Performance and Quality of Service

HCI

Ann-Marie Grissino
April 10, 2020

This PDF was generated from https://docs.netapp.com/us-en/hci/docs/concept_hci_performance.html on October 28, 2020. Always check docs.netapp.com for the latest.

NetApp®

Table of Contents

Performance and Quality of Service 1
- Quality of Service parameters 1
- QoS value limits 2
- QoS performance 2
- QoS policies 3

Performance and Quality of Service

A SolidFire storage cluster has the ability to provide Quality of Service (QoS) parameters on a per-volume basis. You can guarantee cluster performance measured in inputs and outputs per second (IOPS) using three configurable parameters that define QoS: Min IOPS, Max IOPS, and Burst IOPS.

SolidFire Active IQ has a QoS recommendations page that provides advice on optimal configuration and set up of QoS settings.

Quality of Service parameters

IOPS parameters are defined in the following ways:

- **Minimum IOPS** - The minimum number of sustained inputs and outputs per second (IOPS) that the storage cluster provides to a volume. The Min IOPS configured for a volume is the guaranteed level of performance for a volume. Performance does not drop below this level.
- **Maximum IOPS** - The maximum number of sustained IOPS that the storage cluster provides to a volume. When cluster IOPS levels are critically high, this level of IOPS performance is not exceeded.
- **Burst IOPS** - The maximum number of IOPS allowed in a short burst scenario. If a volume has been running below the Max IOPS, burst credits are accumulated. When performance levels become very high and are pushed to maximum levels, short bursts of IOPS are allowed on the volume.

  Element software uses Burst IOPS when a cluster is running in a state of low cluster IOPS utilization.

  A single volume can accrue Burst IOPS and use the credits to burst above their Max IOPS up to their Burst IOPS level for a set "burst period." A volume can burst for up to 60 seconds if the cluster has the capacity to accommodate the burst. A volume accrues one second of burst credit (up to a maximum of 60 seconds) for every second that the volume runs below its Max IOPS limit.

  Burst IOPS are limited in two ways:

  - A volume can burst above its Max IOPS for a number of seconds equal to the number of burst credits that the volume has accrued.
  - When a volume bursts above its Max IOPS setting, it is limited by its Burst IOPS setting. Therefore, the burst IOPS never exceeds the burst IOPS setting for the volume.
- **Effective Max Bandwidth** - The maximum bandwidth is calculated by multiplying the number of IOPS (based on the QoS curve) by the IO size.

  Example: QoS parameter settings of 100 Min IOPS, 1000 Max IOPS, and 1500 Burst IOPs have the

following effects on quality of performance:

- Workloads are able to reach and sustain a maximum of 1000 IOPS until the condition of workload contention for IOPS becomes apparent on the cluster. IOPS are then reduced incrementally until IOPS on all volumes are within the designated QoS ranges and contention for performance is relieved.
- Performance on all volumes is pushed toward the Min IOPS of 100. Levels do not drop below the Min IOPS setting but could remain higher than 100 IOPS when workload contention is relieved.
- Performance is never greater than 1000 IOPS, or less than 100 IOPS for a sustained period. Performance of 1500 IOPS (Burst IOPS) is allowed, but only for those volumes that have accrued burst credits by running below Max IOPS and only allowed for a short periods of time. Burst levels are never sustained.

QoS value limits

Here are the possible minimum and maximum values for QoS.

Parameters	Min value	Default	4 4KB	5 8KB	6 16KB	262KB
Min IOPS	50	50	15,000	9,375*	5556*	385*
Max IOPS	100	15,000	200,000**	125,000	74,074	5128
Burst IOPS	100	15,000	200,000**	125,000	74.074	5128

