



# 管理工作負載

## ONTAP 9

NetApp  
February 12, 2026

# 目錄

管理工作負載 .....	1
識別 ONTAP 中的剩餘效能容量 .....	1
識別 ONTAP 中的高流量用戶端或檔案 .....	2
保證QoS的處理量 .....	4
透過 ONTAP 中的 QoS 總覽來保證傳輸量 .....	4
啟用或停用ONTAP吞吐量底線 v2 .....	9
ONTAP儲存 QoS 工作流程 .....	10
在 ONTAP 中設定 QoS 的處理量上限 .....	11
在 ONTAP 中設定具有 QoS 的處理量層 .....	14
在 ONTAP 中使用調適性 QoS 原則群組 .....	16
在 ONTAP 中設定調適性原則群組範本 .....	19

# 管理工作負載

## 識別 ONTAP 中的剩餘效能容量

效能容量（或稱「\_E餘量」）會測量在資源上的工作負載效能開始受到延遲影響之前、您可以在節點或集合體上放置多少工作。瞭解叢集上可用的效能容量、有助於您配置及平衡工作負載。

開始之前

此工作需要進階權限層級命令。

關於這項工作

您可以將下列值用於 `-object` 用於收集和顯示保留空間統計資料的選項：

- 對於 CPU、`resource_headroom_cpu`。
- 對於集合體、`resource_headroom_aggr`。

您也可以使用 System Manager 和 Active IQ Unified Manager 整套功能來完成這項工作。

步驟

1. 變更為進階權限層級：

```
set -privilege advanced
```

2. 開始即時保留空間統計資料收集：

```
statistics start -object resource_headroom_cpu|aggr
```

如"[指令參考資料ONTAP](#)"需詳細 `statistics start` 資訊，請參閱。

3. 顯示即時保留空間統計資訊：

```
statistics show -object resource_headroom_cpu|aggr
```

4. 返回管理權限：

```
set -privilege admin
```

範例

下列範例顯示叢集節點的平均每小時保留空間統計資料。

您可以透過減去來計算節點的可用效能容量 `current_utilization` 來自的計數器 `optimal_point_utilization` 計數器。在此範例中、是的使用率容量 `CPU_sti2520-213` 為 -14%（72% 至 86%）、表示過去一小時的 CPU 平均利用率過高。

您可以指定 `ewma_daily`、`ewma_weekly` 或 `ewma_monthly` 以獲得較長時間內的相同資訊平均值。

```
sti2520-2131454963690::*> statistics show -object resource_headroom_cpu
-raw -counter ewma_hourly
(statistics show)
```

```
Object: resource_headroom_cpu
Instance: CPU_sti2520-213
Start-time: 2/9/2016 16:06:27
End-time: 2/9/2016 16:06:27
Scope: sti2520-213
```

Counter	Value
-----	-----
ewma_hourly	-
current_ops	4376
current_latency	37719
current_utilization	86
optimal_point_ops	2573
optimal_point_latency	3589
optimal_point_utilization	72
optimal_point_confidence_factor	1

```
Object: resource_headroom_cpu
Instance: CPU_sti2520-214
Start-time: 2/9/2016 16:06:27
End-time: 2/9/2016 16:06:27
Scope: sti2520-214
```

Counter	Value
-----	-----
ewma_hourly	-
current_ops	0
current_latency	0
current_utilization	0
optimal_point_ops	0
optimal_point_latency	0
optimal_point_utilization	71
optimal_point_confidence_factor	1

2 entries were displayed.

如"[指令參考資料ONTAP](#)"需詳細 `statistics show` 資訊，請參閱。

## 識別 ONTAP 中的高流量用戶端或檔案

您可以使用ONTAP「支援物件」技術來識別造成大量叢集流量的用戶端或檔案。識別出這些「頂尖」用戶端或檔案之後、您可以重新平衡叢集工作負載、或採取其他步驟來解決問

題。

開始之前

您必須是叢集管理員才能執行此工作。

步驟

1. 檢視存取叢集的主要用戶端：

```
statistics top client show -node node_name -sort-key sort_column -interval  
seconds_between_updates -iterations iterations -max number_of_instances
```

下列命令會顯示存取的主要用戶端 cluster1：

```
cluster1::> statistics top client show  
  
cluster1 : 3/23/2016 17:59:10  
  
                Client Vserver                Node Protocol    *Total  
-----
```

Client	Vserver	Node	Protocol	*Total Ops
172.17.180.170	vs4	siderop1-vs4	nfs	668
172.17.180.169	vs3	siderop1-vs3	nfs	337
172.17.180.171	vs3	siderop1-vs3	nfs	142
172.17.180.170	vs3	siderop1-vs3	nfs	137
172.17.180.123	vs3	siderop1-vs3	nfs	137
172.17.180.171	vs4	siderop1-vs4	nfs	95
172.17.180.169	vs4	siderop1-vs4	nfs	92
172.17.180.123	vs4	siderop1-vs4	nfs	92
172.17.180.153	vs3	siderop1-vs3	nfs	0

