



使用 **QoS** 策略组信息管理性能 Active IQ Unified Manager

NetApp
January 15, 2026

This PDF was generated from https://docs.netapp.com/zh-cn/active-iq-unified-manager/performance-checker/concept_how_storage_qos_can_control_workload_throughput.html on January 15, 2026. Always check docs.netapp.com for the latest.

目录

- 使用 QoS 策略组信息管理性能 1
 - 存储 QoS 如何控制工作负载吞吐量 1
 - 查看所有集群上可用的所有 QoS 策略组 2
 - 查看同一 QoS 策略组中的卷或 LUN 2
 - 查看应用于特定卷或 LUN 的 QoS 策略组设置 3
 - 查看性能图表以比较同一 QoS 策略组中的卷或 LUN 3
 - 不同类型的 QoS 策略如何显示在吞吐量图表中 4
 - 在性能资源管理器中查看工作负载 QoS 最小和最大设置 5

使用 QoS 策略组信息管理性能

通过 Unified Manager，您可以查看所监控的所有集群上可用的服务质量（QoS）策略组。这些策略可能是使用 ONTAP 软件（System Manager 或 ONTAP 命令行界面）或 Unified Manager 性能服务级别策略定义的。Unified Manager 还会显示已分配 QoS 策略组的卷和 LUN。

有关调整 QoS 设置的详细信息，请参见 ["性能管理概述"](#)

存储 QoS 如何控制工作负载吞吐量

您可以创建服务质量（QoS）策略组来控制其所含工作负载的每秒 I/O 数（IOPS）或吞吐量（MB/秒）限制。如果工作负载位于没有设置限制的策略组中，例如默认策略组，或者设置的限制不符合您的需求，您可以增加此限制或将工作负载移动到具有所需限制的新策略组或现有策略组。

可以将“传统”QoS 策略组分配给单个工作负载，例如单个卷或 LUN。在这种情况下，工作负载可以使用完整吞吐量限制。此外，还可以将 QoS 策略组分配给多个工作负载；在这种情况下，工作负载之间的吞吐量限制为“共享”。例如，如果为三个工作负载分配的 QoS 限制为 9,000 IOPS，则合并 IOPS 将限制为超过 9,000 IOPS。

也可以将“自适应”QoS 策略组分配给单个或多个工作负载。但是，即使分配给多个工作负载，每个工作负载也会获得完整的吞吐量限制，而不是与其他工作负载共享吞吐量值。此外，自适应 QoS 策略会根据卷大小为每个工作负载自动调整吞吐量设置，从而在卷大小发生变化时保持 IOPS 与 TB 数的比率。例如，如果在自适应 QoS 策略中将峰值设置为 5,000 IOPS/TB，则一个 10 TB 卷的最大吞吐量将为 50,000 IOPS。如果稍后将卷的大小调整为 20 TB，则自适应 QoS 会将最大 IOPS 调整为 100,000 IOPS。

从 ONTAP 9.5 开始，您可以在定义自适应 QoS 策略时包括块大小。如果工作负载使用的块大小非常大，最终使用的吞吐量百分比非常大，则此功能可以有效地将此策略从 IOPS/TB 阈值转换为 MB/秒阈值。

对于共享组 QoS 策略，如果策略组中所有工作负载的 IOPS 或 MB/秒超过设置的限制，则策略组会限制工作负载以限制其活动，从而降低策略组中所有工作负载的性能。如果策略组限制生成动态性能事件，则事件问题描述将显示相关策略组的名称。

在“性能：所有卷”视图中，您可以按 IOPS 和 MB/秒对受影响的卷进行排序，以查看哪些工作负载的使用率可能最高，而导致事件发生。在“性能 / 卷资源管理器”页面中，您可以选择卷上的其他卷或 LUN，以便与受影响的工作负载 IOPS 或 MBps 吞吐量使用量进行比较。

通过将过度使用节点资源的工作负载分配给限制性更强的策略组设置，策略组会限制工作负载以限制其活动，从而减少该节点上资源的使用。但是，如果您希望工作负载能够使用更多节点资源，则可以增加策略组的值。

您可以使用 System Manager，ONTAP 命令或 Unified Manager 性能服务级别来管理策略组，包括以下任务：

- 创建策略组
- 在策略组中添加或删除工作负载
- 在策略组之间移动工作负载
- 更改策略组的吞吐量限制

- 将工作负载移动到其他聚合和 / 或节点

查看所有集群上可用的所有 QoS 策略组

您可以显示 Unified Manager 正在监控的集群上所有可用的 QoS 策略组的列表。其中包括传统 QoS 策略，自适应 QoS 策略以及由 Unified Manager 性能服务级别策略管理的 QoS 策略。

