通过 QoS 保证吞吐量

ONTAP 9

NetApp
February 12, 2026

This PDF was generated from https://docs.netapp.com/zh-cn/ontap/performance-admin/guarantee-throughput-qos-task.html on February 12, 2026. Always check docs.netapp.com for the latest.

目录

通过 QoS 保证吞吐量 1
通过ONTAP中的QoS概述保证吞吐量 1
吞吐量上限(QoS最大值) 1
吞吐量下限(QoS最小值) 1
共享和非共享QoS策略组 2
自适应 QoS 2
常规支持 3
支持的工作负载达到吞吐量上限 4
支持吞吐量下限的工作负载 4
自适应 QoS 支持的工作负载 5
工作负载和策略组的最大数量 5
启用或禁用ONTAP吞吐量底线 v2 5
ONTAP存储 QoS 工作流程 6
在ONTAP中使用QoS设置吞吐量上限 7
在ONTAP中使用QoS设置吞吐量下限 10
在ONTAP中使用自适应QoS策略组 12
默认自适应 QoS 策略组 13
存储对象策略组分配限制 13
在ONTAP中设置自适应策略组模板 15

通过 QoS 保证吞吐量

通过ONTAP中的QoS概述保证吞吐量

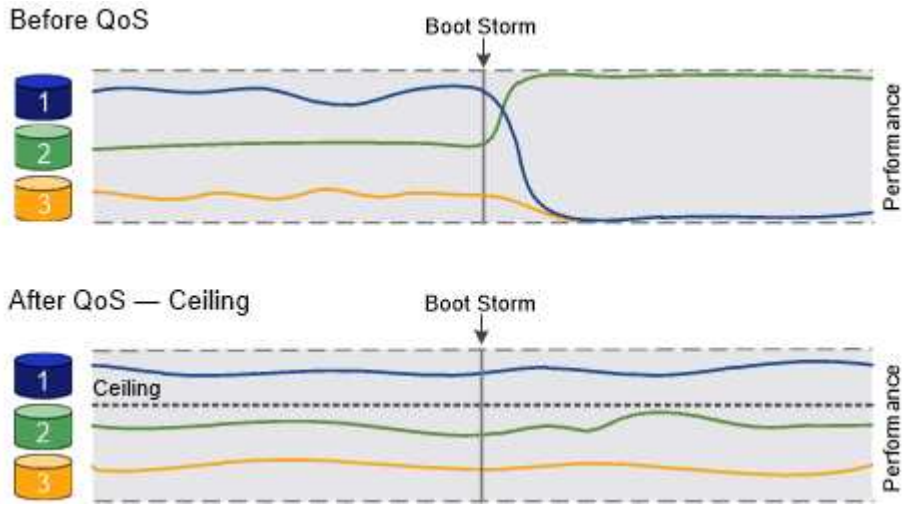
您可以使用存储服务质量（ QoS ）来保证关键工作负载的性能不会因争用资源的工作负载而降级。您可以为争用资源的工作负载设置吞吐量上限，以限制其对系统资源的影响，也可以为关键工作负载设置吞吐量上限，以确保满足最小吞吐量目标，而不管争用资源的工作负载有何需求。您甚至可以为同一工作负载设置上限和下限。

吞吐量上限(QoS最大值)

吞吐量上限会将工作负载的吞吐量限制为最大 IOPS 或 MBps 数，或者 IOPS 和 MBps 数。在下图中，工作负载 2 的吞吐量上限可确保它不会 " 抢占 " 工作负载 1 和 3 。

策略组定义了一个或多个工作负载的吞吐量上限。工作负载表示 _storage 对象： _ 卷，文件， qtree 或 LUN ，或者 SVM 中的所有卷，文件， qtree 或 LUN 的 I/O 操作。您可以在创建策略组时指定上限，也可以等到监控工作负载之后再指定上限。

> 工作负载的吞吐量可能会超出指定上限 10% ，尤其是在工作负载的吞吐量发生快速变化时。要处理突发事件，上限可能会超过 50% 。令牌累积率高达 150% 时，单个节点上发生突发



吞吐量下限(QoS最小值)

吞吐量下限可确保工作负载的吞吐量不会低于最小IOPS或MBps数或IOPS和MBps数。在下图中，工作负载 1 和工作负载 3 的吞吐量下限可确保满足最小吞吐量目标，而不管工作负载 2 的需求如何。