如"指令參考資料ONTAP"需詳細 `statistics top client show` 資訊，請參閱。

2. 檢視叢集上存取的主要檔案：

```
statistics top file show -node node_name -sort-key sort_column -interval  
seconds_between_updates -iterations iterations -max number_of_instances
```

下列命令會顯示在上存取的最上層檔案 cluster1：

```

cluster1::> statistics top file show

cluster1 : 3/23/2016 17:59:10

                                *Total
                                File Volume Vserver          Node      Ops
-----
/vol/vol1/vm170-read.dat      vol1      vs4 siderop1-vs4      22
/vol/vol1/vm69-write.dat     vol1      vs3 siderop1-vs3       6
/vol/vol2/vm171.dat          vol2      vs3 siderop1-vs3       2
/vol/vol2/vm169.dat          vol2      vs3 siderop1-vs3       2
/vol/vol2/p123.dat           vol2      vs4 siderop1-vs4       2
/vol/vol2/p123.dat           vol2      vs3 siderop1-vs3       2
/vol/vol1/vm171.dat          vol1      vs4 siderop1-vs4       2
/vol/vol1/vm169.dat          vol1      vs4 siderop1-vs4       2
/vol/vol1/vm169.dat          vol1      vs4 siderop1-vs3       2
/vol/vol1/p123.dat           vol1      vs4 siderop1-vs4       2

```

如"指令參考資料ONTAP"需詳細 `statistics top file show` 資訊，請參閱。

## 保證QoS的處理量

透過 **ONTAP** 中的 **QoS** 總覽來保證傳輸量

您可以使用儲存服務品質 (QoS) 來保證關鍵工作負載的效能不會因競爭工作負載而降級。您可以在競爭的工作負載上設定處理量上限、以限制其對系統資源的影響、或為關鍵工作負載設定處理量下限、以確保其符合最低處理量目標、無論競爭的工作負載有何需求。您甚至可以針對相同的工作負載設定上限和樓層。

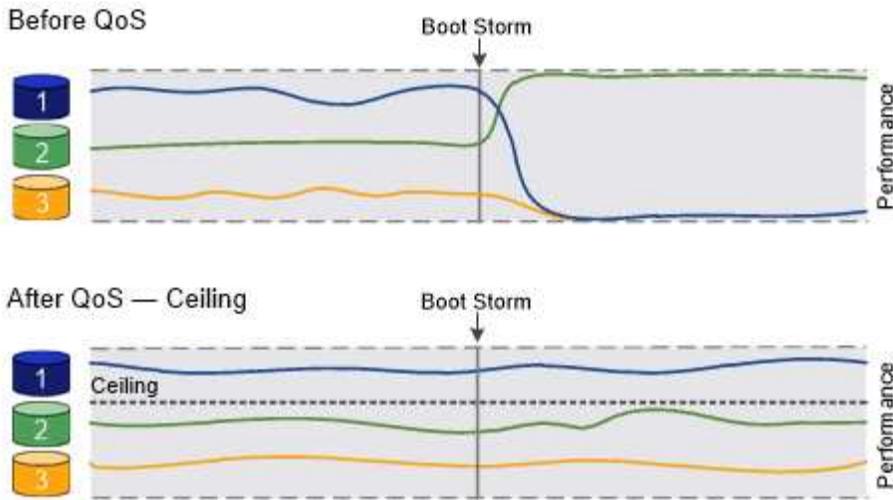
處理量上限 ( **QoS** 上限)

處理量上限會將工作負載的處理量限制為最大IOPS或Mbps、或IOPS和Mbps。在下圖中、工作負載2的處理量上限可確保工作負載1和3不會「凸起」。

原則群組 定義一或多個工作負載的處理量上限。工作負載代表 storage 物件的I/O作業：磁碟區、檔案、qtrees或LUN、或SVM中的所有磁碟區、檔案、qtrees或LUN。您可以在建立原則群組時指定上限、也可以等到監控工作負載之後再指定上限。



工作負載的處理量可能超過指定上限10%、尤其是當工作負載處理量發生快速變化時。處理突發的上限可能超過50%。當權杖累積率高達150%時、單一節點上就會發生突發事件



### 處理量層（QoS 最低）

處理量層保證工作負載的處理量不會低於 IOPS 或 Mbps 的最低數量、或 IOPS 和 Mbps。在下圖中、工作負載1和工作負載3的處理量層級可確保它們符合最低處理量目標、無論工作負載2的需求為何。