步骤

1. 在左侧导航窗格中，单击 * 存储 * > * QoS 策略组 *。

默认情况下，将显示 " 性能：传统 QoS 策略组 " 视图。

2. 查看每个可用传统 QoS 策略组的详细配置设置。
3. 单击 Expand 按钮（▼）以查看有关该策略组的更多详细信息。
4. 在 " 视图 " 菜单中，选择其他选项之一以查看所有自适应 QoS 策略组或使用 Unified Manager 性能服务级别创建的所有 QoS 策略组。

查看同一 QoS 策略组中的卷或 LUN

您可以显示已分配给同一 QoS 策略组的卷和 LUN 的列表。

如果传统 QoS 策略组在多个卷中为 s 共享，则查看某些卷是否过度使用为策略组定义的吞吐量可能会很有用。它还有助于您确定是否可以在不对其他卷产生负面影响的情况下将其他卷添加到策略组。

对于自适应 QoS 策略和 Unified Manager 性能服务级别策略，此操作有助于查看正在使用策略组的所有卷或 LUN，以便您可以查看在更改 QoS 策略的配置设置后哪些对象会受到影响。

步骤

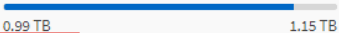
1. 在左侧导航窗格中，单击 * 存储 * > * QoS 策略组 *。

默认情况下，将显示 " 性能：传统 QoS 策略组 " 视图。

2. 如果您对传统策略组感兴趣，请继续查看此页面。否则，请选择其他视图选项之一以显示 Unified Manager 性能服务级别创建的所有自适应 QoS 策略组或所有 QoS 策略组。
3. 在您感兴趣的 QoS 策略中，单击展开按钮（▼）以查看更多详细信息。

View Adaptive QoS Policy Groups Search Quality of Service

Schedule Report

QoS Policy Group	Cluster	SVM	Min Through...	Max Through...	Absolute Min...	Block Size	Asso
▼ julia_vs2_cifs_Performance	opm-simplicity	julia_vs2_cifs	2048.0 IOPS/TB	4096.0 IOPS/TB	500IOPS		1
▲ julia_vs1_nfs_Performance	opm-simplicity	julia_vs1_nfs	2048.0 IOPS/TB	4096.0 IOPS/TB	500IOPS		2
Details Allocated Capacity  0.99 TB 1.15 TB Associated Objects 2 Volumes 0 LUNs Events None							
▼ julia_nfs_extreme_Extreme_Performance	ocum-mobility-01-02	julia_nfs_extreme	6144.0 IOPS/TB	12288.0 IOPS/TB	1000IOPS	any	1
▼ julia_extreme_jan16_aqos	ocum-mobility-01-02	julia_nfs_extreme	10000.0 IOPS/TB	12000.0 IOPS/TB	1000IOPS	any	1

- 单击卷或 LUN 链接以查看使用此 QoS 策略的对象。

此时将显示卷或 LUN 的性能清单页面，其中列出了正在使用 QoS 策略的对象。

查看应用于特定卷或 LUN 的 QoS 策略组设置

您可以查看已应用于卷和 LUN 的 QoS 策略组，也可以链接到性能 /QoS 策略组视图以显示每个 QoS 策略的详细配置设置。

下面显示了查看应用于卷的 QoS 策略的步骤。查看 LUN 的此信息的步骤与此类似。

步骤

- 在左侧导航窗格中，单击 * 存储 * > * 卷 *。

默认情况下，将显示运行状况：所有卷视图。

- 在视图菜单中，选择 * 性能： QoS 策略组中的卷 *。
- 找到要查看的卷并向右滚动，直到看到 * QoS 策略组 * 列。
- 单击 QoS 策略组名称。

