



White Paper

Clustered Data ONTAP Quality of Service Performance Isolation for Multi-Tenant Environments

Amrita Das, Mike McNamara, NetApp
August 2013 | WP-7191

Executive Summary

Storage quality of service (QoS) is a new feature in clustered Data ONTAP® to enable delivery of predictable storage performance for business-critical applications in a shared infrastructure environment. By assigning not-to-exceed performance capacity limits to collections of storage objects (files, LUNs, volumes), you can proactively control resources consumed by less critical or rogue workloads, thus preserving the performance of the business-critical workloads. Similarly, by assigning limits to a storage virtual machine (which maps to a tenant) in a multi-tenant shared infrastructure, a service provider could prevent tenants from affecting each other's performance, as well as enforce a "get no more than what you pay for" model. Storage QoS works well in almost any shared infrastructure environment, including enterprise data centers, private clouds, and public clouds. By enabling you to consolidate your workloads in a shared infrastructure and increase utilization, while still meeting the performance needs of your business-critical workloads, it lowers capital expenses and reduces troubleshooting activities, freeing up your IT team for other work.

TABLE OF CONTENTS

1	Introduction	3
2	Applying NetApp QoS in Your IT Environment	3
2.1	Storage QoS Benefits	4
2.2	Storage QoS Use Cases	4
	Create Storage Performance Limits for Enterprise Workloads	4
	Limit Maximum Storage Performance for Cloud Tenants	4
	Throttle the Performance of Misbehaving Applications	5
3	How QoS Works	5
3.1	QoS Policy Groups	6
4	Conclusion	6
5	Learn More	6

LIST OF FIGURES

Figure 1)	Storage QoS lets you apply performance capacity limits to files, LUNs, volumes, or SVMs.	3
Figure 2)	In private or public clouds, each tenant is assigned an SVM and a performance limit.	5
Figure 3)	A QoS policy group contains a collection of storage objects such as SVMs, volumes, LUNs, or files.	6

1 Introduction

In today's highly virtualized data centers and cloud environments, it has become common to have multiple applications, workgroups, or tenants share the same physical server, network, and storage resources.

Although shared infrastructure increases utilization and lowers costs, it creates new challenges when it comes to making sure that one workload or tenant doesn't affect another. Whether you're running mixed workloads in an enterprise data center, you've made the transition to a private cloud delivering IT as a service (ITaaS), or you're a cloud service provider, you're probably aware of these challenges.

Shared infrastructures in enterprise environments have to prevent application workloads from affecting each other, in order to deliver consistent performance for critical applications while at the same time making sure that important data protection processes and other maintenance tasks also complete.

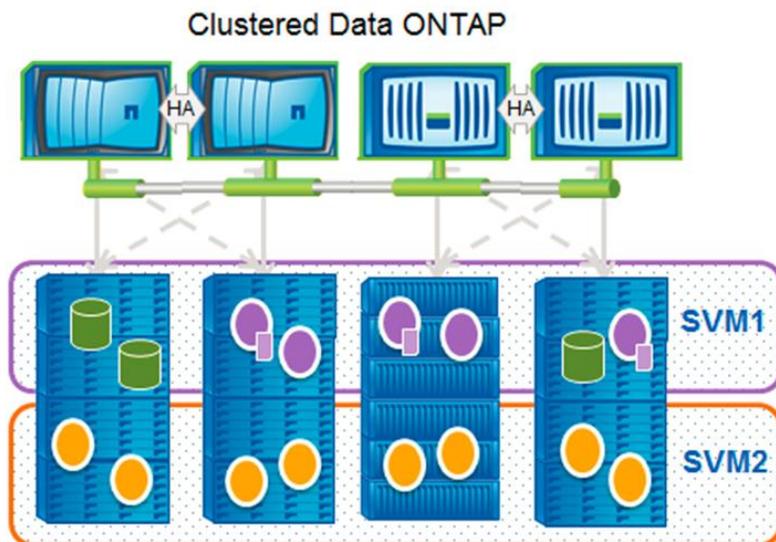
Private and public clouds need to prevent tenant activities from affecting each other, while meeting performance expectations and avoiding degradation as the number of tenants grows. In these environments it's important to make sure that tenants receive no more and no less than what they've paid for.

NetApp developed storage quality of service (QoS) to address concerns for busy storage environments of all types. Our vision for QoS includes delivering consistent performance by preventing workloads from affecting each other and by providing the ability to isolate and throttle resource-intensive workloads. NetApp aims to simplify workload management by providing operational visibility and the ability to monitor and analyze application workloads. NetApp is the only storage vendor today that delivers storage QoS across a whole range of products from entry to enterprise and across both SAN and NAS.

