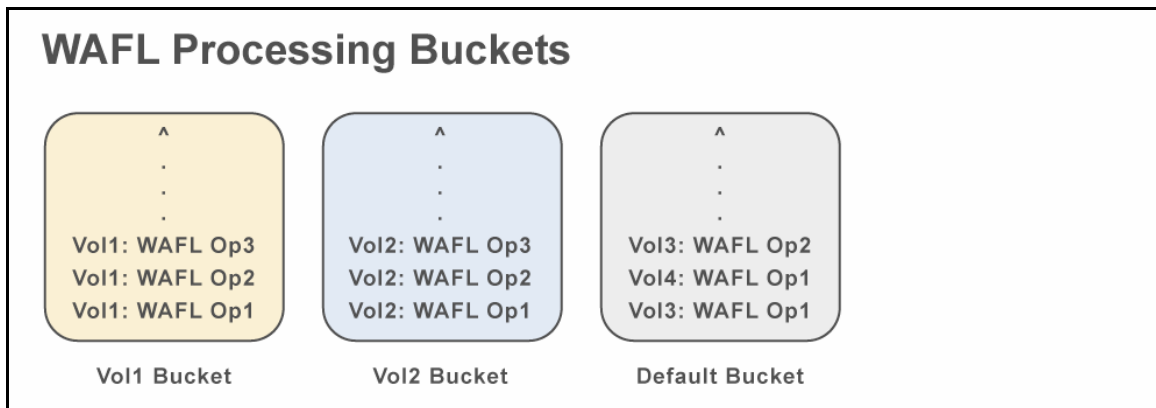


Figure 1) WAFL Processing Buckets

The figure above shows the WAFL processing buckets for Vol1, Vol2, and the Default Bucket. Vol1 and Vol2 have FlexShare priority configurations and as a result have dedicated processing buckets.

User versus System

FlexShare provides a configuration option for user vs. system priority. This allows system-initiated operations to be controlled relative to user-initiated operations. This configuration is available on a per-volume basis.

FlexShare determines whether a WAFL operation is a user or system operation based on the origin of the WAFL operation. If the origin of a WAFL operation is a data access protocol, the operation is considered to be a user operation. All other WAFL operations are considered system operations.

The table below lists some important user and system operations.

USER OPERATIONS	SYSTEM OPERATIONS
Data access operations using: <ul style="list-style-type: none">▪ NFS▪ CIFS▪ iSCSI▪ FCP▪ HTTP▪ FTP	<ul style="list-style-type: none">▪ SnapMirror▪ SnapVault®▪ WAFL Scanners▪ vol clone/vol split▪ SnapRestore®▪ NDMP

Buffer Cache Policies

Data ONTAP uses the cache to store buffers in memory for rapid access. When the cache is full and space is required for a new buffer, Data ONTAP uses a modified least-recently-used (LRU) algorithm to determine which buffers should be discarded from the cache.

FlexShare can modify how the default buffer cache policy behaves by providing hints for the buffers associated with a volume. FlexShare provides hints to Data ONTAP by specifying which information should be kept in the cache and which information should be reused.

FlexShare caching policies, if configured properly based on application workloads, can significantly enhance overall system performance. The buffer cache policy configuration is based on a per-volume setting.

Modes of Operation

The following modes of operation are available with FlexShare.

1. FlexShare service is off.

By default, the FlexShare service is off. The system behavior is identical to previous versions of Data ONTAP when FlexShare was not available.

2. FlexShare service is on; no individual priorities set.

When FlexShare service is enabled, FlexShare provides equal priority to all volumes and equal user versus system priority. FlexShare continues to use the default caching policy.

3. FlexShare service is on; individual volume priorities set.

When FlexShare service is on and one or more individual volume priorities are set, FlexShare begins to prioritize operations between different volumes.

2.2 How FlexShare Schedules WAFL Operations

FlexShare impacts the order in which WAFL operations are processed by the storage system. FlexShare determines the order WAFL operations will be processed based on the priority configuration. FlexShare gives higher priority to WAFL operations originating from higher priority volumes.

When the FlexShare service is on, the prioritization processing described in this section is always in effect.

Volume Level Priorities

The impact of FlexShare volume level priority can best be understood by comparing one storage system with the FlexShare service off with a second storage system with the FlexShare service on.

When the FlexShare service is *off*, the system processes the requests in the order in which they arrive.