如範例所示、處理量上限會直接調節處理量。處理量最低層會間接調節處理量、將已設定最低層的工作負載設為優先順序。

您可以在建立原則群組時指定樓層、也可以等到監控工作負載之後再指定樓層。

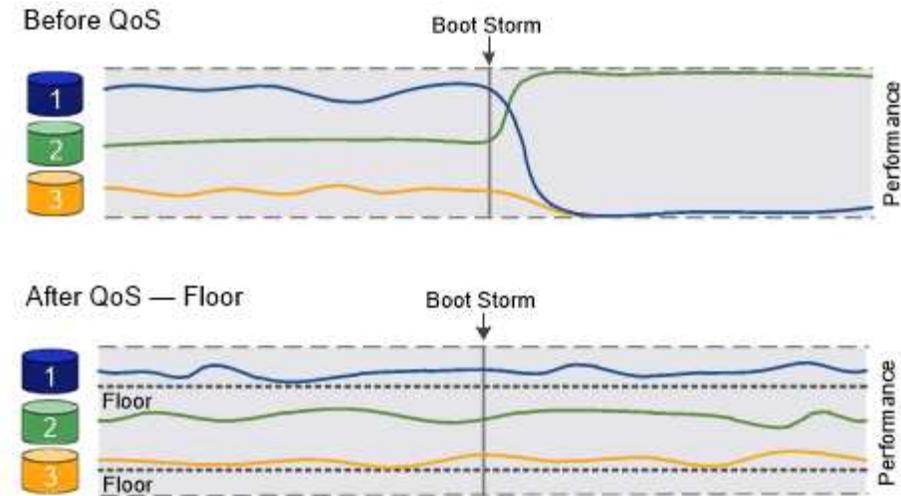
從 ONTAP 9.13.1 開始[[adaptive-qos-templates](#)]，您可以使用設定 SVM 範圍的處理量層。在 9.13.1 之前的 ONTAP 版本中、定義處理量層的原則群組無法套用至 SVM。



在更新於 ONTAP VMware 版本 9.7 之前的版本中、當有足夠的可用效能容量時、就能保證處理量的樓層數。

在 VMware 9.7 及更新版本中、即使可用的效能容量不足、也能保證處理量的樓層數。ONTAP 這種新的樓層行為稱為 Floor v2。為了達成保證、第 v2 層級可能會導致工作負載延遲更高、而不會出現處理量層或工作環境超過場地設定的情況。第 v2 層同時適用於 QoS 和調適性 QoS。

ONTAP 9.7P6 及更新版本提供啟用 / 停用第 2 層樓的新行為選項。工作負載在執行關鍵作業時、可能會低於指定的樓層、例如 `volume move trigger-cutover`。即使有足夠的可用容量且未執行關鍵作業、工作負載的處理量仍可能低於指定樓層達 5%。如果場地配置過度、而且沒有效能容量、則部分工作負載可能會落在指定樓層以下。



### 共享和非共享的 QoS 原則群組

從ONTAP S得9.4開始、您可以使用非共享的QoS原則群組、來指定定義的處理量上限或樓層分別套用至每個成員工作負載。共享原則群組的行為取決於原則類型：

- 對於處理量上限、指派給共用原則群組的工作負載總處理量不得超過指定上限。
- 對於處理量層、共用原則群組只能套用至單一工作負載。

### 調適性QoS

通常、您指派給儲存物件的原則群組值會固定。當儲存物件大小變更時、您需要手動變更值。例如、增加磁碟區上使用的空間量、通常需要為磁碟區指定的處理量上限相應增加。

Adaptive QoS 會自動將原則群組值調整為工作負載大小、並隨著工作負載大小的變更、維持IOPS與TBs的比率。當您在大型部署中管理數百或數千個工作負載時、這是一項重大優勢。

您通常會使用調適性QoS來調整處理量上限、但也可以使用它來管理處理量層（當工作負載大小增加時）。工作負載大小是以儲存物件的已配置空間或儲存物件所使用的空間表示。



在ONTAP 更新版本的更新版本中、可在處理量層使用已用的空間。不支援ONTAP 使用於效能不符合更新版本的資料層。

- allocated space 原則會根據儲存物件的名義大小、維持IOPS/TB|GB比率。如果比率為100 IOPS/GB、則150 GB的磁碟區只要磁碟區維持該大小、就會有15,000 IOPS的處理量上限。如果磁碟區大小調整為300 GB、調適性QoS會將處理量上限調整為30、000 IOPS。
- used space 原則（預設值）會根據儲存效率前的實際資料量、維持IOPS/TB|GB比率。如果比率為100 IOPS/GB、則儲存100 GB資料的150 GB磁碟區的處理量上限為10、000 IOPS。隨著使用空間量的變化、調適性QoS會根據比率調整處理量上限。