> 如示例所示，吞吐量上限会直接限制吞吐量。吞吐量下限会优先考虑设置了下限的工作负载，从而间接限制吞吐量。

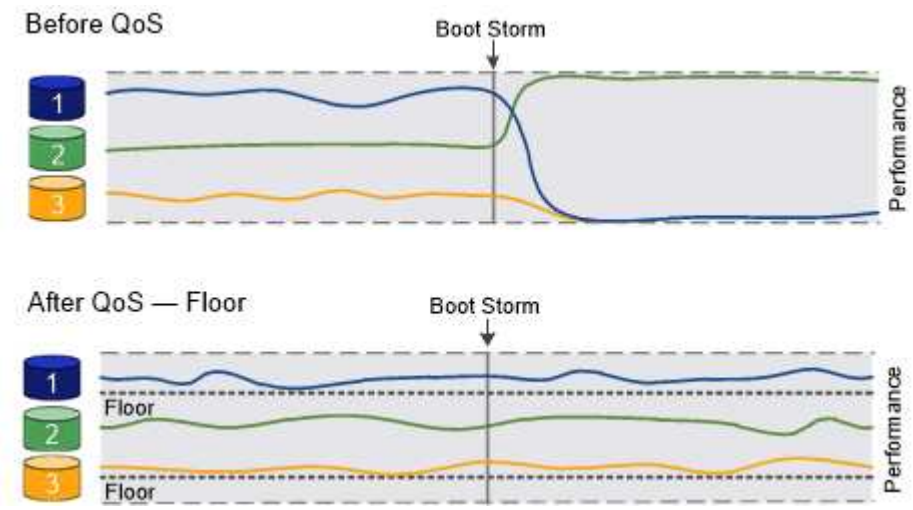
您可以在创建策略组时指定下限，也可以等到监控工作负载之后再指定下限。

从ONTAP 9.13.1开始，您可以使用在SVM范围设置吞吐量下限[adaptive-qos-templates]。在9.13.1之前的ONTAP 版本中、定义吞吐量下限的策略组不能应用于SVM。

在 ONTAP 9.7 之前的版本中，如果有足够的可用性能容量，则可以保证吞吐量下限。

在 ONTAP 9.7 及更高版本中，即使可用性能容量不足，也可以保证吞吐量下限。这种新的楼层行为称为 Floor v2 。为了满足保证要求，对于没有吞吐量下限的工作负载或超出下限设置的工作负载， Floor v2 可能会导致延迟更高。第 2 层适用场景 QoS 和自适应 QoS 。

ONTAP 9.7P6及更高版本提供了启用/禁用楼层v2新行为的选项。在执行关键操作(如)期间、工作负载可能会低于指定的下限 `volume move trigger-cutover`。即使有足够的可用容量且未执行关键操作，工作负载的吞吐量也可能会低于指定下限 5% 。如果楼层配置过度，并且没有性能容量，则某些工作负载可能会低于指定的楼层。



共享和非共享QoS策略组

从 ONTAP 9.4 开始，您可以使用 *non-shared* QoS 策略组分别指定每个成员工作负载的已定义吞吐量上限或下限适用场景。*shared* 策略组的行为取决于策略类型：

- 对于吞吐量上限，分配给共享策略组的工作负载的总吞吐量不能超过指定的上限。
- 对于吞吐量下限，共享策略组只能应用于单个工作负载。

自适应 QoS

通常，分配给存储对象的策略组值是固定的。当存储对象的大小发生变化时，您需要手动更改此值。例如，增加卷上的已用空间量通常需要相应地增加为卷指定的吞吐量上限。

Adaptive QoS 会自动将策略组值扩展到工作负载大小，并在工作负载大小发生变化时保持 IOPS 与 TBSGB 的比率。如果您要在大型部署中管理数百或数千个工作负载，则这是一项显著优势。

通常，您可以使用自适应 QoS 来调整吞吐量上限，但也可以使用它来管理吞吐量下限（当工作负载大小增加时）。工作负载大小表示为存储对象分配的空间或存储对象使用的空间。

在 ONTAP 9.5 及更高版本中，已用空间可用于吞吐量下限。在 ONTAP 9.4 及更早版本中，吞吐量下限不支持此功能。

- 已分配空间策略会根据存储对象的标称大小保持 IOPS/TBGB 比率。如果此比率为 100 IOPS/GB ，则只要 150 GB 卷保持此大小，其吞吐量上限将为 15 ， 000 IOPS 。如果将卷大小调整为 300 GB ，则自适应

QoS 会将吞吐量上限调整为 30 ， 000 IOPS 。

- 已用空间策略（默认值）会根据存储效率之前存储的实际数据量保持 IOPS/TBGB 比率。如果此比率为 100 IOPS/GB ，则存储了 100 GB 数据的 150 GB 卷的吞吐量上限为 10 ， 000 IOPS 。随着已用空间量的变化，自适应 QoS 会根据比率调整吞吐量上限。