根据是传统 QoS 策略，自适应 QoS 策略还是使用 Unified Manager 性能服务级别创建的 QoS 策略，系统将显示相应的服务质量页面。

- 查看 QoS 策略组的详细配置设置。
- 单击 Expand 按钮 (▼) 以查看有关该策略组的更多详细信息。

查看性能图表以比较同一 QoS 策略组中的卷或 LUN

您可以查看同一 QoS 策略组中的卷和 LUN，然后在一个 IOPS，MB/ 秒或 IOPS/TB 图表上比较性能，以确定是否存在任何问题。

下面显示了比较同一 QoS 策略组中卷的性能的步骤。查看 LUN 的此信息的步骤与此类似。

步骤

1. 在左侧导航窗格中，单击 * 存储 * > * 卷 *。

默认情况下，将显示运行状况：所有卷视图。

2. 在视图菜单中，选择 * 性能： QoS 策略组中的卷 *。
3. 单击要查看的卷的名称。

此时将显示卷的性能资源管理器页面。

4. 在查看和比较菜单中，选择 * 同一 QoS 策略组中的卷 *。

下表列出了共享相同 QoS 策略的其他卷。

5. 单击 * 添加 * 按钮将这些卷添加到图表中，以便比较图表中所有选定卷的 IOPS ， MB/ 秒， IOPS/TB 和其他性能计数器。

您可以更改时间范围以查看默认 72 小时以外的其他时间间隔的性能。

不同类型的 QoS 策略如何显示在吞吐量图表中

您可以在性能资源管理器和工作负载分析 IOPS ， IOPS/TB 和 MB/ 秒图表中查看已应用于卷或 LUN 的 ONTAP 定义的服务质量（ QoS ）策略设置。图表中显示的信息因应用于工作负载的 QoS 策略类型而异。

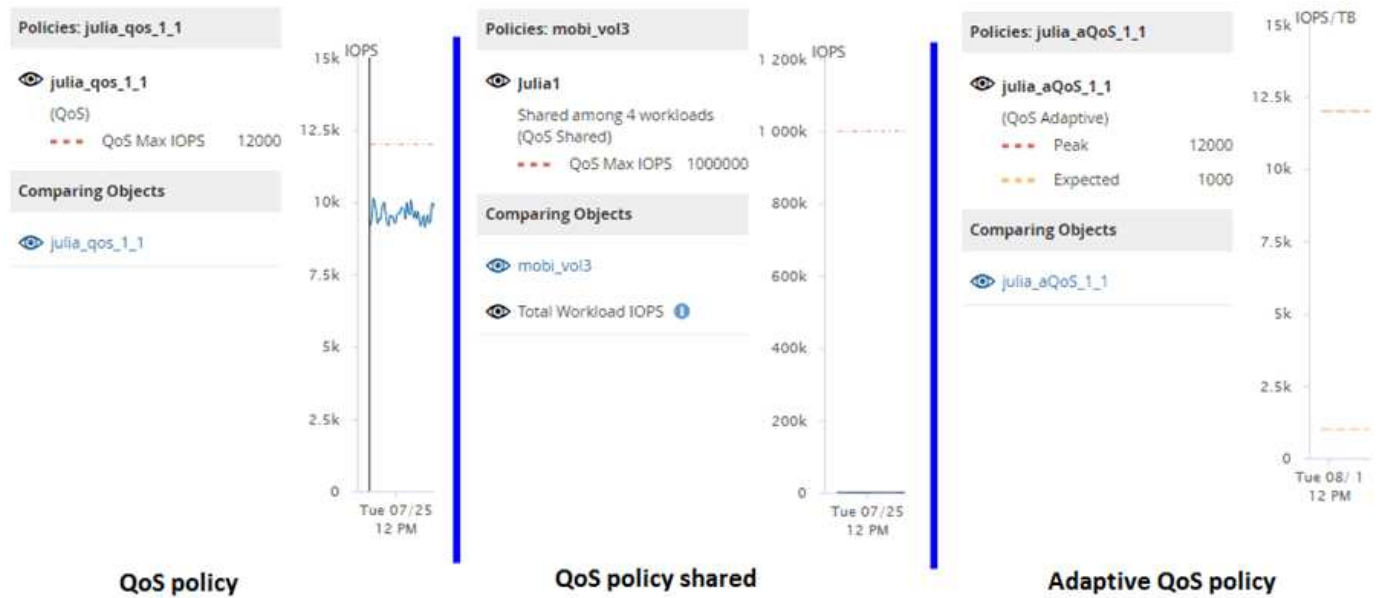
最大吞吐量（或 "峰值"）设置用于定义工作负载可以使用的最大吞吐量，从而限制对争用系统资源的工作负载的影响。最小吞吐量（或 "预期"）设置定义了工作负载必须使用的最小吞吐量，以便无论争用工作负载的需求如何，关键工作负载都能满足最小吞吐量目标。

IOPS 和 MB/ 秒的共享和非共享 QoS 策略使用术语 minimum 和 maximum 来定义下限和上限。ONTAP 9.3 中引入的 IOPS/TB 自适应 QoS 策略使用术语 "expected" 和 "peak" 来定义下限和上限。

虽然 ONTAP 允许您创建这两种类型的 QoS 策略，但根据这些策略应用于工作负载的方式， QoS 策略将通过三种方式显示在性能图表中。

策略类型	功能	Unified Manager 界面中的指示符
分配给单个工作负载的 QoS 共享策略，或分配给单个或多个工作负载的 QoS 非共享策略	每个工作负载都可以使用指定的吞吐量设置	显示 "（ QoS ） `"
分配给多个工作负载的 QoS 共享策略	所有工作负载都共享指定的吞吐量设置	显示 "（ QoS 共享） `"
分配给单个或多个工作负载的自适应 QoS 策略	每个工作负载都可以使用指定的吞吐量设置	显示 "（ QoS 自适应） `"