從功能支援的9.5開始ONTAP、您可以為應用程式指定I/O區塊大小、以IOPS和Mbps來表示處理量限制。Mbps限制是根據區塊大小乘以IOPS限制計算而得。例如、IOPS限制為6144IOPS/TB的I/O區塊大小為32K、會產生192 MBps的Mbps限制。

處理量上限和樓層的行為如下：

- 當工作負載指派給調適性QoS原則群組時、上限或樓層會立即更新。

- 調整調適性QoS原則群組中的工作負載大小時、上限或樓層大約會在五分鐘內更新。

在進行更新之前、處理量必須增加至少10 IOPS。

調適性QoS原則群組永遠不會共用：定義的處理量上限或樓層會個別套用至每個成員的工作負載。

從 ONTAP 9.6 開始、透過 SSD 的 ONTAP Select Premium 即可支援處理量層。

#### 調適性原則群組範本

從 ONTAP 9.13.1 開始，您可以在 SVM 上設定調適性 QoS 範本。可調整的原則群組範本可讓您設定 SVM 中所有磁碟區的處理量層和上限。

自適應原則群組範本只能在建立 SVM 之後設定。使用 `vserver modify` 命令 `-qos-adaptive-policy -group-template` 設定原則的參數。

當您設定調適性原則群組範本時、在設定原則之後建立或移轉的磁碟區會自動繼承原則。指派原則範本時、SVM 上現有的任何磁碟區都不會受到影響。如果停用 SVM 上的原則、任何後來移轉到 SVM 或在 SVM 上建立的磁碟區都不會收到原則。停用調適性原則群組範本不會影響繼承原則範本的磁碟區、因為這些磁碟區會保留原則範本。

如需詳細資訊、請參閱 [設定調適性原則群組範本](#)。

#### 一般支援

下表顯示支援處理量上限、處理量層和調適性QoS的差異。

資源或功能	處理量上限	處理量最低	處理量層v2	調適性QoS
版本9 ONTAP	全部	9.2 及更新版本	9.7 及更新版本	9.3 及更新版本
平台	全部	<ul style="list-style-type: none"> <li>• AFF</li> <li>• C190 <sup>1</sup></li> <li>• ONTAP Select Premium 搭配 SSD <sup>1</sup></li> </ul>	<ul style="list-style-type: none"> <li>• AFF</li> <li>• C190</li> <li>• 搭載SSD的高階版ONTAP Select</li> </ul>	全部
通訊協定	全部	全部	全部	全部
FabricPool	是的	是的、如果分層原則設定為「無」、而且雲端中沒有區塊。	是的、如果分層原則設定為「無」、而且雲端中沒有區塊。	否
SnapMirror同步	是的	否	否	是的

<sup>1</sup> C190 與 ONTAP Select 支援從 ONTAP 9.6 版本開始。

## 處理量上限支援的工作負載

下表顯示ONTAP 支援各個版本的工作負載、以支援不同版本的處理量上限。不支援根磁碟區、負載共用鏡像和資料保護鏡像。

工作負載支援	更新版本ONTAP	ONTAP 9.7 至 9.4	ONTAP 9.3 及更早版本
Volume	是的	是的	是的
檔案	是的	是的	是的
LUN	是的	是的	是的
SVM	是的	是的	是的
流通量FlexGroup	是的	是的	是（僅限 ONTAP 9.3）
qtree <sup>1</sup>	是的	否	否
每個原則群組有多個工作負載	是的	是的	是的
非共用原則群組	是的	是的	否

<sup>1</sup> 從ONTAP 9.9.1 開始，啟用 SMB 的FlexVol和FlexGroup磁碟區中的 qtree 也支援 SMB 存取。從ONTAP 9.8 開始，啟用 NFS 的FlexVol和FlexGroup磁碟區中的 qtree 支援 NFS 存取。

## 處理量層的支援工作負載

下表顯示ONTAP 支援各個版本之資料中心的工作負載。不支援根磁碟區、負載共用鏡像和資料保護鏡像。

工作負載支援	ONTAP 9.13.1 及更新版本	零點9.8至9.13.0 ONTAP	ONTAP 9.4 至 9.7	ONTAP 9.3
Volume	是的	是的	是的	是的
檔案	是的	是的	是的	是的
LUN	是的	是的	是的	是的
SVM	是的	否	否	否
流通量FlexGroup	是的	是的	是的	否
qtree <sup>1</sup>	是的	是的	否	否
每個原則群組有多個工作負載	是的	是的	是的	否
非共用原則群組	是的	是的	是的	否

<sup>1</sup> 從 ONTAP 9.8 開始，FlexVol 的 qtree 和啟用 NFS 的 FlexGroup 磁碟區都支援 NFS 存取。從ONTAP 推出支