从 ONTAP 9.5 开始，您可以为应用程序指定 I/O 块大小，以便以 IOPS 和 MBps 为单位表示吞吐量限制。MBps 限制是通过块大小乘以 IOPS 限制计算得出的。例如，如果 I/O 块大小为 32 K ，而 IOPS 限制为 6144IOPS/TB ，则 MBps 限制为 192 MBps 。

吞吐量上限和下限均会出现以下行为：

- 将工作负载分配给自适应 QoS 策略组后，上限或下限将立即更新。
- 调整自适应 QoS 策略组中的工作负载大小后，上限或下限大约会在五分钟内更新。

在进行更新之前，吞吐量必须至少增加 10 IOPS 。

自适应 QoS 策略组始终为非共享组：定义的吞吐量上限或每个成员工作负载的下限适用场景。

从ONTAP 9.6开始、采用SSD的ONTAP Select 高级版支持吞吐量下限。

自适应策略组模板

从ONTAP 9.13.1开始、您可以在SVM上设置自适应QoS模板。通过自适应策略组模板、您可以为SVM中的所有卷设置吞吐量下限和上限。

只有在创建SVM之后、才能设置自适应策略组模板。使用 `vserver modify` 命令 `-qos-adaptive-policy-group-template` 参数以设置策略。

设置自适应策略组模板时、在设置策略后创建或迁移的卷会自动继承策略。分配策略模板时、SVM上现有的任何卷不受影响。如果在SVM上禁用此策略、则此后迁移到SVM或在此SVM上创建的任何卷都不会收到此策略。禁用自适应策略组模板不会影响继承策略模板的卷、因为它们会保留策略模板。

有关详细信息，请参见 设置自适应策略组模板。

常规支持

下表显示了在支持吞吐量上限，吞吐量下限和自适应 QoS 方面的差异。

资源或功能	吞吐量上限	吞吐量下限	吞吐量下限 **v2**	自适应 **QoS**
ONTAP 9 版本	全部	9.2及更高版本	9.7及更高版本	9.3及更高版本
平台	全部	• AFF • C190 [1] • ONTAP Select高级版、含SSD [1]	• AFF • C190 • 采用 SSD 的 ONTAP Select 高级版	全部
协议	全部	全部	全部	全部

资源或功能	吞吐量上限	吞吐量下限	吞吐量下限 **v2**	自适应 **QoS**
FabricPool	是的。	是，如果分层策略设置为 " 无 " 且云中没有块。	是，如果分层策略设置为 " 无 " 且云中没有块。	否
SnapMirror 同步	是的。	否	否	是的。

从ONTAP 9.6版开始支持[1] C190和ONTAP Select。

支持的工作负载达到吞吐量上限

下表按 ONTAP 9 版本显示了工作负载对吞吐量上限的支持。不支持根卷，负载共享镜像和数据保护镜像。

工作负载支持	**ONTAP 9.8**及更高版本	**ONTAP 9.7** 至 **9.4**	**ONTAP 9.3**及更早版本
Volume	是的。	是的。	是的。
文件	是的。	是的。	是的。
LUN	是的。	是的。	是的。
SVM	是的。	是的。	是的。
FlexGroup 卷	是的。	是的。	是(仅限ONTAP 9.3)
qtrees [1]	是的。	否	否
每个策略组具有多个工作负载	是的。	是的。	是的。
非共享策略组	是的。	是的。	否

[1] 从ONTAP 9.9.1 开始，启用 SMB 的FlexVol和FlexGroup卷中的 qtree 也支持 SMB 访问。从ONTAP 9.8 开始，启用 NFS 的FlexVol和FlexGroup卷中的 qtree 支持 NFS 访问。

支持吞吐量下限的工作负载

下表按 ONTAP 9 版本显示了吞吐量下限的工作负载支持。不支持根卷，负载共享镜像和数据保护镜像。

工作负载支持	**ONTAP 9.13.1**及更高版本	**ONTAP 9.8** 到 **9.13.0**	**ONTAP 9.4**到**9.7**	**ONTAP 9.3**
Volume	是的。	是的。	是的。	是的。
文件	是的。	是的。	是的。	是的。
LUN	是的。	是的。	是的。	是的。