2 Applying NetApp QoS in Your IT Environment

A NetApp® storage cluster running clustered Data ONTAP can be subdivided into distinct storage virtual machines (SVMs), each governed by its own rights and permissions. SVMs make multi-tenancy possible by securely isolating individual tenants—for instance, in a service provider environment—or individual applications, workgroups, business units, and so on. Each application or tenant typically has its own SVM, and that SVM can be managed directly by the application owner or tenant using the SVM.

Figure 1) Storage QoS lets you apply performance capacity limits to individual files, LUNs, volumes, or an entire SVM.



Any time you put numerous workloads on a storage system or storage cluster, there is the possibility that excessive activity from one workload can affect other workloads. Storage QoS workload management (available in clustered Data ONTAP 8.2 and later) allows you to define service levels by creating policies that control the resources that can be consumed by individual files, volumes, or LUNs—or entire SVMs—to manage performance spikes and improve customer satisfaction. QoS gives you the ability to set not-to-exceed performance capacity limits (defined in terms of a maximum value for IOPS or MB/sec) on a group of files or volumes or LUNs within an SVM or on the entire SVM. This capability will enable enterprise IT administrators and service providers to consolidate many workloads or tenants on a cluster without fear that the most important workloads will suffer or that activity in one tenant partition will affect another. As a result, you can push your storage infrastructure to higher levels of utilization, increasing your overall efficiency and decreasing capital costs.

2.1 Storage QoS Benefits

NetApp storage QoS works with both SAN and NAS storage, and it runs across the entire NetApp FAS product line from entry to enterprise. You can incorporate third-party storage into your NetApp storage environment using NetApp V-Series. With V-Series as a front end to SAN arrays from EMC, HP, HDS, and others, your existing storage gets the full benefit of QoS limits and other NetApp capabilities.

Storage QoS offers significant benefits for all types of storage environments. It lets you:

- Achieve greater levels of consolidation
- Set limits corresponding to service-level agreements (SLAs)
- Add additional workloads with less risk of interference
- Make sure your customers get what they pay for, but not more

2.2 Storage QoS Use Cases

There are many possible use cases for storage QoS. The following examples illustrate some of the ways that QoS can be used.

Create Storage Performance Limits for Enterprise Workloads

The traditional approach to providing QoS for workloads with different service-level needs has been to house each application in its own dedicated infrastructure silo. However, this leads to lower resource utilization and higher costs.

Storage QoS makes it possible to run multiple workloads on a more cost-effective, shared infrastructure by creating limits for each significant workload. This can be accomplished by assigning each application workload to its own SVM and assigning a performance capacity (IOPS or MB/sec) limit on that SVM or by assigning a performance capacity limit on the group of volumes or LUNs associated with an application workload.

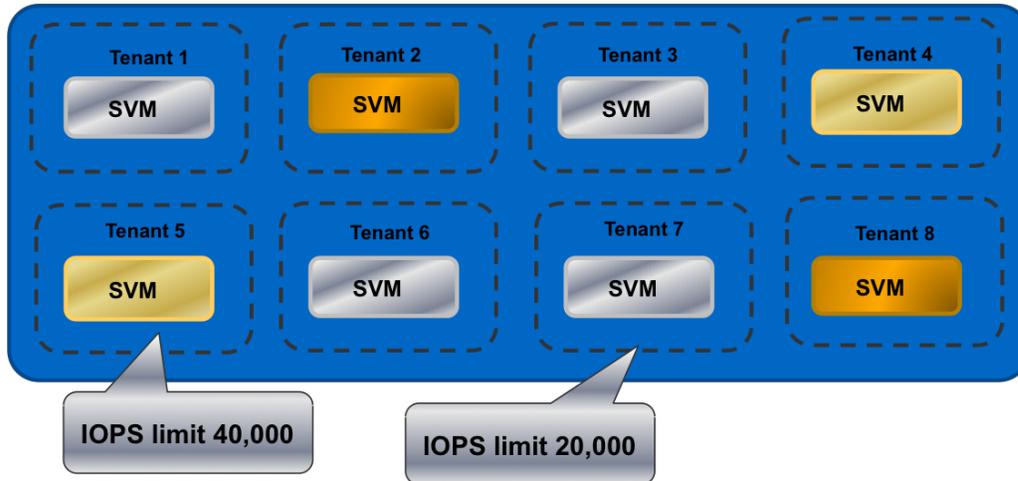
Limit Maximum Storage Performance for Cloud Tenants

In a multi-tenant cloud environment—whether private or public—the first tenants on a particular resource might see a level of performance in excess of their SLA. This can create a perception of performance degradation as additional tenants are added and performance decreases.

Storage QoS allows you to avoid this problem by assigning a performance capacity limit to each tenant in accordance with that tenant's SLA. That way, a tenant can't exceed the set performance limit, even when resources are available, and the tenant is therefore less likely to notice changes in perceived performance over time.