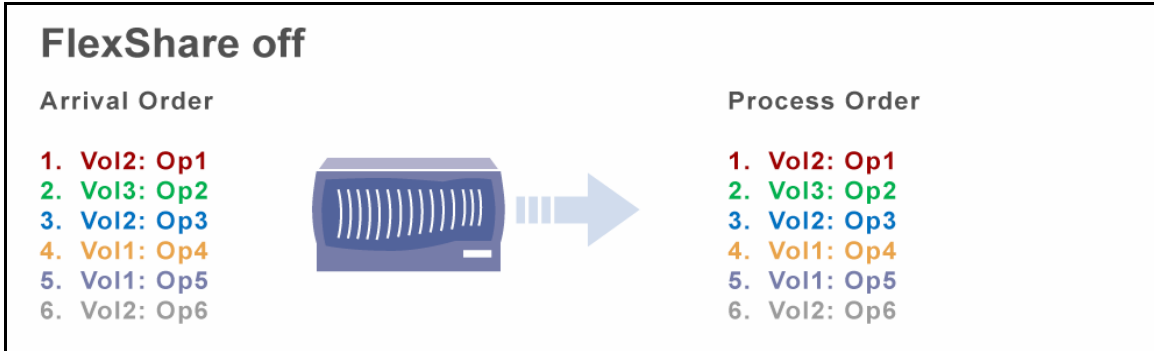


Figure 2) FlexShare off

The figure above depicts the order in which tasks arrive to be processed and the order in which they are processed by a storage system. The order of tasks processed is exactly the same as the order in which tasks arrive.

When FlexShare service is *on*, FlexShare intelligently chooses the order tasks are processed to best meet the priority configuration. On average, FlexShare is more likely to pick a WAFL operation originating from a high priority volume than a WAFL operation originating from a low priority volume. FlexShare ensures that all WAFL operations will be processed regardless of the priority configuration, but FlexShare is more likely to choose higher priority operations to be processed before lower priority operations.

Figure 3 provides a simple example of how FlexShare can impact the order in which tasks are processed based on the priority level configurations.

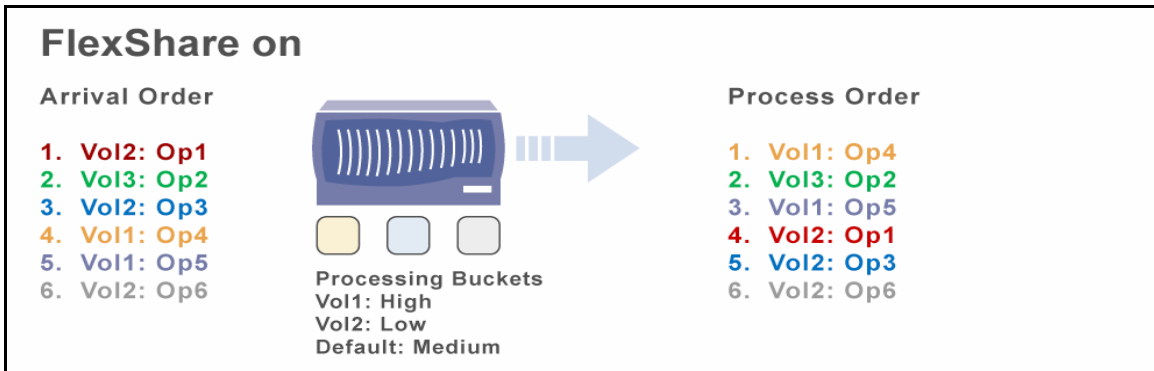


Figure 3) FlexShare on

The figure depicts a *possible* ordering of tasks when FlexShare service is on. The order tasks arrive is different than the order tasks are processed by the storage system. FlexShare orders tasks for processing taking into account the priority configuration. In this example, Vol1 has higher priority configuration than the other volumes and therefore its WAFL operations are preferentially processed.

Volume Level and System Priorities

FlexShare orders WAFL operations to be processed based on the following:

1. The configured volume priority
2. The configured user versus system priority

The order of the steps above is important in determining when WAFL operations are executed. First, the WAFL operations are prioritized based on the volume priorities. The priority of the processing buckets, which contain the WAFL operations, is the first factor that is considered. Second, the WAFL operations are prioritized based on the configured user versus system priority. The items in the individual processing buckets are ordered with respect to the user versus system priority.

Figure 4 and Figure 5 depict an example of how FlexShare chooses WAFL operations to execute based on the priority level and system configurations.

