QSAN
Official Document

QoS (Quality of Service)

White Paper

July 2022

ANNOUNCEMENT

Copyright

QSAN believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

Trademarks

- QSAN, the QSAN logo, QSAN.com, XCubeFAS, XCubeSAN, XCubeNXT, XCubeNAS, XCubeDAS, XEVO, SANOS, and QSM are trademarks or registered trademarks of QSAN Technology, Inc.
- Microsoft, Windows, Windows Server, and Hyper-V are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark of Linus Torvalds in the United States and/or other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Mac and OS X are trademarks of Apple Inc., registered in the U.S. and other countries.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- VMware, ESXi, and vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other countries.
- Citrix and Xen are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries.
- Other trademarks and trade names used in this document to refer to either the entities claiming the marks and names or their products are the property of their respective owners.

TABLE OF CONTENTS

FIGURES

TABLES

NOTICES

Information contained in this document has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

PREFACE

Executive Summary

QoS (Quality of Service) ensures that a particular application always gets a specific predefined performance level. Adjusting bandwidth and performing automatic tuning of the I/O performance makes sure the limitation of the throughput, IOPS per volume or required response time per application will be achieved.

Audience

This document is applicable for QSAN customers and partners who are interested in learning about QoS to ensure that your applications work smoothly and consistently performance at anytime. It assumes the reader is familiar with QSAN products and has general IT experience, including knowledge as a system or network administrator. If there is any question, please refer to the user manuals of products, or contact QSAN support for further assistance.

Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: https://www.qsan.com/technical_support
- Via Telephone: +886-2-77206355
- (Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Skype ID: qsan.support
- (Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com

Information, Tip, and Caution

This document uses the following symbols to draw attention to important safety and operational information.

INFORMATION

INFORMATION provides useful knowledge, definition, or terminology for reference.

TIP

TIP provides helpful suggestions for performing tasks more effectively.

CAUTION

CAUTION indicates that failure to take a specified action could result in damage to the system.

1. OVERVIEW

QoS is a term used in the field of packet switched networks to refer to the probability that a network will satisfy a given business contract. In many cases, informally refers to the probability that a packet will pass between two points in the network. QoS has a control mechanism that provides different priorities for different users or different data streams, or ensures that the performance of the data stream reaches a certain level according to the requirements of the application.

QoS features exist in everything from network devices to hypervisors to storage. When multiple workloads share limited resources, QoS helps control how resources are shared and prevents the noisy applications from damaging the performance of all other applications on the same system.

In the early days of the Internet, it didn't realize the needs of QoS application, so the entire Internet operation was like a "best-effort" system. According to the sender and the receiver, many things happen during the packet transmissions from the start point to the end point, and the following problematic results are produced:

- **Lost packet:** When the packet arrives at a buffer full, it means that the transmission failed. The system will decide to discard part, or discard all packets according to the status of the network, and the application at the receiving end must request retransmission at this time, which may also cause a serious delay in the overall transmission.
- **Latency:** It may take a long time to transfer the packet to the destination because it can be delayed by a long queue. In short, the delay is very difficult to predict.
- **Incorrect transmission sequence:** When a group of related packets are routed through the Internet, different packets may choose different ways, which will result in different delays for each packet. The order in which the last packet arrives at the destination is inconsistent with the order in which the packet is sent from the sender.

Although some specific forms of network data flow need to define service quality. QoS is developed to set priorities for data flows and set guaranteed limits of effectiveness, throughput, latency, etc. QoS guarantees are important for networks with limited bandwidth, especially for streaming multimedia applications such as VoIP (Voice over Internet Protocol) and IPTV (Internet Protocol Television), as these applications often require a fixed transmission rate and are sensitive to latency.

1.1. QoS in Storage System

The same situation exists in the storage system. QoS refers to service priority and resource reservation control mechanisms. QoS is the ability to provide different priorities for different volumes to guarantee a certain level of performance for a data stream.

Generally, a single service does not need QoS because it has the whole system resources. But for business-critical applications in multi-tenant, QoS is a key technology to deliver consistent primary storage performance and this predictable performance is also an important indicator.

1.2. Benefit of QoS

If different applications share the same storage platform, we can predict that the following issues may occur:

- Workloads with I/O conflicts, such as OLTP (Online Transaction Processing) and data warehousing.
- Tiering storage access restrictions, such as development and production applications.
- Peak demand processing for critical applications and maintenance activities.

Although raw storage performance is important, predictable and consistent performance delivery ensures that every application has the required resources to run without interruptions. QoS feature enables the storage system when serving these workloads.

- Prioritize the storage bandwidth and capacity resources.
- Consistently deliver predictable performance across multiple applications.
- Reduce unpredictable I/O patterns.
- Eliminate manual adjustments to adapt to changing workload demands.
- Allow application scale to grow without disrupting the system.

QoS feature makes it easier for users to assign I/O priorities. And it ensures a better service consolidation when serving these workloads.

1.3. QoS Use Cases

QoS is best for multi-tenant systems running different workloads with unpredictable requirements. Multi-tenant systems allow the flexibility to meet your business needs. Therefore, QoS is particularly effective at the following applications:

- Mission critical database applications: real-time transactional database (Exchange server, SQL server, RDBMS, DB2), OLTP (On-Line Transaction Processing).
- Consolidation a virtualization: virtualized data center, VMware, Hyper-V, Citrix.
- Private cloud computing and big data analytics.

2. THEORY OF OPERATION

Mention the evolution of the storage system, in order to meet different application performance requirements, the storage industry developed SSD cache and auto tiering technologies based on the traditional disk drives. These solutions apply complex algorithms and predictive methods to transfer data to the right media at the right time to improve performance. These approaches are costly and complex, and only apply to some form of certain data distribution, but may not meet the predictable performance requirements of mission-critical applications.

QoS is developed to resolve this difference. Each volume is configured with priority, target response time, or maximum IOPS and maximum throughput values that are strictly enforced within the system. The priority and target response time provides a guarantee for performance, as system resources allow. The maximum IOPS and maximum throughput values limit the allocation of performance and deliver consistent performance to workloads. The QoS feature ensures that multiple applications can run on a single storage platform without affecting each other.