Another advantage of storage QoS is that it makes it simple to establish tiers of service based on SLAs. For instance, bronze service might correspond to a limit of 10,000 IOPS, silver to a limit of 20,000 IOPS, and gold to a limit of 40,000 IOPS.

Figure 2) In private or public clouds, each tenant is assigned an SVM and a performance limit. You can provide different service levels by establishing tiered limits (for example, bronze, silver, and gold).



Throttle the Performance of Misbehaving Applications

Another potential use case for storage QoS is to use it to throttle back an application that is using too much of the system performance capacity. In this scenario, you don't set QoS limits on workloads initially. Every workload can use as many resources as are available and all receive best effort.

However, when situations arise in which an application consumes too much of the system performance capacity—perhaps because of a programming error or unexpected spike in usage—storage QoS lets you react quickly.

Identifying the application, contacting the owner, and correcting the application can be a time-consuming process. A simpler and faster solution is to simply use storage QoS to throttle the performance capacity consumed by the I/Os associated with the application.

Data ONTAP tools allow you to quickly identify the overactive storage objects and isolate the workload by assigning a performance capacity limit low enough to eliminate resource contention.

3 How QoS Works

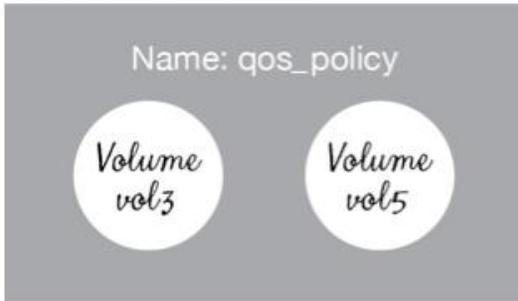
Storage QoS workload management allows you to control the resources that can be consumed by storage objects (such as volumes, LUNs, VMDKs, or SVMs) to manage performance spikes and improve customer satisfaction. You set throughput limits expressed in terms of MB/sec (for sequential workloads) or I/O operations per second (for transactional workloads) to achieve fine-grained control. When a limit is set on an SVM, the limit is shared for all objects within that SVM. This allows you to set a performance limit on a particular tenant or application, but it leaves the tenant or application free to manage the assigned resources however it chooses within that limit.

3.1 QoS Policy Groups

Storage QoS is applied by creating policy groups and applying limits to each policy group. For instance, a policy group can contain a single SVM or a collection of volumes or LUNs (within an SVM) used by an application.

Figure 3) A QoS policy group contains a collection of storage objects such as SVMs, volumes, LUNs, or files. The limit on a policy group applies to all the objects in the group collectively, even when the objects are on different nodes in the cluster.

QoS Policy Group



Limit: IOPS < 2000 per policy group

In virtual environments, a policy group could contain one or more VMDK files or LUNs within a datastore. The limit applied to a policy group is a combined limit for all the objects the policy group contains. The scheduler actively controls work so that resources are apportioned fairly to all objects in a group.

Note that the objects need not be on the same cluster node, and if an object moves within the cluster, the policy limit remains in effect. You can set a limit on a policy group in terms of IOPS or MB/sec, but not both.

4 Conclusion

Clustered Data ONTAP delivers the features and capabilities that enable your storage environment to succeed. Whether you are managing a traditional enterprise IT environment, a private cloud, or a public cloud, storage QoS gives you another important tool for optimizing storage.

With QoS you can increase utilization of storage resources by consolidating more workloads while minimizing the risk of workloads affecting each other. You can prevent tenants and applications from consuming all available resources, improving the end-user experience. Predefining service levels allows you to offer different levels of service to different consumers and applications.

With storage QoS you are able to use your storage infrastructure at maximum efficiency, so you need less total storage, and your capital expenses (capex) go down. Your IT team becomes more proactive and less reactive, helping to decrease operating expenses (opex) as well.

5 Learn More

To find out more about storage QoS, check out the following resources:

- ESG Lab Spotlight: NetApp Clustered Data ONTAP 8.2 Storage QoS <http://www.netapp.com/us/media/esg-lab-spotlight-ontap-qos.pdf>
- Technical Report-3982: NetApp Clustered Data ONTAP 8.2 <http://www.netapp.com/us/system/pdf-reader.aspx?pdfuri=tcm:10-60249-16&m=tr-3982.pdf>

Refer to the [Interoperability Matrix Tool \(IMT\)](#) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®



www.netapp.com

© 2013 NetApp, Inc. All rights reserved. No portions of this document may be reproduced without prior written consent of NetApp, Inc. Specifications are subject to change without notice. NetApp, the NetApp logo, Go further, faster, and Data ONTAP are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. WP-7191-0813