工作负载支持	**ONTAP 9.13.1**及更高版本	**ONTAP 9.8** 到 **9.13.0**	**ONTAP 9.4**到**9.7**	**ONTAP 9.3**
SVM	是的。	否	否	否
FlexGroup 卷	是的。	是的。	是的。	否
qtrees [1]	是的。	是的。	否	否
每个策略组具有多个工作负载	是的。	是的。	是的。	否
非共享策略组	是的。	是的。	是的。	否

[1]从ONTAP 9.8开始、在启用了NFS的FlexVol和FlexGroup卷中的qtrees支持NFS访问。从 ONTAP 9.1.1 开始，启用了 SMB 的 FlexVol 和 FlexGroup 卷的 qtree 也支持 SMB 访问。

自适应 **QoS** 支持的工作负载

下表显示了 ONTAP 9 版本对自适应 QoS 的工作负载支持。不支持根卷，负载共享镜像和数据保护镜像。

工作负载支持	**ONTAP 9.13.1**及更高版本	**ONTAP 9.4** 到 **9.13.0**	**ONTAP 9.3**
Volume	是的。	是的。	是的。
文件	是的。	是的。	否
LUN	是的。	是的。	否
SVM	是的。	否	否
FlexGroup 卷	是的。	是的。	否
每个策略组具有多个工作负载	是的。	是的。	是的。
非共享策略组	是的。	是的。	是的。

工作负载和策略组的最大数量

下表按 ONTAP 9 版本显示了工作负载和策略组的最大数量。

工作负载支持	**ONTAP 9.4**及更高版本	**ONTAP 9.3**及更早版本
每个集群的最大工作负载数	40、000	12、000
每个节点的最大工作负载数	40、000	12、000
最大策略组数	12、000	12、000

启用或禁用**ONTAP**吞吐量底线 **v2**

您可以在 AFF 上启用或禁用吞吐量下限 v2 。默认值为 enabled 。如果启用了楼层 v2 ，

则在控制器大量使用时，以其他工作负载更高的延迟为代价，可以满足吞吐量下限。第 2 层适用场景 QoS 和自适应 QoS 。

步骤

1. 更改为高级权限级别：

```
set -privilege advanced
```

2. 输入以下命令之一：

如果您要 ...	使用以下命令：
禁用楼层 v2	`qos settings throughput-floors-v2 -enable false`
启用楼层 v2	`qos settings throughput-floors-v2 -enable true`

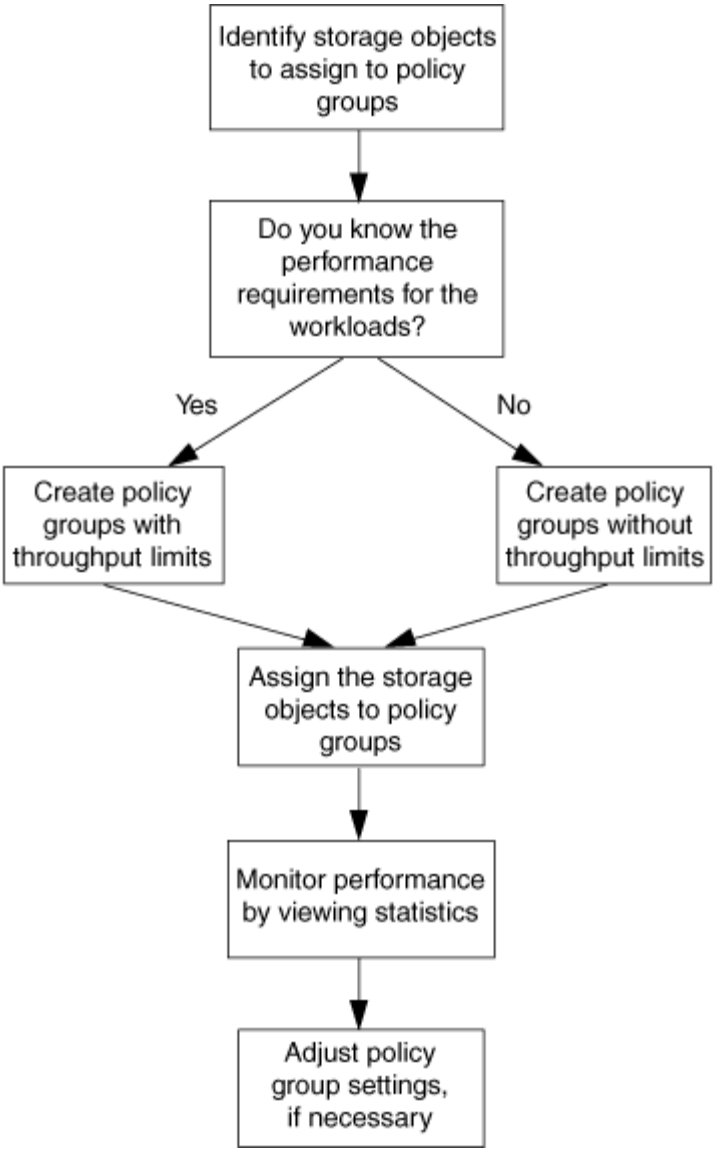
> 要在 MetroCluster 集群中禁用吞吐量下限 v2 ，必须运行
>
> `qos settings throughput-floors-v2 -enable false`
>
> 命令。

```
cluster1::*> qos settings throughput-floors-v2 -enable false
```