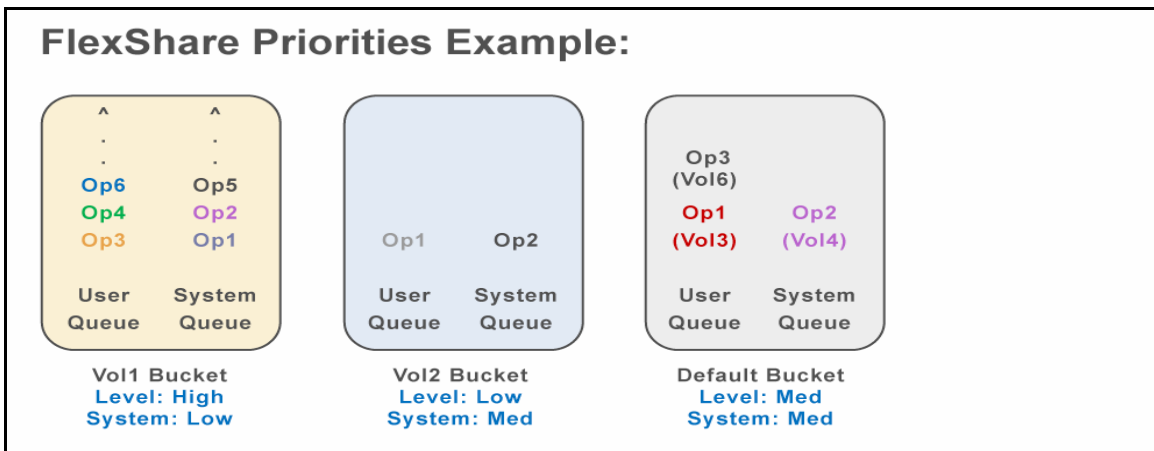


Figure 4) FlexShare Priorities

The figure above depicts what the processing buckets could look like for a storage system as they arrive for processing. Vol1 is configured with a high priority level and low system priority. Vol2 is configured with low priority level and medium system priority. Vol1 and Vol2 are the only volumes that have FlexShare priority configurations and as a result have dedicated processing buckets.

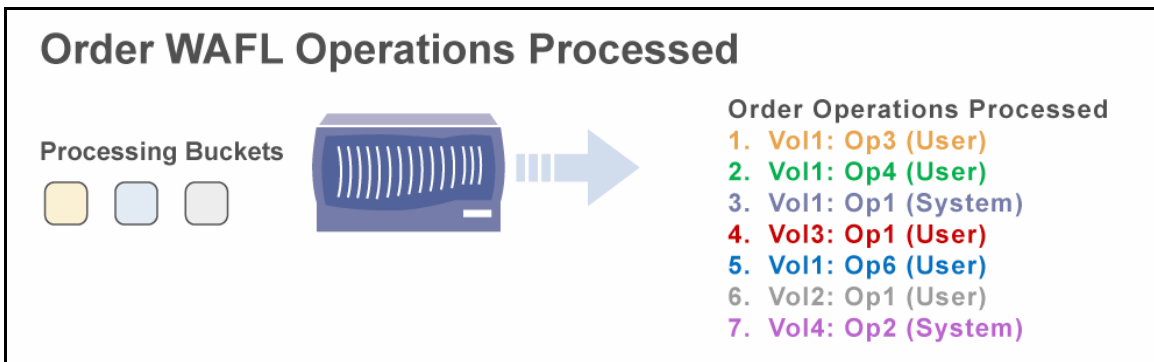


Figure 5) Order WAFL Operations Processed

The figure shows a *possible* order FlexShare would choose to process the WAFL operations from Figure 2. FlexShare orders the operations to be processed given the relative volume priorities and per-volume user versus system priority.

The example in Figure 4 and Figure 5 highlights some important points about the heuristics that FlexShare uses.

- FlexShare provides relative priority based on the volume priority configurations for the different volumes. FlexShare will preferentially choose WAFL operations to be processed from Vol1 before any other volume since Vol1's priority is set the highest.
- FlexShare takes into account the User versus System priority on a per-volume basis. Out of the WAFL operations processed for Vol1, FlexShare will preferentially choose User operations before System operations since the system priority for Vol1 was set to Low. FlexShare can choose User operations from the Vol1 processing bucket even if they were added to the bucket *after* System operations. For example, FlexShare chose to process "Vol1: Op3 (User)" earlier even though this operation was added to the Vol1 Bucket after "Vol1: Op1 (System)".
- FlexShare will choose lower priority operations before higher priority operations, but this happens less frequently.