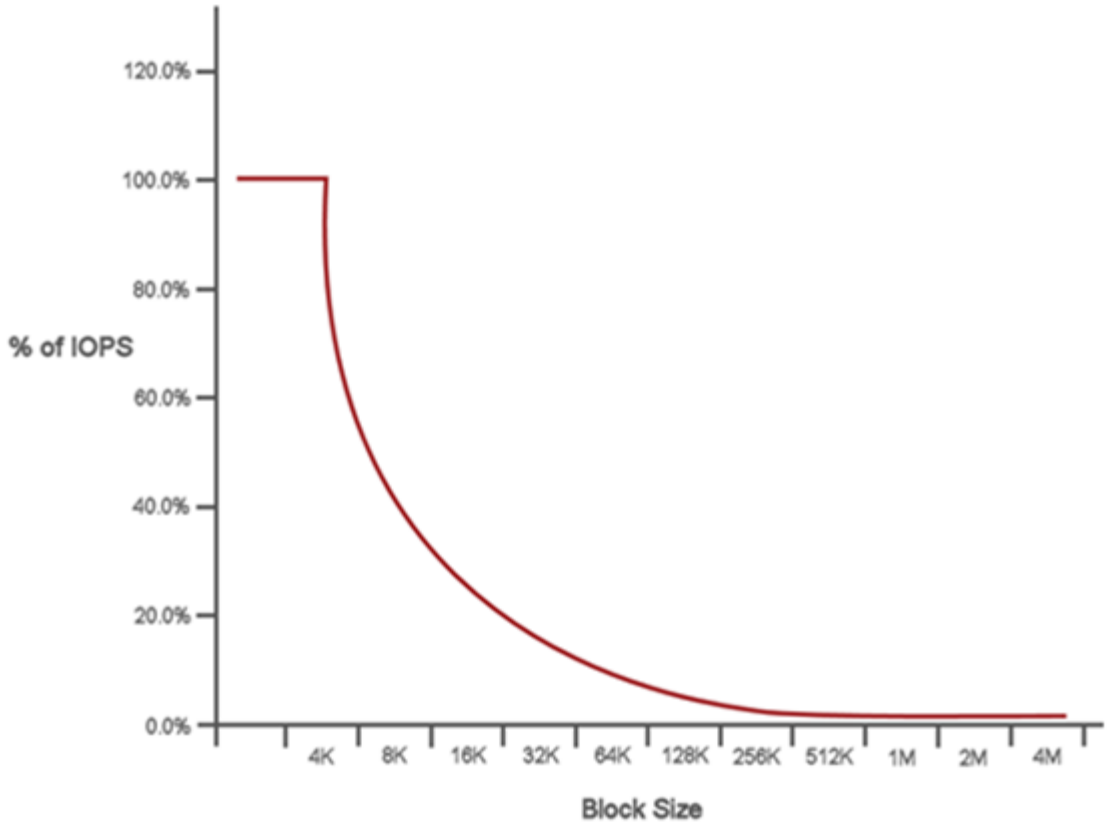
*These estimations are approximate.
**Max IOPS and Burst IOPS can be set as high as 200,000; however, this setting is allowed only to effectively uncap the performance of a volume. Real-world maximum performance of a volume is limited by cluster usage and per-node performance.

QoS performance

The QoS performance curve shows the relationship between block size and the percentage of IOPS.

Block size and bandwidth have a direct impact on the number of IOPS that an application can obtain. Element software takes into account the block sizes it receives by normalizing block sizes to 4k. Based on workload, the system might increase block sizes. As block sizes increase, the system increases bandwidth to a level necessary to process the larger block sizes. As bandwidth increases the number of IOPS the system is able to attain decreases.

The QoS performance curve shows the relationship between increasing block sizes and the decreasing percentage of IOPS:



As an example, if block sizes are 4k, and bandwidth is 4000 KBps, the IOPS are 1000. If block sizes increase to 8k, bandwidth increases to 5000 KBps, and IOPS decrease to 625. By taking block size into account, the system ensures that lower priority workloads that use higher block sizes, such as backups and hypervisor activities, do not take too much of the performance needed by higher priority traffic using smaller block sizes.

QoS policies

A QoS policy enables you to create and save a standardized quality of service setting that can be applied to many volumes. You can create, edit, and delete QoS policies.

QoS policies are best for service environments, for example, with database, application, or infrastructure servers that rarely reboot and need constant equal access to storage. QoSSIOC automation is best for light use VMs, such as virtual desktops or specialized kiosk-type VMs, that may be rebooted, powered on, or powered off daily or several times a day. QoSSIOC automation and QoS policies should not be used together. If you are using QoS policies, do not enable QoSSIOC. QoSSIOC will override and adjust QoS values for volume QoS settings.

You can view QoS policies on the **Management** > **QoS Policies** page from the NetApp Element Management extension point.

The selected cluster must be Element 10.0 or later to use QoS policies; otherwise, QoS policy functions are not available.

Find more information

- NetApp HCI Documentation Center
- NetApp HCI Resources page

Copyright Information

Copyright © 2020 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.