有关的详细信息 `qos settings throughput-floors-v2`，请参见"ONTAP 命令参考"。

ONTAP存储 QoS 工作流程

如果您已经知道要使用 QoS 管理的工作负载的性能要求，则可以在创建策略组时指定吞吐量限制。否则，您可以等到监控工作负载之后再指定限制。



在ONTAP中使用QoS设置吞吐量上限

您可以使用 `max-throughput` 用于定义存储对象工作负载吞吐量上限(QoS最大值)的策略组字段。您可以在创建或修改存储对象时应用策略组。

开始之前

- 您必须是集群管理员才能创建策略组。
- 要将策略组应用于 SVM ，您必须是集群管理员。

关于此任务

- 从 ONTAP 9.4 开始，您可以使用 *non-shared* QoS 策略组来指定定义的吞吐量上限适用场景每个成员工作负载。否则，策略组为 _shared： _ 分配给策略组的工作负载的总吞吐量不能超过指定的上限。

  设置 `-is-shared=false` 。 `qos policy-group create` 命令以指定非共享策略组。

- 您可以指定 IOPS ， MB/ 秒或 IOPS ， MB/ 秒上限的吞吐量限制如果同时指定 IOPS 和 MB/ 秒，则会强制执行首先达到的限制。

> 如果为同一工作负载设置了上限和下限，则只能以 IOPS 为单位指定上限的吞吐量限制。

- 受 QoS 限制的存储对象必须包含在策略组所属的 SVM 中。多个策略组可以属于同一个 SVM 。
- 如果某个存储对象的包含对象或子对象属于某个策略组，则不能将该存储对象分配给该策略组。
- QoS 最佳实践是将策略组应用于相同类型的存储对象。

步骤

1. 创建策略组。

   `qos policy-group create -policy-group` *`policy_group`* `-vserver` *`SVM`* `-max -throughput` *`number_of_iops`*`|Mb/S|iops,Mb/S -is-shared true|false`

   有关的详细信息 `qos policy-group create`，请参见"ONTAP 命令参考"。

   您可以使用 `qos policy-group modify`命令调整吞吐量上限。

   以下命令将创建共享策略组 `pg-vs1` 最大吞吐量为5、000次IOPS：

   ```
   cluster1::> qos policy-group create -policy-group pg-vs1 -vserver vs1
   -max-throughput 5000iops -is-shared true
   ```

   以下命令将创建非共享策略组 `pg-vs3` 最大吞吐量为100 IOPS和400 KB/秒：

   ```
   cluster1::> qos policy-group create -policy-group pg-vs3 -vserver vs3
   -max-throughput 100iops,400KB/s -is-shared false
   ```

   以下命令将创建非共享策略组 `pg-vs4` 无吞吐量限制：

   ```
   cluster1::> qos policy-group create -policy-group pg-vs4 -vserver vs4
   -is-shared false
   ```

   有关的详细信息 `qos policy-group modify`，请参见"ONTAP 命令参考"。

2. 将策略组应用于 SVM ，文件，卷或 LUN ：

   `storage_object` `create -vserver` *`SVM`* `-qos-policy-group` *`policy_group`*

   有关此过程中所述命令的更多信息，请参见"ONTAP 命令参考"。您可以使用 `*storage_object* modify`命令将其他策略组应用于存储对象。

   以下命令将应用策略组 `pg-vs1` 到SVM `vs1`：

   ```
   cluster1::> vserver create -vserver vs1 -qos-policy-group pg-vs1
   ```

以下命令将应用策略组 pg-app 到卷 app1 和 app2：

```
cluster1::> volume create -vserver vs2 -volume app1 -aggregate aggr1
-qos-policy-group pg-app
```

```
cluster1::> volume create -vserver vs2 -volume app2 -aggregate aggr1
-qos-policy-group pg-app
```