援SMB的支援範圍起、FlexVol 從推出支援SMB的功能、即可從功能支援功能支援使用功能的功能性功能、從功能性功能表中的qtree和FlexGroup 功能表中、存取SMB。

### 支援調適性QoS的工作負載

下表顯示ONTAP 支援各更新版本的調適性QoS的工作負載。不支援根磁碟區、負載共用鏡像和資料保護鏡像。

工作負載支援	ONTAP 9.13.1 及更新版本	零點9.4至9.13.0 ONTAP	ONTAP 9.3
Volume	是的	是的	是的
檔案	是的	是的	否
LUN	是的	是的	否
SVM	是的	否	否
流通量FlexGroup	是的	是的	否
每個原則群組有多個工作負載	是的	是的	是的
非共用原則群組	是的	是的	是的

### 工作負載和原則群組的最大數量

下表顯示ONTAP 各個版本的工作負載和原則群組數量上限。

工作負載支援	更新版本ONTAP	ONTAP 9.3 及更早版本
每個叢集的工作負載上限	40、000	12、000
每個節點的工作負載上限	40、000	12、000
原則群組上限	12、000	12、000

### 啟用或停用ONTAP吞吐量底線 v2

您可以啟用AFF 或停用支援速度的v2。預設為啟用。啟用第v2層時、如果控制器使用頻繁、而其他工作負載的延遲較高、則可滿足處理量層級的需求。第v2層同時適用於QoS和調適性QoS。

#### 步驟

1. 變更為進階權限層級：

```
set -privilege advanced
```

2. 輸入下列其中一個命令：

如果您想要...	使用此命令：
停用樓層v2	<code>qos settings throughput-floors-v2 -enable false</code>
啟用樓層v2	<code>qos settings throughput-floors-v2 -enable true</code>