下图举例说明了这三个选项在计数器图表中的显示方式。



如果已在 IOPS 中定义的正常 QoS 策略显示在工作负载的 IOPS/TB 图表中，则 ONTAP 会将 IOPS 值转换为 IOPS/TB 值，Unified Manager 会在 IOPS/TB 图表中显示该策略以及文本 "QoS, defined in IOPS"。

当以 IOPS/TB 为单位定义的自适应 QoS 策略显示在工作负载的 IOPS 图表中时，ONTAP 会将 IOPS/TB 值转换为 IOPS 值，Unified Manager 会在 IOPS 图表中显示该策略以及文本 "QoS Adaptive - Used, 以 IOPS/TB 为单位定义或 "QoS Adaptive - Allocated, defined in IOPS/TB"，具体取决于峰值 IOPS 分配设置的配置方式。如果分配设置设置为 "allocated-space"，则会根据卷大小计算峰值 IOPS。如果分配设置设置为 "used-space"，则会根据卷中存储的数据量并考虑存储效率来计算峰值 IOPS。



只有当卷使用的逻辑容量大于或等于 128 GB 时，IOPS/TB 图表才会显示性能数据。如果已用容量在选定时间范围内降至 128 GB 以下，则图表中会显示空隙。

在性能资源管理器中查看工作负载 QoS 最小和最大设置

您可以在性能资源管理器图表中查看卷或 LUN 上 ONTAP 定义的服务质量（QoS）策略设置。最大吞吐量设置限制了争用工作负载对系统资源的影响。最小吞吐量设置可确保关键工作负载满足最小吞吐量目标，而不管竞争工作负载的需求如何。

只有在 ONTAP 中配置了 QoS 吞吐量 "minimum" 和 "maximum" IOPS 和 MB/秒设置后，这些设置才会显示在计数器图表中。只有运行 ONTAP 9.2 或更高版本软件的系统以及 AFF 系统才提供最小吞吐量设置，并且目前只能针对 IOPS 进行设置。

自适应 QoS 策略从 ONTAP 9.3 开始可用，并使用 IOPS/TB 而非 IOPS 来表示。这些策略会根据每个工作负载的卷大小自动调整 QoS 策略值，从而在卷大小发生变化时保持 IOPS 与 TB 数的比率。您只能将自适应 QoS 策略组应用于卷。QoS 术语 "expected" 和 "peak" 用于自适应 QoS 策略，而不是最小值和最大值。

如果工作负载吞吐量在前一小时的每个性能收集期间均超过定义的 QoS 最大策略设置，则 Unified Manager 将针对 QoS 策略违规生成警告事件。在每个收集期间，工作负载吞吐量可能会短时间超过 QoS 阈值，但 Unified Manager 会在图表上显示收集期间的 "Average" 吞吐量。因此，您可能会看到 QoS 事件，而工作负载的吞吐量可能未超过图表中显示的策略阈值。

步骤

1. 在选定卷或 LUN 的 * 性能资源管理器 * 页面中，执行以下操作以查看 QoS 上限和下限设置：

如果您要 ...	执行此操作 ...
查看 IOPS 上限（ QoS 最大值）	在 IOPS 总计或细分图表中，单击 * 缩放视图 * 。
查看 MB/ 秒上限（ QoS 最大值）	在 MB/ 秒总计或细分图表中，单击 * 缩放视图 * 。
查看 IOPS 下限（ QoS 最小值）	在 IOPS 总计或细分图表中，单击 * 缩放视图 * 。
查看 IOPS/TB 上限（ QoS 峰值）	对于卷，在 IOPS/TB 图表中，单击 * 缩放视图 * 。
查看 IOPS/TB 下限（ QoS 预期值）	对于卷，在 IOPS/TB 图表中，单击 * 缩放视图 * 。

- 虚线水平线表示在 ONTAP 中设置的最大或最小吞吐量值。您还可以查看何时对 QoS 值进行了更改。
2. 要查看与 QoS 设置比较的特定 IOPS 和 MB/ 秒值，请将光标移动到图表区域以查看弹出窗口。
- 如果您发现某些卷或 LUN 的 IOPS 或 MB/ 秒非常高，并对系统资源造成压力，则可以使用 System Manager 或 ONTAP 命令行界面调整 QoS 设置，以便这些工作负载不会影响其他工作负载的性能。
- 有关调整QoS设置的详细信息、请参见 ["性能管理概述"](#)

版权信息

版权所有 © 2026 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。