3. 监控策略组性能：

   qos statistics performance show

   有关的详细信息 qos statistics performance show，请参见"ONTAP 命令参考"。

   > 从集群监控性能。请勿使用主机上的工具监控性能。

   以下命令显示策略组性能：

```
cluster1::> qos statistics performance show
Policy Group           IOPS      Throughput   Latency
-------------------- -------- --------------- ----------
-total-                 12316       47.76MB/s  1264.00us
pg_vs1                   5008       19.56MB/s     2.45ms
_System-Best-Effort        62       13.36KB/s     4.13ms
_System-Background         30           0KB/s        0ms
```

4. 监控工作负载性能：

   qos statistics workload performance show

   > 从集群监控性能。请勿使用主机上的工具监控性能。

   以下命令显示工作负载性能：

```
cluster1::> qos statistics workload performance show
Workload          ID     IOPS      Throughput    Latency
--------------- ------ -------- ---------------- ----------
-total-              -    12320        47.84MB/s  1215.00us
app1-wid7967      7967     7219        28.20MB/s   319.00us
vs1-wid12279     12279     5026        19.63MB/s     2.52ms
_USERSPACE_APPS     14       55        10.92KB/s   236.00us
_Scan_Backgro..   5688       20            0KB/s        0ms
```

有关的详细信息 `qos statistics workload performance show`，请参见"ONTAP 命令参考"。

> 您可以使用 `qos statistics workload latency show`命令查看QoS工作负载的详细延迟统计信息。有关的详细信息 `qos statistics workload latency show`，请参见"ONTAP 命令参考"。

在ONTAP中使用QoS设置吞吐量下限

您可以使用 `min-throughput` 用于为存储对象工作负载定义吞吐量下限(QoS最小值)的策略组字段。您可以在创建或修改存储对象时应用策略组。从 ONTAP 9.8 开始，您可以以 IOPS 或 MBps 或 IOPS 和 MBps 为单位指定吞吐量下限。

开始之前

- 您必须是集群管理员才能创建策略组。
- 从ONTAP 9.131开始，您可以使用在SVM级别强制实施吞吐量下限自适应策略组模板。您不能在具有QoS策略组的SVM上设置自适应策略组模板。

关于此任务

- 从 ONTAP 9.4 开始，您可以使用 *non-shared* QoS 策略组来指定将定义的吞吐量下限分别应用于每个成员工作负载。只有在这种情况下，吞吐量下限的策略组才能应用于多个工作负载。

  设置 `-is-shared=false`。 `qos policy-group create` 命令以指定非共享策略组。

- 如果节点或聚合上的性能容量（余量）不足，则工作负载的吞吐量可能会低于指定的下限。
- 受 QoS 限制的存储对象必须包含在策略组所属的 SVM 中。多个策略组可以属于同一个 SVM 。
- QoS 最佳实践是将策略组应用于相同类型的存储对象。
- 定义吞吐量下限的策略组不能应用于 SVM 。

步骤

1. 检查节点或聚合上是否具有足够的性能容量，如中所述 "确定剩余性能容量"。
2. 创建策略组。

   ```
   qos policy-group create -policy group policy_group -vserver SVM -min
   -throughput qos_target -is-shared true|false
   ```

   有关的详细信息 `qos policy-group create`，请参见"ONTAP 命令参考"。

3. 您可以使用 `qos policy-group modify`命令调整吞吐量下限。

   以下命令将创建最小吞吐量为1、000 IOPS的共享策略组 `pg-vs2`：

   ```
   cluster1::> qos policy-group create -policy group pg-vs2 -vserver vs2
   -min-throughput 1000iops -is-shared true
   ```

以下命令将创建非共享策略组 `pg-vs4` 无吞吐量限制：

```
cluster1::> qos policy-group create -policy group pg-vs4 -vserver vs4
-is-shared false
```

有关的详细信息 `qos policy-group modify`，请参见"ONTAP 命令参考"。

4. 将策略组应用于卷或 LUN：

   `*storage_object* create -vserver *SVM* -qos-policy-group *policy_group*`您可以使用 `_storage_object_modify` 命令将不同的策略组应用到存储对象。

   以下命令将应用策略组 `pg-app2` 卷 `app2`：

```
cluster1::> volume create -vserver vs2 -volume app2 -aggregate aggr1
-qos-policy-group pg-app2
```

   有关此过程中所述命令的更多信息，请参见"ONTAP 命令参考"。

5. 监控策略组性能：

   `qos statistics performance show`

   > 从集群监控性能。请勿使用主机上的工具监控性能。

   以下命令显示策略组性能：

```
cluster1::> qos statistics performance show
Policy Group            IOPS      Throughput   Latency
-------------------- -------- --------------- ----------
-total-                  12316       47.76MB/s  1264.00us
pg_app2                   7216       28.19MB/s   420.00us
_System-Best-Effort         62       13.36KB/s     4.13ms
_System-Background          30           0KB/s        0ms
```