若要在MetroCluster 一個不穩定叢集中停用處理量層v2、您必須執行

`qos settings throughput-floors-v2 -enable false`

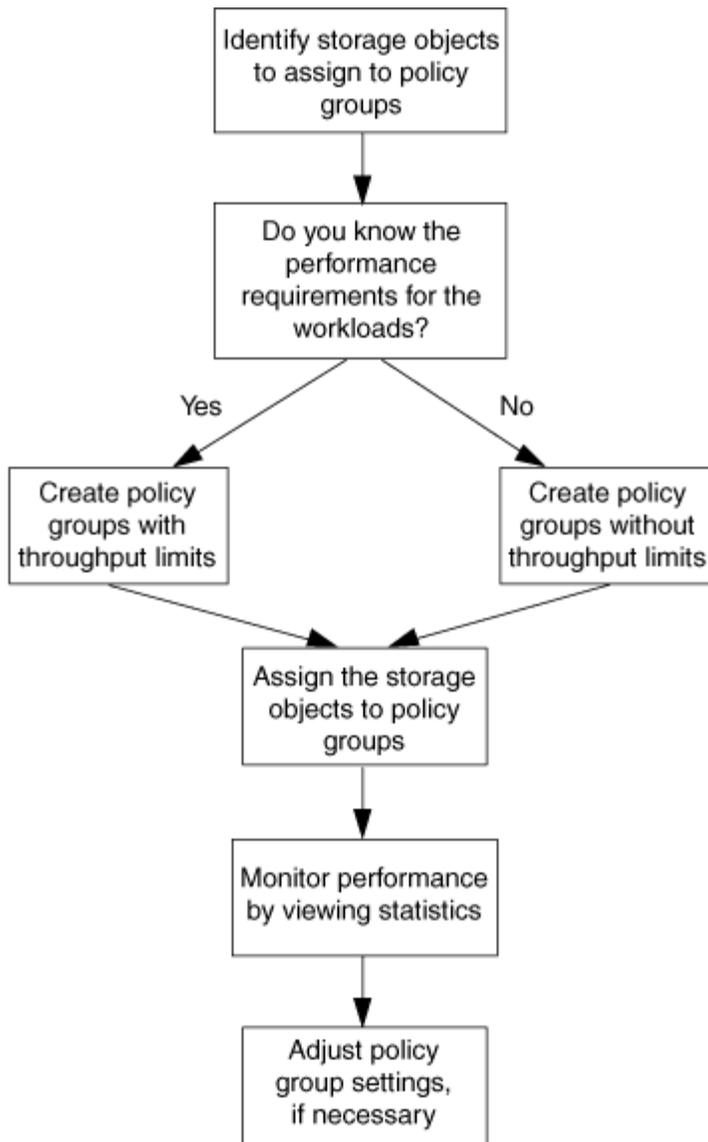
在來源叢集和目的地叢集上執行命令。

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

如"[指令參考資料ONTAP](#)"需詳細 `qos settings throughput-floors-v2` 資訊，請參閱。

## ONTAP儲存 QoS 工作流程

如果您已經知道想要使用QoS管理的工作負載效能需求、可以在建立原則群組時指定處理量限制。否則、您可以等到監控工作負載之後再指定限制。



## 在 ONTAP 中設定 QoS 的處理量上限

您可以使用 `max-throughput` 原則群組的欄位、可定義儲存物件工作負載的處理量上限（QoS Max）。您可以在建立或修改儲存物件時套用原則群組。

### 開始之前

- 您必須是叢集管理員、才能建立原則群組。
- 您必須是叢集管理員、才能將原則群組套用至SVM。

### 關於這項工作

- 從ONTAP S得9.4開始、您可以使用非共享的QoS原則群組、來指定定義的處理量上限會個別套用至每個成員工作負載。否則、原則群組會是 `_shared`：\_指派給原則群組的工作負載總處理量不能超過指定上限。

設定 `-is-shared=false` 適用於 `qos policy-group create` 指定非共用原則群組的命令。

- 您可以指定IOPS、MB/s或IOPS、MB/s等上限的處理量限制如果您同時指定IOPS和MB/s、則會強制執行先達到的限制。



如果您為相同的工作負載設定上限和樓層、則只能指定上限的處理量限制（以IOPS為單位）。

- 受QoS限制的儲存物件必須由原則群組所屬的SVM所包含。多個原則群組可以屬於同一個SVM。
- 如果原則群組包含物件或其子物件屬於原則群組、則無法將儲存物件指派給原則群組。
- 將原則群組套用至相同類型的儲存物件、是QoS最佳實務做法。

## 步驟

### 1. 建立原則群組：

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

如"[指令參考資料ONTAP](#)"需詳細 `qos policy-group create` 資訊，請參閱。

您可以使用 `qos policy-group modify` 命令來調整處理量上限。

下列命令會建立共用原則群組 `pg-vs1` 最大處理量為 5、000 IOPS：

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

下列命令會建立非共用原則群組 `pg-vs3` 最高處理量為 100 IOPS 和 400 kb/S：

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

下列命令會建立非共用原則群組 `pg-vs4` 無處理量限制：

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

如"[指令參考資料ONTAP](#)"需詳細 `qos policy-group modify` 資訊，請參閱。

### 2. 將原則群組套用至SVM、檔案、磁碟區或LUN：

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

如需有關本程序中所述命令"[指令參考資料ONTAP](#)"的詳細資訊，請參閱。您可以使用 `storage\_object modify` 命令將不同的原則群組套用至儲存物件。

下列命令會套用原則群組 `pg-vs1` 至 SVM `vs1`：

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

下列命令會套用原則群組 pg-app 磁碟區 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. 監控原則群組效能：

```
qos statistics performance show
```

如"指令參考資料ONTAP"需詳細 `qos statistics performance show` 資訊，請參閱。



監控叢集的效能。請勿使用主機上的工具來監控效能。

下列命令顯示原則群組效能：

```
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. 監控工作負載效能：

```
qos statistics workload performance show
```



監控叢集的效能。請勿使用主機上的工具來監控效能。

下列命令顯示工作負載效能：

```
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
```

如"[指令參考資料ONTAP](#)"需詳細 `qos statistics workload performance show` 資訊，請參閱。



您可以使用 `qos statistics workload latency show` 命令來檢視 QoS 工作負載的詳細延遲統計資料。如"[指令參考資料ONTAP](#)"需詳細 `qos statistics workload latency show` 資訊，請參閱。

## 在 ONTAP 中設定具有 QoS 的處理量層

您可以使用 `min-throughput` 原則群組的欄位、可定義儲存物件工作負載的處理量層（QoS 最小值）。您可以在建立或修改儲存物件時套用原則群組。從 ONTAP 功能性的問題 9.8 開始、您可以指定處理量層（以 IOPS 或 Mbps 為單位）、或是 IOPS 和 Mbps。

### 開始之前

- 您必須是叢集管理員、才能建立原則群組。
- 從 ONTAP 9.13.1 開始，您可以使用在 SVM 層級強制執行處理量層級 [調適性原則群組範本](#)。您無法在具有 QoS 原則群組的 SVM 上設定調適性原則群組範本。

### 關於這項工作

- 從 ONTAP 9.4 開始、您可以使用非共享的 QoS 原則群組來指定要個別套用定義的處理量層級至每個成員工作負載。這是處理量層的原則群組可套用至多個工作負載的唯一條件。

設定 `-is-shared=false` 適用於 `qos policy-group create` 指定非共用原則群組的命令。

- 如果節點或 Aggregate 上的效能容量（保留空間）不足、則工作負載的處理量可能會低於指定樓層。
- 受 QoS 限制的儲存物件必須由原則群組所屬的 SVM 所包含。多個原則群組可以屬於同一個 SVM。
- 將原則群組套用至相同類型的儲存物件、是 QoS 最佳實務做法。
- 定義處理量層的原則群組無法套用至 SVM。

### 步驟

1. 如所述、檢查節點或集合體上是否有足夠的效能容量 "[識別剩餘的效能容量](#)"。
2. 建立原則群組：

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

如"[指令參考資料ONTAP](#)"需詳細 `qos policy-group create` 資訊，請參閱。

3. 您可以使用 `qos policy-group modify` 命令來調整處理量層。

下列命令會建立最小處理量為 1,000 IOPS 的共用原則群組 pg-vs2：

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

下列命令會建立非共用原則群組 pg-vs4 無處理量限制：

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

如"[指令參考資料ONTAP](#)"需詳細 `qos policy-group modify` 資訊，請參閱。

4. 將原則群組套用至磁碟區或LUN：

`*storage\_object create -vserver SVM -qos-policy-group policy\_group*` 您可以使用 `*\_storage\_object\_modify*` 命令將不同的策略群組套用到儲存物件。

下列命令會套用原則群組 pg-app2 磁碟區 app2：

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

如需有關本程序中所述命令"[指令參考資料ONTAP](#)"的詳細資訊，請參閱。

5. 監控原則群組效能：