After enabling QoS feature, we provide two options to configure. One is priority and target response time; the other is maximum IOPS and maximum throughput. Describe on the following.

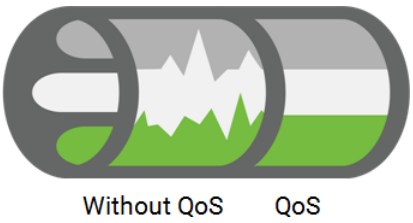

Figure 2-1 Without and With QoS

2.1. Priority and Target Response Time

Set performance limits for each connected server based on priority. By prioritizing data access and dynamically managing any I/O conflicts, you can guarantee the high performance of high-priority applications while using capacity more efficiently.

To define the QoS volumes two parameters can be set:

1. **Priority:** Defines the service level of the volume that represents its priority. The volume priority can be set to high, medium, or low.
2. **Target response time:** Define the performance of each volume as the average of the read and write response times for that volume. The volume with high priority can be set the target response time. After setting, the performance of the volume is adjusted based on the defined target response time.

The QoS mechanism shares less I/O resources to the volumes with low priority. Volumes of higher priorities are granted greater processing I/O capabilities and achieve better response times as they approach the target.

INFORMATION

The volume which sets to the high priority can be set the target response time.

TIP

When the target response time is set properly and the system resources are sufficient, the latency of the volume will reach the target response time.

The system keeps checking the current value and the target response time. It is performed bandwidth adjustment if the measured volume response time does not match the expected target performance. After achieving the performance goal, QoS will continue to monitor the latency and adjust the resources as necessary. If the performance goal is not met, please view the volume configuration and QoS settings and change the tuning parameters properly.

2.2. Maximum IOPS and Maximum Throughput

Other than the priority option, the maximum IOPS and/or the maximum throughput can be configured to limit the volume processing I/O capabilities and bandwidth. It limits the maximum resources of the volume to not exceed, and reserved for other volumes.

To define the QoS volumes two options can be chosen:

1. **Unlimited:** Defines the service level of the volume that is no limit.
2. **Maximum IOPS and Maximum Throughput:** Define the performance of each volume as the maximum IOPS and/or throughput for that volume. The threshold can be set to two or one of. After setting, the performance of the volume is adjusted based on the defined values.

The QoS mechanism monitors the IOPS and throughput of the volume. It will block the I/O processing when the threshold is reached. Volumes with these settings are predictable and consistent performance delivery.

INFORMATION

The maximum IOPS and maximum throughput settings can be set to two or one of.

3. CONFIGURE QOS

3.1. Configure QoS Settings

This section will describe the operations of enabling QoS. Select the **System** tab and the **Settings** subtab, and then click the **QoS Settings** pane to configure the QoS settings.

Figure 3-1 Settings Subtab in the System Tab

3.1.1. Operations on QoS Settings

The options are available in this pane.

Enable QoS Settings

Before using QoS, you have to enable the QoS settings.

1. Click the **QoS Settings** pane to configure the QoS settings

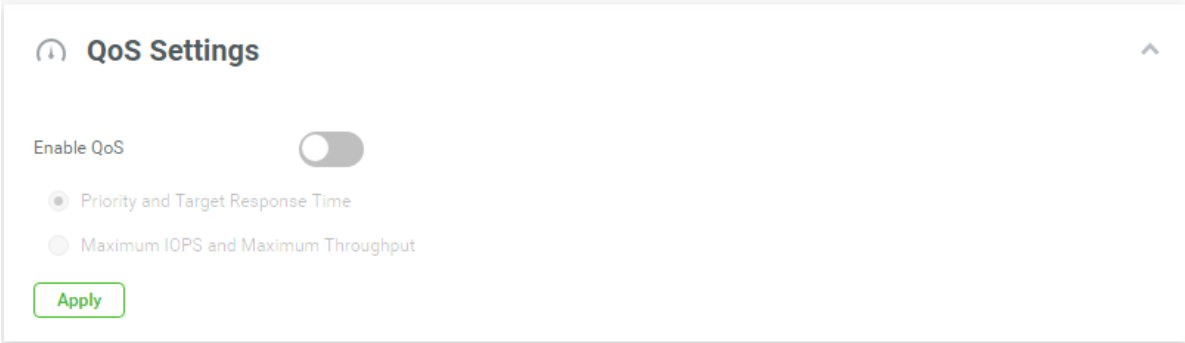

Figure 3-2 QoS Setting is Disabled

2. Click the **Enable QoS** switch to  ON (Enable) to enable.
3. Select an option of **Priority and Target Response Time** or **Maximum IOPS and Maximum Throughput**.
4. Click the **Apply** button to enable.