   有关的详细信息 `qos statistics performance show`，请参见"ONTAP 命令参考"。

6. 监控工作负载性能：

   `qos statistics workload performance show`

   > 从集群监控性能。请勿使用主机上的工具监控性能。

   以下命令显示工作负载性能：

```
cluster1::> qos statistics workload performance show
Workload          ID     IOPS      Throughput    Latency
--------------- ------ -------- ---------------- ----------
-total-              -    12320        47.84MB/s  1215.00us
app2-wid7967      7967     7219        28.20MB/s   319.00us
vs1-wid12279     12279     5026        19.63MB/s     2.52ms
_USERSPACE_APPS     14       55        10.92KB/s   236.00us
_Scan_Backgro..   5688       20            0KB/s        0ms
```

有关的详细信息 `qos statistics workload performance show`，请参见"ONTAP 命令参考"。

> 您可以使用 `qos statistics workload latency show`命令查看QoS工作负载的详细延迟统计信息。有关的详细信息 `qos statistics workload latency show`，请参见"ONTAP 命令参考"。

在ONTAP中使用自适应QoS策略组

您可以使用_自适应 QoS_策略组，根据卷大小自动调整吞吐量上限或下限，从而在卷大小变化时保持 IOPS 与 TB/GB 的比率。在大型部署中管理成百上千个工作负载时，这是一个显著的优势。

开始之前

- 您必须运行ONTAP 9.3或更高版本。自 ONTAP 9.3 起，可使用自适应 QoS 策略组。
- 您必须是集群管理员才能创建策略组。

关于此任务

存储对象可以是自适应策略组或非自适应策略组的成员，但不能同时是这两者。存储对象的 SVM 和策略必须相同。存储对象必须处于联机状态。

自适应 QoS 策略组始终为非共享组：定义的吞吐量上限或每个成员工作负载的下限适用场景。

吞吐量限制与存储对象大小的比率取决于以下字段的交互：

- `expected-iops`是每分配 TB/GB 的最低预期 IOPS。

> `expected-iops`仅在AFF平台上提供担保。 `expected-iops`只有当层策略设置为"无"且云中没有任何块时、才保证适用于FabricPool。 `expected-iops`保证适用于非SnapMirror同步关系中的卷。

- `peak-iops`是每分配或使用的 TB/GB 的最大可能 IOPS。
- `expected-iops-allocation` 指定是将已分配空间(默认值)还是已用空间用于预期IOPS。

> expected-iops-allocation 适用于ONTAP 9.5及更高版本。ONTAP 9.4 及更早版本不支持此功能。

- peak-iops-allocation 指定是将已分配空间用于还是将已用空间(默认值)用于 peak-iops。
- absolute-min-iops 是绝对最小IOPS数。您可以对非常小的存储对象使用此字段。它会同时覆盖这两者 peak-iops 和 / 或 expected-iops 时间 absolute-min-iops 大于计算得出的 expected-iops。

  例如、如果您设置了 expected-iops 计算得出的值为1、000 IOS/TB、并且卷大小小于1 GB expected-iops 将是部分IOP。计算的 peak-iops 将是一个更小的比例。您可以通过设置来避免此问题 absolute-min-iops 以获得实际价值。

- block-size 指定应用程序I/O块大小。默认值为 32 K 。有效值为 8K ， 16K ， 32K ， 64K ，任意。any 表示不强制实施块大小。

默认自适应 QoS 策略组

可用的三个默认自适应 QoS 策略组如下表所示。您可以将这些策略组直接应用于卷。

默认策略组	预期 IOPS/TB	峰值 IOPS/TB	绝对最小 IOPS
extreme	6、144	12、288	1000
performance	2 ， 048	4、096	500
value	128.	512	75

存储对象策略组分配限制

在某些情况下，如果存储对象包含的对象或其子对象属于某个策略组，则无法将存储对象分配给该策略组。

从ONTAP 9.18.1 开始，您可以使用嵌套的 QoS 策略，该策略允许将策略组分配给包含对象（例如 SVM）及其子对象（例如卷）。在多租户环境中，嵌套的 QoS 策略使管理员能够将 SVM 上的 QoS 限制细分到 SVM 内的各个卷和 qtree，从而能够在计算环境中平衡存储资源，同时还能优先处理关键任务工作负载。