```
qos statistics performance show
```



監控叢集的效能。請勿使用主機上的工具來監控效能。

下列命令顯示原則群組效能：

```
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
```

如"指令參考資料ONTAP"需詳細 `qos statistics performance show` 資訊，請參閱。

## 6. 監控工作負載效能：

```
qos statistics workload performance show
```



監控叢集的效能。請勿使用主機上的工具來監控效能。

下列命令顯示工作負載效能：

```
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
```

如"指令參考資料ONTAP"需詳細 `qos statistics workload performance show` 資訊，請參閱。



您可以使用 `qos statistics workload latency show` 命令來檢視 QoS 工作負載的詳細延遲統計資料。如"指令參考資料ONTAP"需詳細 `qos statistics workload latency show` 資訊，請參閱。

## 在 ONTAP 中使用調適性 QoS 原則群組

您可以使用 `_自適應 QoS_` 策略群組，根據磁碟區大小自動調整吞吐量上限或下限，從而在磁碟區大小變化時保持 IOPS 與 TB/GB 的比率。在大型部署中管理數百上千個工作負載時，這是一個顯著的優勢。

### 開始之前

- 您必須執行 ONTAP 9.3 或更新版本。自 ONTAP 功能性 QoS 原則群組開始提供、從功能性的 9.3 開始提供。
- 您必須是叢集管理員、才能建立原則群組。

### 關於這項工作

儲存物件可以是調適性原則群組或非調適性原則群組的成員、但不能同時是兩者。儲存物件的 SVM 和原則必須相同。儲存物件必須處於線上狀態。

調適性 QoS 原則群組永遠不會共用：定義的處理量上限或樓層會個別套用至每個成員的工作負載。

處理量限制與儲存物件大小的比率、取決於下列欄位的互動：

- ``expected-iops`` 是每分配 TB/GB 的最低預期 IOPS。



`expected-iops` 僅在 AFF 平台上保證。`expected-iops` 只有當分層原則設定為「無」且雲端中沒有區塊時、FabricPool 才保證適用。`expected-iops` 保證不屬於 SnapMirror 同步關係的磁碟區。

- `peak-iops` 是每分配或使用的 TB/GB 的最大可能 IOPS。
- `expected-iops-allocation` 指定是將分配的空間（預設）還是已使用的空間用於預期的 IOPS。



`expected-iops-allocation` 可在 ONTAP 9.5 或更新版本中取得。不支援 ONTAP 此功能。

- `peak-iops-allocation` 指定是使用分配的空間還是使用的空間（預設）`peak-iops`。
- `absolute-min-iops` 為 IOPS 的絕對最小值。您可以將此欄位用於非常小的儲存物件。它會同時取代兩者 `peak-iops` 和/或 `expected-iops` 何時 `absolute-min-iops` 大於計算值 `expected-iops`。

例如、如果您設定 `expected-iops` 以 1、000 IOPS / TB 為單位、且磁碟區大小小於 1 GB `expected-iops` 將為分數 IOP。計算所得的 `peak-iops` 將會是更小的一部分。您可以透過設定來避免這種情況 `absolute-min-iops` 至實際值。

- `block-size` 指定應用程式 I/O 區塊大小。預設值為 32K。有效值為 8K、16K、32K、64K、any。任何表示不會強制執行區塊大小。

### 預設自適應 QoS 策略群組

有三個預設的調適性 QoS 原則群組可供使用、如下表所示。您可以將這些原則群組直接套用至磁碟區。

預設原則群組	預期 IOPS / TB	IOPS / TB 尖峰	絕對最小 IOPS
extreme	6、144	12288/3	1000
performance	2、048	4、096	500
value	128/128	512	75

### 儲存物件策略群組分配限制

在某些情況下，如果儲存物件包含的物件或其子物件屬於某個策略群組，則無法將儲存物件指派給該策略群組。

從 ONTAP 9.18.1 開始，您可以使用嵌套的 QoS 策略，該策略允許將策略群組指派給包含物件（例如 SVM）及其子物件（例如磁碟區）。在多租用戶環境中，嵌套的 QoS 策略使管理員能夠將 SVM 上的 QoS 限制細分到 SVM 內的各個磁碟區和 qtree，從而能夠在運算環境中平衡儲存資源，同時還能優先處理關鍵任務工作負載。

以下物件對支援嵌套 QoS 策略：

- SVM 和 SVM 包含的 FlexVol 或 FlexGroup 磁碟區。
- FlexVol 或 FlexGroup 卷以及卷內的 qtree。

對於嵌套的 QoS 策略，採用最嚴格的適用策略。

下表列出了各項限制。

如果您指派...	那麼你就無法將其指派到策略群組...
SVM到原則群組	SVM 中包含的任何儲存物件。   如果您執行的是ONTAP 9.18.1，則 SVM 中所包含的FlexVol和FlexGroup磁碟區_可以_指派給策略群組。
磁碟區至原則群組	包含該磁碟區或任何子 LUN 的 SVM。   如果您執行的是ONTAP 9.18.1 及更高版本，則包含磁碟區的 SVM _可以_指派給策略群組。此外，也可以分配FlexVol或FlexGroup磁碟區中的 qtree 。
LUN至原則群組	包含 LUN 的磁碟區或 SVM
檔案至原則群組	包含文件的磁碟區或 SVM

## 步驟

### 1. 建立可調適的QoS原則群組：