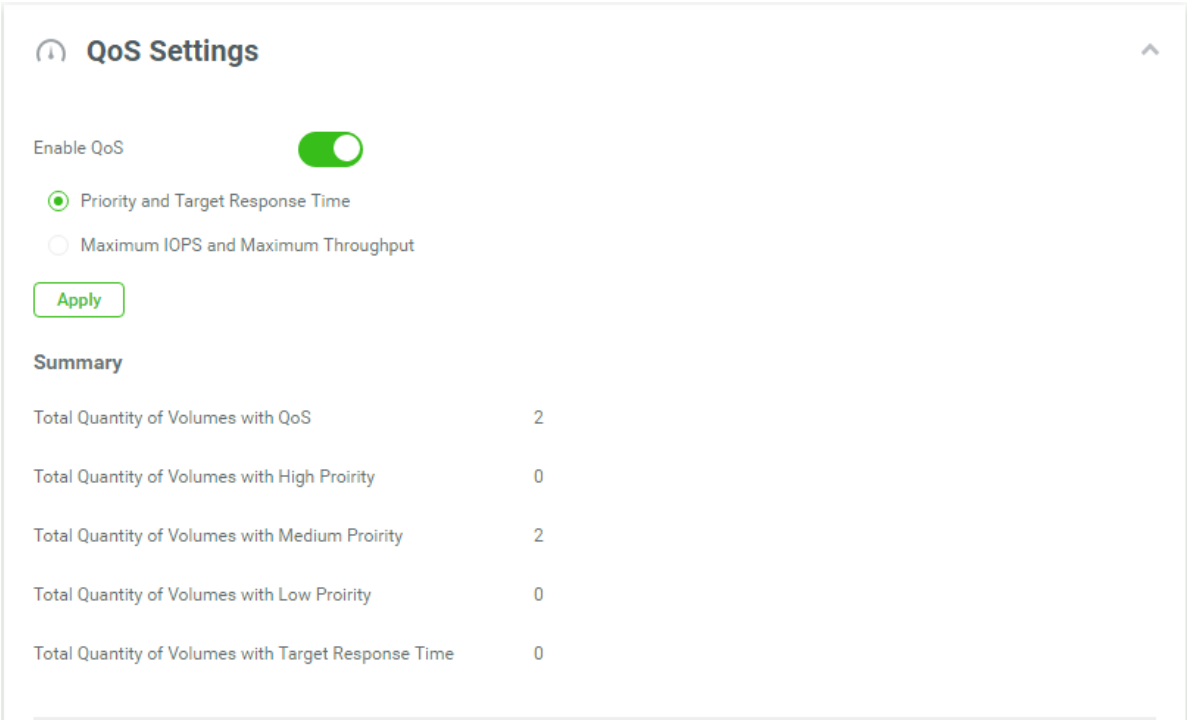

Figure 3-3 QoS Setting with Priority and Target Response

After the QoS setting is enabled, the relevant QoS summary is displayed in the page.

Change QoS Settings

The QoS settings can be changed dynamically.

1. Change the option of QoS setting.
2. Click the **Apply** button to change.

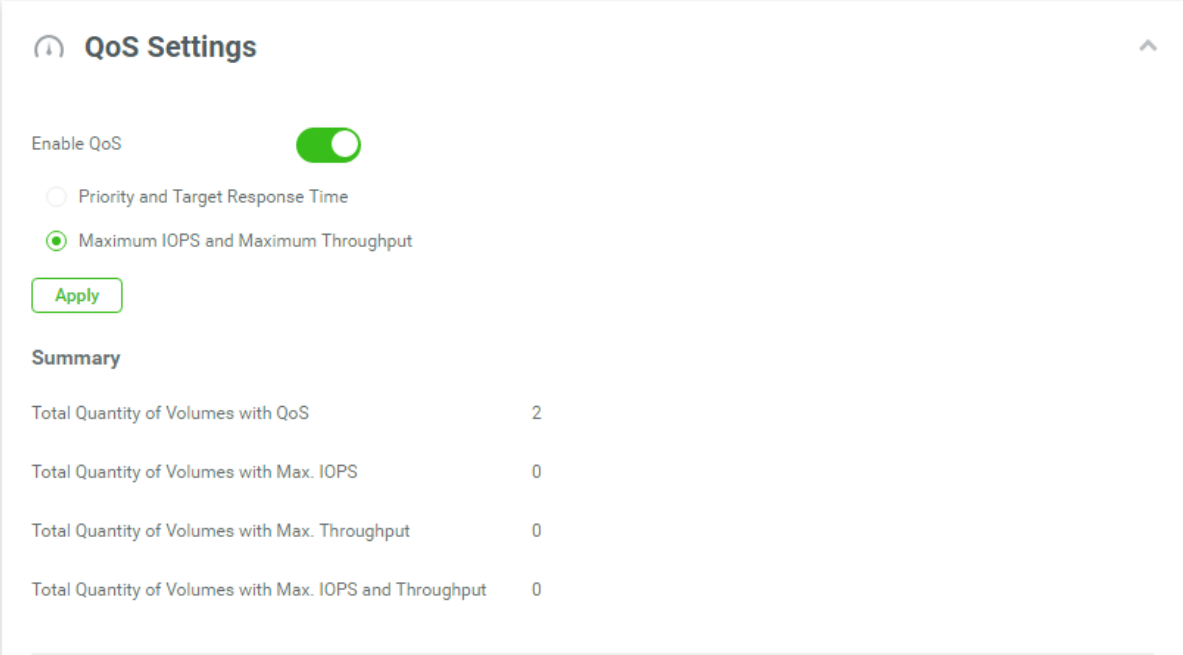

Figure 3-4 QoS Setting with Maximum IOPS and Maximum Throughput

The same, the relevant QoS summary is displayed in the page.

Disable QoS Settings

1. Click the **Enable QoS** switch to OFF (Disable) to disable.
2. Click the **Apply** button to disable.

3.2. Configure QoS Volumes

The **QoS Volumes** table is only visible when the QoS is enabled. It is displayed in the page below. In the **QoS Volumes** table, it can be displayed the status of QoS volumes, or configured the QoS volumes.

3.2.1. List QoS Volumes

The QoS settings can be configured by clicking the icon beside the volume name.