以下对象对支持嵌套 QoS 策略：

- SVM 和 SVM 包含的FlexVol或FlexGroup卷。
- FlexVol或FlexGroup卷以及卷内的 qtree。

对于嵌套的 QoS 策略，采用最严格的适用策略。

下表列出了各项限制。

如果您分配 ...	那么你就无法将其分配到策略组.....
SVM 到策略组	SVM 中包含的任何存储对象。 如果您运行的是ONTAP 9.18.1，则 SVM 中包含的FlexVol和FlexGroup卷_可以_分配给策略组。
卷到策略组	包含该卷或任何子 LUN 的 SVM。 如果您运行的是ONTAP 9.18.1 及更高版本，则包含卷的 SVM _可以_分配给策略组。此外，还可以分配FlexVol或FlexGroup卷中的 qtree。
将LUN分配给策略组	包含 LUN 的卷或 SVM
文件到策略组	包含文件的卷或 SVM

步骤

1. 创建自适应 QoS 策略组：

   ```
   qos adaptive-policy-group create -policy group policy_group -vserver SVM
   -expected-iops number_of_iops/TB|GB -peak-iops number_of_iops/TB|GB -expected
   -iops-allocation-space|used-space -peak-iops-allocation allocated-space|used-
   space -absolute-min-iops number_of_iops -block-size 8K|16K|32K|64K|ANY
   ```

   有关的详细信息 `qos adaptive-policy-group create`，请参见"ONTAP 命令参考"。

   > `-expected-iops-allocation` 和 `-block-size` 适用于ONTAP 9.5及更高版本。ONTAP 9.4 及更早版本不支持这些选项。

   以下命令将创建自适应QoS策略组 `adpg-app1` 使用 `-expected-iops` 设置为300 IOS/TB、 `-peak-iops` 设置为1、000 IOS/TB、 `-peak-iops-allocation` 设置为 `used-space`，和 `-absolute-min-iops` 设置为50 IOPS：

   ```
   cluster1::> qos adaptive-policy-group create -policy group adpg-app1
   -vserver vs2 -expected-iops 300iops/tb -peak-iops 1000iops/TB -peak-iops
   -allocation used-space -absolute-min-iops 50iops
   ```

2. 将自适应 QoS 策略组应用于卷：

   ```
   volume create -vserver SVM -volume volume -aggregate aggregate -size number_of
   TB|GB -qos-adaptive-policy-group policy_group
   ```

   有关的详细信息 `volume create`，请参见"ONTAP 命令参考"。

以下命令将应用自适应QoS策略组 `adpg-app1` 到卷 `app1`：

```
cluster1::> volume create -vserver vs1 -volume app1 -aggregate aggr1
-size 2TB -qos-adaptive-policy-group adpg-app1
```

以下命令将应用默认自适应QoS策略组 `extreme` 到新卷 `app4` 和到现有卷 `app5`。为策略组适用场景卷定义的吞吐量上限 `app4` 和 `app5` 单独：

```
cluster1::> volume create -vserver vs4 -volume app4 -aggregate aggr4
-size 2TB -qos-adaptive-policy-group extreme
```

```
cluster1::> volume modify -vserver vs5 -volume app5 -qos-adaptive-policy
-group extreme
```

在ONTAP中设置自适应策略组模板

从ONTAP 9.13.1开始、您可以使用自适应策略组模板在SVM级别强制实施吞吐量下限和上限。

关于此任务

- 自适应策略组模板是默认策略 `apg1`。可以随时修改此策略。它只能通过命令行界面或ONTAP REST API进行设置、并且只能应用于现有SVM。
- 自适应策略组模板仅会影响在SVM上创建或迁移到该SVM的卷。SVM上的现有卷将保留其现有状态。

  如果禁用自适应策略组模板、SVM上的卷将保留其现有策略。只有随后在SVM上创建或迁移到SVM的卷才会受此功能的影响。

- 您不能在具有QoS策略组的SVM上设置自适应策略组模板。
- 自适应策略组模板专为AFF 平台而设计。可以在其他平台上设置自适应策略组模板、但该策略可能不会强制实施最小吞吐量。同样、您可以向FabricPool聚合或不支持最小吞吐量的聚合中的SVM添加自适应策略组模板、但不会强制实施吞吐量下限。
- 如果SVM采用MetroCluster 配置或SnapMirror关系、则会在镜像的SVM上强制实施自适应策略组模板。

步骤

1. 修改SVM以应用自适应策略组模板： `vserver modify -qos-adaptive-policy-group-template apg1`
2. 确认已设置策略： `vserver show -fields qos-adaptive-policy-group`

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。