Impact of WAFL Operation Scheduling

The impact of FlexShare rescheduling WAFL operations is generally only noticeable when the system is under heavy load. If the system is not loaded, the number of outstanding operations is small enough that FlexShare prioritization will not noticeably impact the system performance. To better understand this, imagine there is only one outstanding WAFL operation that needs to be processed out of all the processing buckets. In this case, FlexShare will not have to do any intelligent prioritization. It will simply pick the one outstanding WAFL operation to be processed. The order of processing items can have a significant impact to end users only when the system is loaded and there are sufficient items in the different processing buckets.

FlexShare does not impact the running time of WAFL operations. Once a WAFL operation is dispatched to execute, FlexShare work with the WAFL operation is complete. If there is a WAFL operation that has been dispatched or is already in progress, FlexShare will not interrupt that WAFL operation even if higher priority WAFL operations arrive in the system. FlexShare only controls the order in which WAFL operations are dispatched to be processed, but once they are dispatched they are out of the control of FlexShare.

2.3 How FlexShare Manages System Resources

FlexShare automatically controls how system resources are used by the storage system based on the volume priority level and per-volume buffer cache policy configurations. The storage administrator does not have to configure any other options to take advantage of the system resource management FlexShare provides.

FlexShare has several mechanisms to control how system resources are used. The WAFL operation ordering contributes to how system resources get used, but it is not the only means. FlexShare employs cache management and other intelligent schemes to control the different system resources.

FlexShare prioritizes, but does not guarantee the availability of system resources. FlexShare does not pre-partition or exclusively reserve system resources.

Cache Management

FlexShare provides hints to the Data ONTAP buffer cache manager by specifying which information should be kept in the cache and which information should be reused. FlexShare provides the following important information to the buffer cache manager:

- FlexShare recommends that data items in the cache that originated from a volume with a *reuse* setting be the first to be removed from the cache. Buffers containing user data are proactively aged as soon as the data has been sent to the client.
- FlexShare recommends that data items in the cache with a *keep* setting be preferentially kept in the cache.

The buffer cache manager preferentially keeps the items in the cache marked with a keep setting. However, if the cache is full and all items in the cache have a keep setting, the least-recently-used data item will be removed from the cache. It is important to note that cached data with a keep setting can be removed from the cache if the cache is full and all items in the cache belong to volumes with a keep configuration.

For optimal performance, it is recommended to set the volume cache policies appropriately. Refer to *Section 4: FlexShare Best Practices* for more information.

System Resource Usage

FlexShare prioritization controls the following system resources:

- CPU
- Disk I/O
- NVRAM
- Memory

This section highlights the mechanism by which FlexShare controls the critical system resource usage.

CPU

Higher priority volumes have their WAFL operations preferentially scheduled for CPU processing. This is primarily impacted based on how FlexShare controls the order in which the WAFL operations are chosen to execute. Refer to *Section 2.2: How FlexShare Schedules WAFL Operations* for more details on how FlexShare processes the volume and system priorities.

Disk I/O

Higher priority volumes are allowed more concurrent disk reads than lower priority volumes. FlexShare maintains a maximum number of concurrent disk reads allowed per-volume. The higher the priority of the volume, the higher the maximum number of concurrent disk reads allowed.

The amount of concurrent disk reads is automatically set based on the volume level priority. The amount of concurrent disk reads for a volume can be viewed using the advanced counters described in Section 5.

WAFL uses NVRAM to keep a log of write requests that need to be written to disk. During a Consistency Point (CP), the same set of data is copied from system memory to disk. FlexShare prioritizes disk writes by controlling how NVRAM is used. This is described in the next section.

NVRAM

FlexShare controls the amount of NVRAM consumption based on volume priority. A volume's priority dictates how much NVRAM can be consumed relative to other volumes. This is essential in maintaining priority for writes during a Consistency Point (CP) operation. If a low priority volume has exhausted its amount of writes allocated for NVRAM, it will have to wait until the current CP is completed. High priority volumes have significantly larger NVRAM limits and, therefore, their writes are generally unaffected during a CP.

The amount of NVRAM consumption is automatically set based on the volume level priority. The amount of NVRAM consumption for a volume can be viewed using the advanced counters described in Section 5.