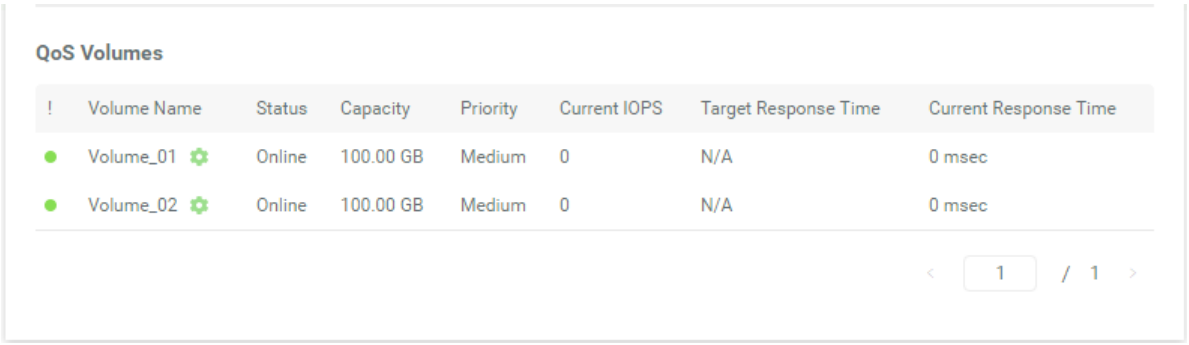

Figure 3-5 List QoS Volumes with Priority and Target Response

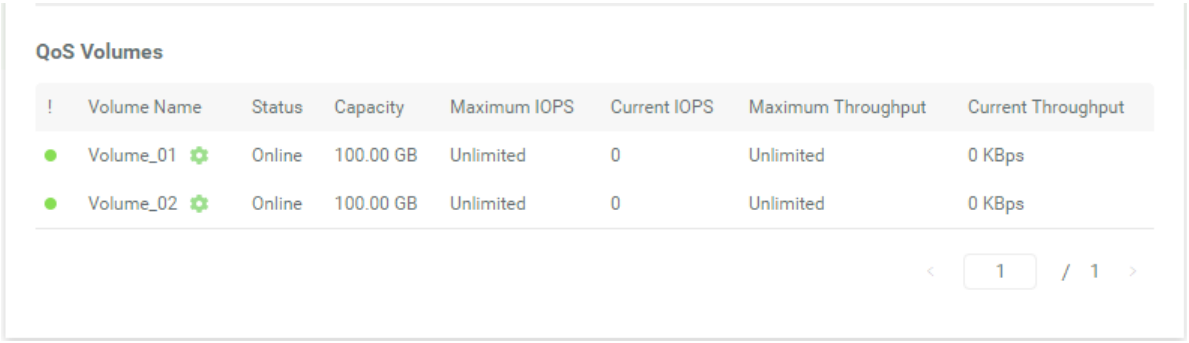

Figure 3-6 List QoS Volumes with Maximum IOPS and Maximum Throughput

This table shows the column descriptions.

Table 3-1 QoS Volume Descriptions

COLUMN NAME	DESCRIPTION
!	The status of the disk group: • Green Color / Normal: The volume is good. • Orange Color / Abnormal: The pool is unhealthy and incomplete. The cause may be a disk loss or failure. • Red Light / Warning: The disk group has failed.
Volume Name	The volume name.
Status	The status of the volume: • Online: The volume is online.

	• Offline: The volume is offline. • Erasing: The volume is being erased. • Initiating: The volume is being initialized. • Rebuilding: The volume is being rebuilt. • Migrating: The volume is being migrated. • Rollback: The volume is being rolled back. • Parity Checking: The volume is being parity check. • Relocating: The volume is being relocated. • EE Rebuilding: The volume is being RAID EE rebuilt.
Capacity	Total capacity of the volume.

Table 3-2 QoS Volume Descriptions with Priority and Target Response Time

COLUMN NAME	DESCRIPTION
Priority	The priority of the volume: • High • Middle • Low
Current IOPS	Current IOPS of the volume.
Target Response Time	Target response time setting, N/A is not configured.
Current Response Time	Current response time of the volume.

Table 3-3 QoS Volume Descriptions with Maximum IOPS and Maximum Throughput

COLUMN NAME	DESCRIPTION
Maximum IOPS	The maximum IOPS setting.
Current IOPS	Current IOPS of the volume.

Maximum Throughput	The maximum throughput setting.
Current Throughput	Current throughput of the volume.

3.2.2. Operations on QoS Volumes

Click the ⚙ icon beside the volume name to list the drop down options. These options are available in the QoS volume.

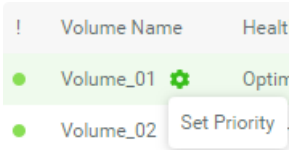

Figure 3-7 QoS Volume Options

Set Priority and Target Response Time

1. If the QoS setting is **Priority and Target Response Time**, select a volume, and then click **Set Priority** to configure.

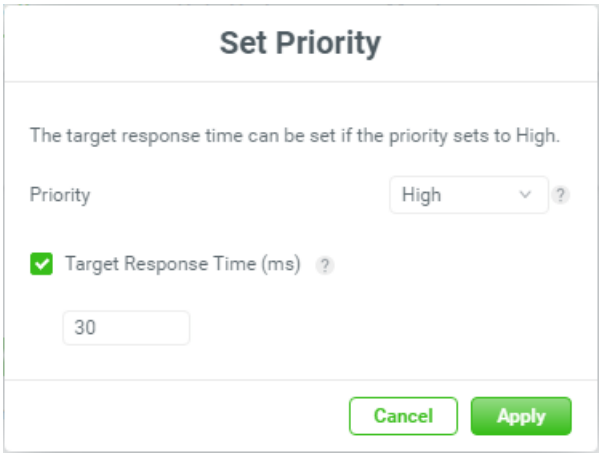

Figure 3-8 Set Priority

2. Select a **Priority** from the drop-down list. The default setting is **Middle**. The **Target Response Time** can be configured when the **Priority** is set to **High**.
3. If necessary, check the **Target Response Time** checkbox and enter a number. The number is between 1 and 10,000 milliseconds.
4. Click the **Apply** button to set.

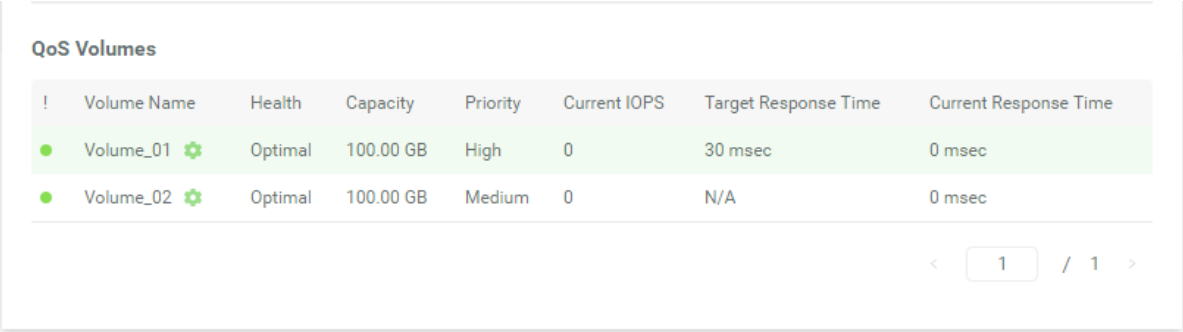

Figure 3-9 Priority and Target Response Time are Set

Set Maximum IOPS and Maximum Throughput

1. If the QoS setting is Maximum IOPS and Maximum Throughput, select a volume, and then click Set Maximum IOPS and Maximum Throughput to configure.