```
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
```

如"[指令參考資料ONTAP](#)"需詳細 `qos adaptive-policy-group create` 資訊，請參閱。



-expected-iops-allocation 和 -block-size 可在 ONTAP 9.5 或更新版本中取得。上述選項不受ONTAP 支援於支援的版本不包括在內。

下列命令會建立調適性 QoS 原則群組 adpg-app1 與 -expected-iops 設為 300 IOPS / TB 、 -peak -iops 設為 1、000 IOPS / TB 、 -peak-iops-allocation 設定為 used-space 和 -absolute-min-iops 設為 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. 將調適性QoS原則群組套用至磁碟區：

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

如"[指令參考資料ONTAP](#)"需詳細 `volume create` 資訊，請參閱。

下列命令會套用調適性 QoS 原則群組 adpg-app1 至 Volume app1：

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

下列命令會套用預設的調適性 QoS 原則群組 extreme 到新的 Volume app4 和現有的 Volume app5。為原則群組定義的處理量上限會套用至磁碟區 app4 和 app5 個別：

```
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 中設定調適性原則群組範本

從 ONTAP 9.13.1 開始，您可以使用調適性原則群組範本，在 SVM 層級強制執行處理量樓層和天花板。

關於這項工作

- 調適性原則群組範本是預設原則 apg1。您可以隨時修改原則。只能使用 CLI 或 ONTAP REST API 設定、而且只能套用至現有的 SVM。
- 調適性原則群組範本只會影響在設定原則之後、在 SVM 上建立或移轉到 SVM 的磁碟區。SVM 上的現有磁碟區會保留其現有狀態。

如果停用調適性原則群組範本、SVM 上的磁碟區會保留其現有原則。只有後續在 SVM 上建立或移轉到 SVM 的磁碟區、才會受到停用影響。

- 您無法在具有 QoS 原則群組的 SVM 上設定調適性原則群組範本。
- 調適性原則群組範本是專為 AFF 平台所設計。可在其他平台上設定調適性原則群組範本、但原則可能無法強制執行最低處理量。同樣地、您也可以將調適性原則群組範本新增至 FabricPool Aggregate 中的 SVM、或是新增至不支援最低處理量的 Aggregate、但不會強制執行處理量層。
- 如果 SVM 是在 MetroCluster 組態或 SnapMirror 關係中、則會在鏡射 SVM 上強制執行調適性原則群組範本。

步驟

1. 修改 SVM 以套用調適性原則群組範本：

```
vserver modify -qos-adaptive-policy-group
-template apg1
```

2. 確認已設定原則：`vserver show -fields qos-adaptive-policy-group`

## 版權資訊

Copyright © 2026 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變更本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

## 商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。