Memory

The memory consumption is dictated by the configured buffer cache policy for the volume. This is described in detail with the description of the cache management.

3. FlexShare Administration

FlexShare can be administered using the CLI or the Manage ONTAP API. This section describes the important configuration and status commands for the CLI, the CLI commands that impact FlexShare configuration, and details about the Manage ONTAP API.

The content in this section provides an overview of the typical commands and options. Comprehensive details on the FlexShare CLI are provided in the *System Administration Guide*. Refer to the *System Administration Guide* or the *na_priority* man page for additional information.

The default values that are assigned when FlexShare is initially enabled are:

- Volume Level: Medium
- System: Medium
- Cache: default

FlexShare configuration can be dynamically changed at any time the system is running. Configuration changes take effect as soon as they are issued on the system. There is no overhead to change configuration options. Configuration changes stay active across system reboots. The default values assigned by FlexShare can be modified as well.

3.1 FlexShare CLI Overview

The *priority* command is the CLI command that provides all configuration and status information related to FlexShare.

Basics

Issue the *priority* command without any arguments to display the priority command options.

```
NetApp1> priority
The following commands are available; for more information
type "priority help <command>"
delete           off                set                show
help             on
```

Use the *help* option to find out more information about a command.

```
NetApp1> priority help on
priority on
- Start priority scheduler.
```

Refer to the *na_priority* man page or the *System Administration Guide* for detailed information.

```
NetApp1> man na_priority
na_priority(1)                                na_priority(1)

NAME
    na_priority - commands for managing priority scheduling.

SYNOPSIS
    priority command argument ...

DESCRIPTION
    The priority family of commands manages the priority
    scheduling policy on an appliance.
    .
    .
    .
```

Enable Service

To see the status of the FlexShare service, use the *show* command.

```
NetApp1> priority show
Priority scheduler is stopped.

Priority scheduler system settings:
    io_concurrency: 8
```

The FlexShare service is off by default.

Note: The *io_concurrency* setting displayed in the *priority show* output represents the average number of concurrent suspended operations per disk for a volume. This is an advanced option and should not be modified unless recommended by NetApp personnel.

To enable FlexShare service, use the *on* command.

```
NetApp1> priority on
Priority scheduler starting.
NetApp1> Fri Mar 17 22:07:11 GMT [wafl.priority.enable:info]: Priority scheduling is
being enabled
```

To verify the FlexShare service is enabled, use the *show* command.

```
NetApp1> priority show
Priority scheduler is running.

Priority scheduler system settings:
    io_concurrency: 8
```

To disable FlexShare service, use the *off* command.

```
NetApp1> priority off
Priority scheduler has stopped.
NetApp1> Fri Mar 17 22:52:28 GMT [wafl.priority.disable:info]: Priority scheduling
is being disabled
```

Priority Settings

The *set* command is used to configure volume priorities. Configuration for *level*, *system*, and *cache* can be specified. At least one configuration option from *level*, *system*, or *cache* must be specified. Options that are not explicitly set inherit the default setting.

The *level* option is configured on a per-volume basis. A volume with a higher priority level will be given more resources than a volume with a lower priority level.

The *system* option is configured on a per-volume basis. It controls the balance of system versus user priority given to a volume.

Valid *level* and *system* options include:

- VeryHigh
- High
- Medium
- Low
- VeryLow

The *system* option can also take a number as a numeric percentage from 1 to 100 for the system priority.

The *cache* option is configured on a per-volume basis. It controls the buffer cache policy for the volume.

Valid cache options include:

- reuse
- keep
- default

The example below sets the volume level priority to High. The volume will inherit the default settings for *system* and *cache*.

```
NetApp1> priority set volume vol1 level=High
NetApp1> priority show volume -v vol1
Volume: vol1
    Enabled: on
    Level: High
    System: Medium
    Cache: n/a
```

Note: The *Cache: n/a* output represents the default cache configuration.

The example below explicitly sets the level, system, and cache configuration.

```
NetApp1> priority set volume vol2 level=Low system=Low cache=reuse
NetApp1> priority show volume -v vol2
Volume: vol2
    Enabled: on
    Level: Low
    System: Low
    Cache: reuse
```

FlexShare maintains a default configuration that applies to the *Default* processing bucket. All the volumes with priority configurations inherit the default settings unless explicitly configured.

```
NetApp1> priority show default -v
Default:
    Level: Medium
    System: Medium
```