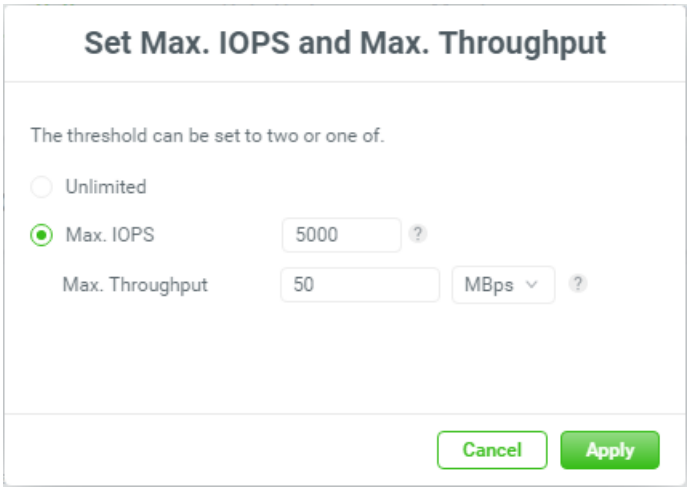

Figure 3-10 Set Maximum IOPS and Maximum Throughput

2. Select the Unlimited or Maximum IOPS and Maximum Throughput radio button.

3. If it is set to **Maximum IOPS and Maximum Throughput**, enter numbers for the maximum IOPS and/or the maximum throughput. The threshold can be set to two or one of. The maximum IOPS number is between 100 and 10,000,000. The maximum throughput number is between 50KBps and 100GBps.
4. Click the **Apply** button to set.

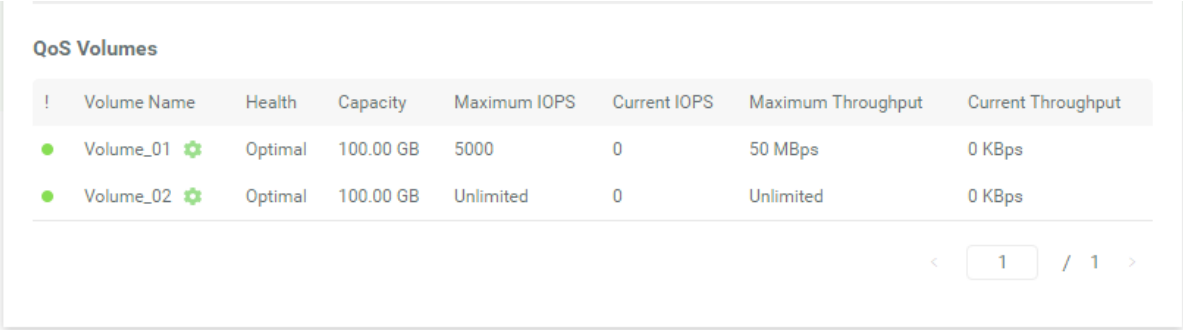

Figure 3-11 Maximum IOPS and Maximum Throughput are set

3.3. Volume Performance Monitoring

Select the Performance pane to monitor performance.

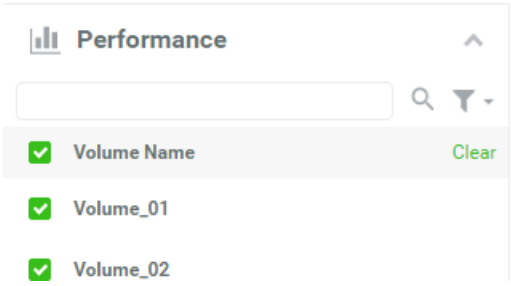

Figure 3-12 List All Volumes in the Performance Pane

All volumes are listed. You can select one or more volumes to view performance.

Figure 3-13 Performance Graph

4. TEST RESULTS

4.1. Case 1: Volumes with QoS Priority

The following example shows the adjustments that are performed with the QoS priority setting for volumes.

Test Equipments and Configurations

- Server
  - Model: Dell T630 (CPU: Intel Xeon E5-2620 v3 / RAM: 64 GB)
    FC HBA: QLogic QLE2672
    OS: Windows Server 2012 R2
- Storage
  - Model: XF2026D
    Memory: 32 GB (4 x 8 GB) per controller
    Host Card: 2 x HQ-16F4S2 (4-port 16 Gb Fibre Channel)
    Firmware 1.4.0
    HDD: 24 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
  - HDD Pool: 3 x **RAID 0** Pool with 8 x NL-SAS HDDs in Controller 1
  - HDD Volume: 3 x 100 GB (**Vol-1** in Pool-1, **Vol-2** in Pool-2, **Vol-3** in Pool-3)
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 3
  - Outstanding (Queue Depth): 128
  - Access Specifications: **Sequential Read, 64 KB**
- QoS Settings

TIME / VOL NAME	0 MIN 0 SEC	2 MIN 3 SEC	2 MIN 24 SEC
Vol -1	Middle	High	High
Vol-2	Middle	Middle	Middle
Vol-3	Middle	Middle	Low

- Test Scenario
  - First, we create three pools using RAID 0, and create one volume in each pool.
  - Enable the QoS setting to **Priority and Target Response Time**.
  - At the beginning, the QoS priority settings for three volumes are set to **Middle**.
  - After 2 minutes and 3 seconds, set the QoS priority of the Vol-1 to **High**.
  - After 2 minutes and 24 seconds, set the QoS priority of the Vol-3 to **Low**.

Test Results

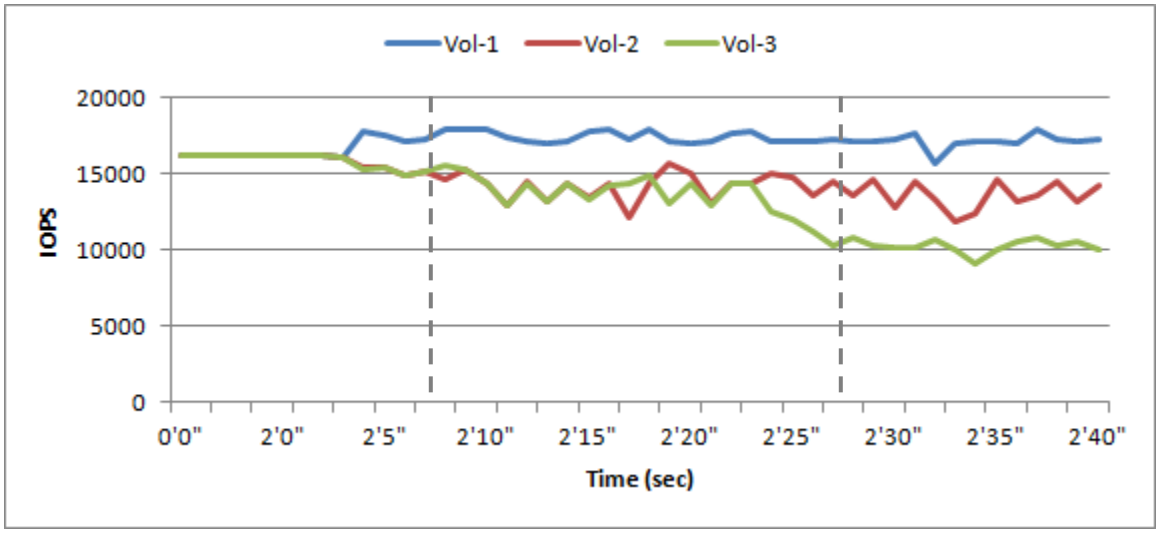

Figure 4-1 Volumes with QoS Priority

When the QoS priority of Vol-1 is set to **High** at 2 minutes and 3 seconds, Vol-1 obtains a higher IOPS and is displayed as a blue line. At 2 minutes and 24 seconds, the QoS priority of Vol-3 is set to **Low**. Therefore, the IOPS of Vol-3 is reduced and is displayed as green line.

Summary

- QoS priority results are in line with expectations.

4.2. Case 2: A Volume with Target Response Time

The following example shows the adjustments that are performed with the target response time for a volume.

Test Equipments and Configurations

- Server
  - Model: Dell T630 (CPU: Intel Xeon E5-2620 v3 / RAM: 64 GB)
    FC HBA: QLogic QLE2672
    OS: Windows Server 2012 R2
- Storage
  - Model: XF2026D
    Memory: 32 GB (4 x 8 GB) per controller
    Host Card: 2 x HQ-16F4S2 (4-port 16 Gb Fibre Channel)
    Firmware 1.4.0
    HDD: 24 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
  - HDD Pool: 3 x **RAID 0** Pool with 8 x NL-SAS HDDs in Controller 1
  - HDD Volume: 3 x 100 GB (**Vol-1** in Pool-1, **Vol-2** in Pool-2, **Vol-3** in Pool-3)
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 3
  - Outstanding (Queue Depth): 128
  - Access Specifications: **Sequential Write, 256 KB**
- QoS Settings

TIME / VOL NAME	0 MIN 0 SEC	2 MIN 3 SEC
Vol -1	Middle	High Target Response Time = 25 ms
Vol-2	Middle	Middle
Vol-3	Middle	Middle

- Test Scenario
  - First, we create three pools using RAID 0, and create one volume in each pool.
  - Enable the QoS setting to **Priority and Target Response Time**.
  - At the beginning, the QoS priority settings for three volumes are set to **Middle**.
  - After 2 minutes and 3 seconds, set the QoS priority of the Vol-1 to **High**, and set the Target Response Time to **25 ms**.

Test Results

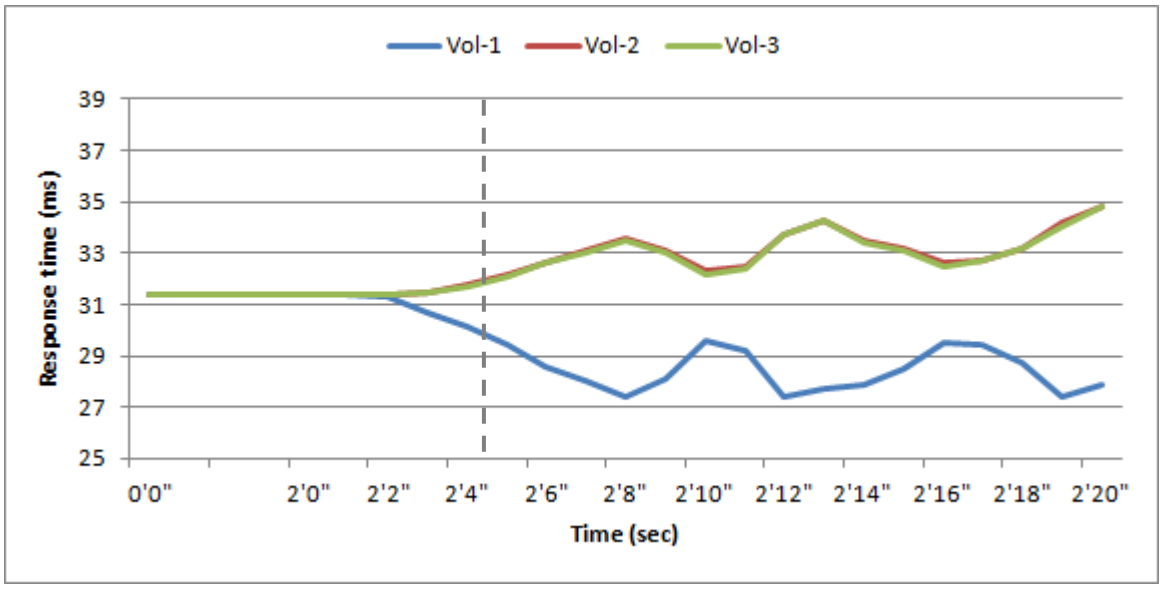

Figure 4-2 A Volume with Target Response Time

When the QoS priority of Vol-1 is set to **High** and the Target Response Time is set to **25 ms** at 2 minutes and 3 seconds, Vol-1 obtains a lower response time and is displayed as the blue line. Although Vol-1 gained more resources, it ran for approximately 28 ms due to system limitations.

Summary

- The result of configuring the volume target response time is effective.

4.3. Case 3: Bandwidth Adjustment with Maximum IOPS

The following example shows the adjustments that are performed with the maximum IOPS for volumes.

Test Equipments and Configurations

- Server
  - Model: Dell T630 (CPU: Intel Xeon E5-2620 v3 / RAM: 64 GB)
    FC HBA: QLogic QLE2672
    OS: Windows Server 2012 R2
- Storage
  - Model: XF2026D
    Memory: 32 GB (4 x 8 GB) per controller
    Host Card: 2 x HQ-16F4S2 (4-port 16 Gb Fibre Channel)
    Firmware 1.4.0
    HDD: 24 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
  - HDD Pool: 3 x **RAID 0** Pool with 8 x NL-SAS HDDs in Controller 1
  - HDD Volume: 3 x 100 GB (**Vol-1** in Pool-1, **Vol-2** in Pool-2, **Vol-3** in Pool-3)
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 3
  - Outstanding (Queue Depth): 128
  - Access Specifications: **Sequential Read, 64 KB**
- QoS Settings

TIME / VOL NAME	0 MIN 0 SEC	2 MIN 3 SEC	2 MIN 30 SEC
Vol -1	Unlimited	Unlimited	Unlimited
Vol-2	Unlimited	Unlimited	10,000 IOPS
Vol-3	Unlimited	5,000 IOPS	5,000 IOPS

- Test Scenario
  - First, we create three pools using RAID 0, and create one volume in each pool.
  - Enable the QoS setting to **Maximum IOPS and Maximum Throughput**.
  - At the beginning, the maximum IOPS settngs for three volumes are Unlimited.
  - After 2 minutes and 3 seconds, set the maximum IOPS of the Vol-3 to **5,000**.
  - After 2 minutes and 30 seconds, set the maximum IOPS of the Vol-2 to **10,000**.

Test Results

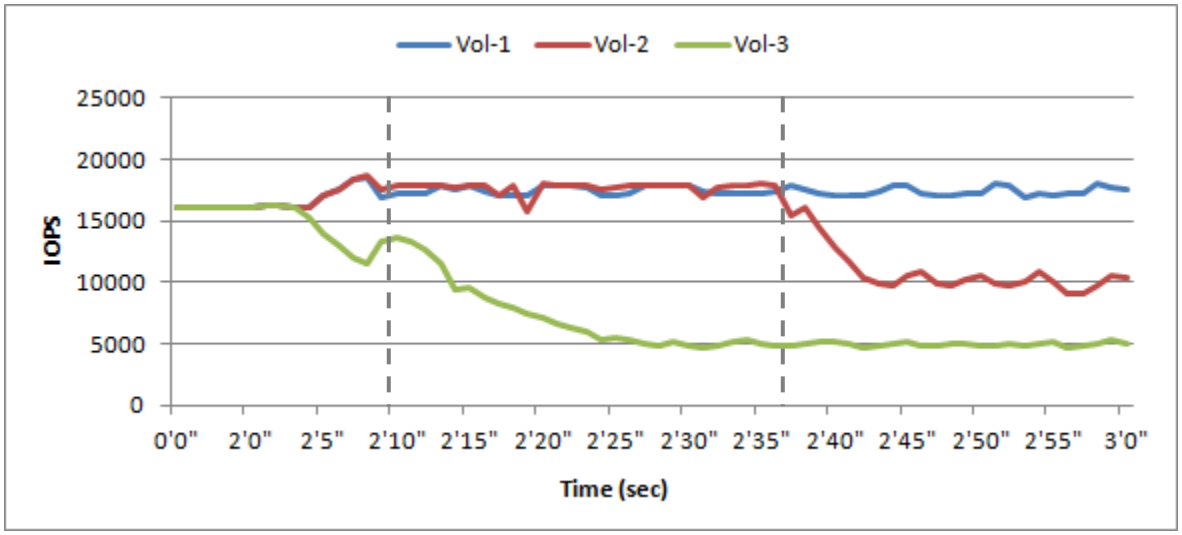

Figure 4-3 Volumes with Maximum IOPS

When the maximum IOPS of Vol-3 is set to **5,000** at 2 minutes and 3 seconds, the IOPS of Vol-3 is reduced to 5,000 and is displayed as green line. At 2 minutes and 30 seconds, the maximum IOPS of Vol-2 is set to **10,000**. The IOPS of Vol-2 is limited to 10,000 and is displayed as red line.

Summary

- The result of configuring the maximum IOPS for the volume is as expected.

4.4. Case 4: Bandwidth Adjustment with Maximum Throughput

The following example shows the adjustments that are performed with the maximum throughput for volumes.

Test Equipments and Configurations

- Server
  - Model: Dell T630 (CPU: Intel Xeon E5-2620 v3 / RAM: 64 GB)
    FC HBA: QLogic QLE2672
    OS: Windows Server 2012 R2
- Storage
  - Model: XF2026D
    Memory: 32 GB (4 x 8 GB) per controller
    Host Card: 2 x HQ-16F4S2 (4-port 16 Gb Fibre Channel)
    Firmware 1.4.0
    HDD: 24 x Seagate Constellation ES, ST500NM0001, 500 GB, SAS 6 Gb/s
  - HDD Pool: 3 x **RAID 0** Pool with 8 x NL-SAS HDDs in Controller 1
  - HDD Volume: 3 x 100GB (**Vol-1** in Pool-1, **Vol-2** in Pool-2, **Vol-3** in Pool-3)
- I/O Pattern
  - Tool: IOmeter V1.1.0
  - Workers: 3
  - Outstanding (Queue Depth): 128
  - Access Specifications: **Sequential Write, 64 KB**
- QoS Settings

TIME / VOL NAME	0 MIN 0 SEC	2 MIN 3 SEC	2 MIN 30 SEC
Vol -1	Unlimited	Unlimited	Unlimited
Vol-2	Unlimited	Unlimited	800 MB/s
Vol-3	Unlimited	400 MB/s	400 MB/s

- Test Scenario
  - First, we create three pools using RAID 0, and create one volume in each pool.
  - Enable the QoS setting to **Maximum IOPS and Maximum Throughput**.
  - At the beginning, the maximum throughput settings for three volumes are **Unlimited**.
  - After 2 minutes and 3 seconds, set the maximum throughput of the Vol-3 to **400 MB/s**.
  - After 2 minutes and 30 seconds, set the maximum throughput of the Vol-2 to **800 MB/s**.

Test Results

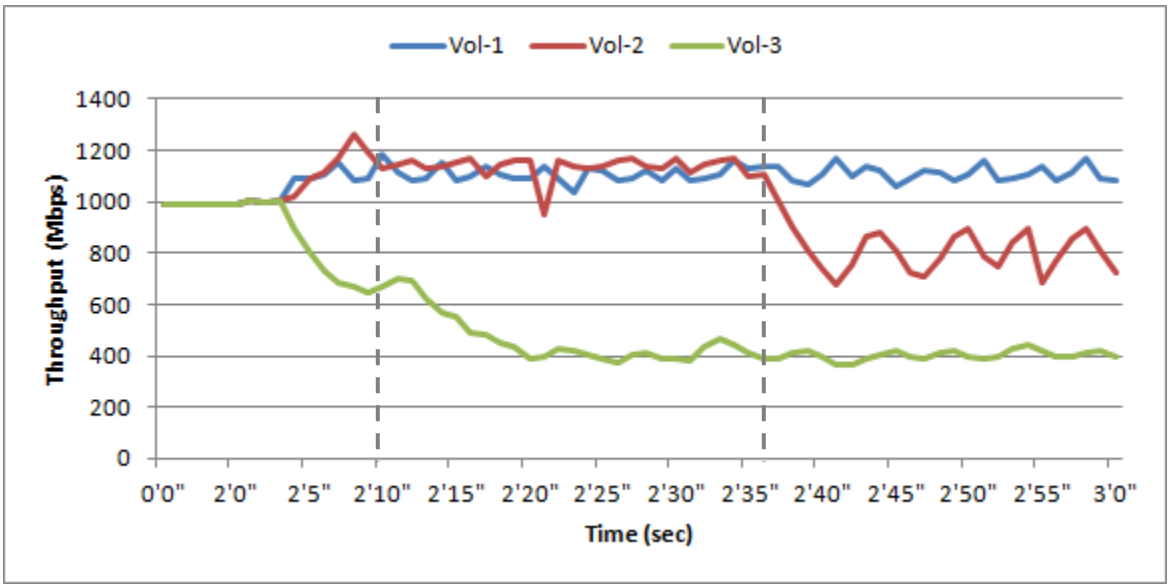

Figure 4-4 Volumes with Maximum Throughput

When the maximum throughput of Vol-3 is set to **400 MB/s** at 2 minutes and 3 seconds, the throughput of Vol-3 is reduced to 400 MB/s and is displayed as green line. At 2 minutes and 30 seconds second, the maximum throughput of Vol-2 is set to **800 MB/s**. The throughput of Vol-2 is limited to 800 MB/s and is displayed as red line.

Summary

- The result of configuring the maximum throughput for the volume is as expected.

5. INTEGRATED QTIERING

QTiering (Auto Tiering) function can be further enhanced by using QoS on a flexible tiering pool. If the target response time of the volume cannot be satisfied by the QoS function, the volume with hot data can be automatically moved to a faster tier by the auto tiering mechanism. The quota shares of a volume in different storage tiers are automatically adjusted to achieve the required response time. By integrating QoS and auto tiering, you can provide an efficient and automated way to get the best performance from your business applications.

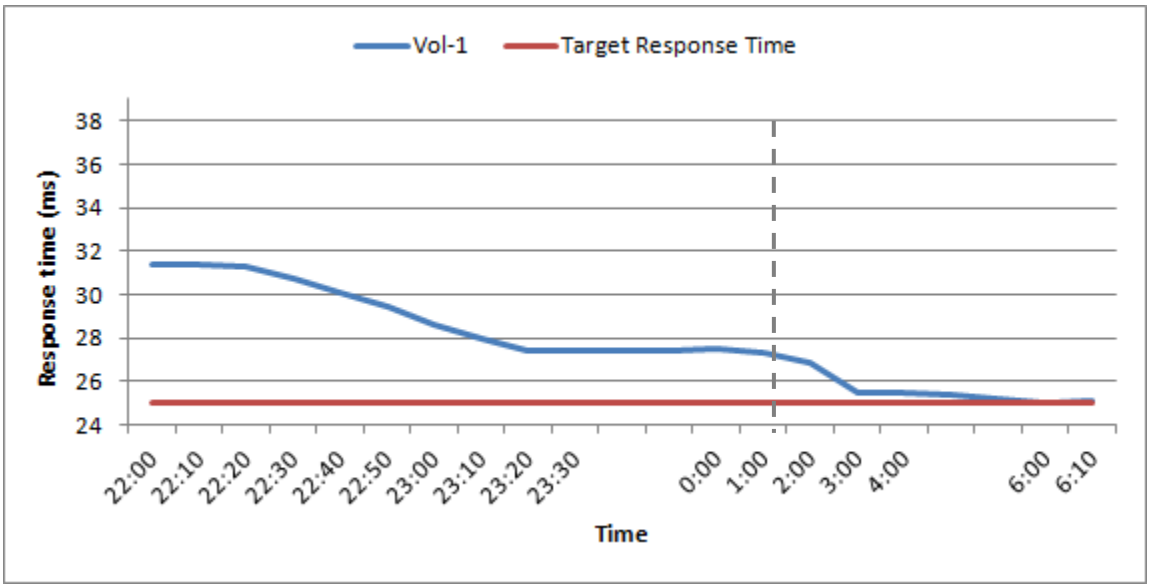

Figure 5-1 QoS Integrated with Auto Tiering

The figure shows how auto tiering affects the response time of volumes that have been adjusted by QoS. Before auto tiering relocations, the volume cannot achieve the required target response time. At 0:00, auto tiering relocations occur and hot data is relocated to faster tiers. You can see significant performance improvements over time and the volume will meet the target response time.

6. CONCLUSION

The above test cases show that using QoS can significantly adjust bandwidth and perform automatic adjustment of I/O performance to ensure throughput, IOPS per volume or the response time required per application. QoS is one of the defining principles of QSAN storage software. Customers can ensure that a particular application always gets a specific predefined performance level.

7. APPENDIX

7.1. Apply To

- XEVO firmware 1.1.0 and later
- SANOS firmware 1.4.0 and later

7.2. Reference

- XEVO Software Manual
- QoS Tutorial in XEVO
- SANOS Software Manual
- QoS Tutorial in SANOS