



Gerenciar workloads

ONTAP 9

NetApp
January 17, 2025

Índice

- Gerenciar workloads 1
 - Identificar a capacidade de performance restante 1
 - Identificar clientes ou arquivos de alto tráfego 2
 - Garantir taxa de transferência com QoS 4

Gerenciar workloads

Identificar a capacidade de performance restante

A capacidade de desempenho, ou *headroom*, mede quanto trabalho você pode colocar em um nó ou agregado antes que o desempenho das cargas de trabalho no recurso comece a ser afetado pela latência. Conhecer a capacidade de performance disponível no cluster ajuda você a provisionar e equilibrar workloads.

O que você vai precisar

São necessários comandos avançados de nível de privilégio para esta tarefa.

Sobre esta tarefa

Você pode usar os seguintes valores para a `-object` opção de coletar e exibir estatísticas de headroom:

- Para CPUs, `resource_headroom_cpu`.
- Para agregados `resource_headroom_aggr`, .

Você também pode concluir esta tarefa usando o Gerenciador de sistema e o Active IQ Unified Manager.

Passos

1. Alterar para nível de privilégio avançado:

```
set -privilege advanced
```

2. Iniciar a coleção de estatísticas de headroom em tempo real:

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

Para obter a sintaxe completa do comando, consulte a página `man`.

3. Apresentar informações estatísticas em tempo real do espaço livre:

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

Para obter a sintaxe completa do comando, consulte a página `man`.

4. Retornar ao privilégio administrativo:

```
set -privilege admin
```

Exemplo

O exemplo a seguir exibe as estatísticas médias horárias do espaço livre para nós de cluster.

Você pode calcular a capacidade de desempenho disponível para um nó subtraindo o `current_utilization` contador do `optimal_point_utilization` contador. Neste exemplo, a capacidade de utilização para `CPU_sti2520-213` é de -14% (72%-86%), o que sugere que a CPU foi superutilizada em média na última hora.

Pode ter especificado `ewma_daily`, `ewma_weekly` ou `ewma_monthly` obter a mesma média das

informações durante períodos de tempo mais longos.

```
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.

Identificar clientes ou arquivos de alto tráfego

Você pode usar a tecnologia ONTAP active Objects para identificar clientes ou arquivos responsáveis por uma quantidade desproporcionalmente grande de tráfego de cluster.

Depois de identificar esses "principais" clientes ou arquivos, você pode reequilibrar as cargas de trabalho do cluster ou tomar outras medidas para resolver o problema.

O que você vai precisar

Você deve ser um administrador de cluster para executar esta tarefa.

Passos

1. Veja os principais clientes que acessam o cluster:

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

Para obter a sintaxe completa do comando, consulte a página man.

O comando a seguir exibe os principais clientes acessando cluster1:

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

2. Veja os principais arquivos acessados no cluster:

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

Para obter a sintaxe completa do comando, consulte a página man.

O comando a seguir exibe os principais arquivos acessados no 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
```

Garantir taxa de transferência com QoS

Garanta a taxa de transferência com uma visão geral de QoS

Use a qualidade do serviço (QoS) de storage para garantir que a performance de workloads essenciais não seja degradada pelos workloads da concorrência. Você pode definir um throughput *ceiling* em uma carga de trabalho concorrente para limitar seu impacto nos recursos do sistema ou definir um throughput *floor* para uma carga de trabalho crítica, garantindo que ele atenda aos objetivos mínimos de taxa de transferência, independentemente da demanda por cargas de trabalho concorrentes. Você pode até mesmo definir um teto e piso para a mesma carga de trabalho.

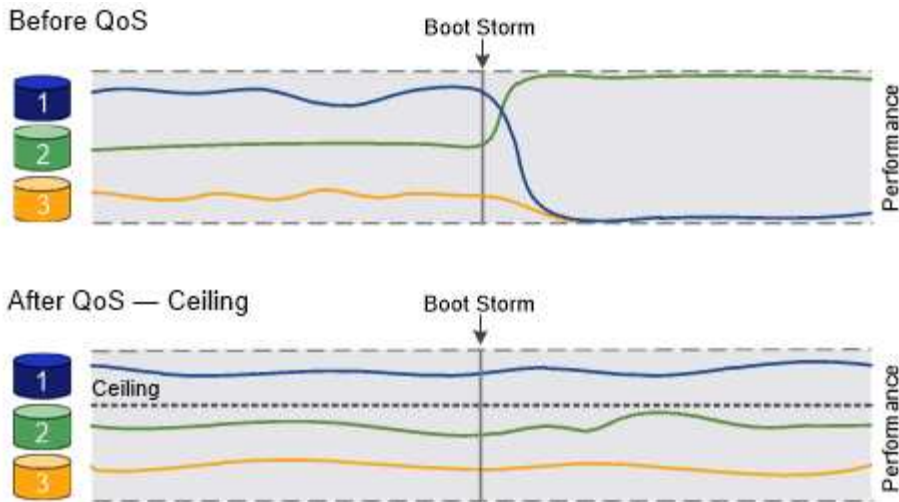
Sobre limites máximos de taxa de transferência (QoS Max)

Um limite máximo de taxa de transferência limita a taxa de transferência de um workload a um número máximo de IOPS ou Mbps, ou IOPS e Mbps. Na figura abaixo, o limite de taxa de transferência para a carga de trabalho 2 garante que não "bully" as cargas de trabalho 1 e 3.

Um *grupo de políticas* define o limite máximo de taxa de transferência para uma ou mais cargas de trabalho. Um workload representa as operações de e/S de um *objeto de storage*: um volume, arquivo, qtree ou LUN, ou todos os volumes, arquivos, qtrees ou LUNs em um SVM. Você pode especificar o limite máximo ao criar o grupo de políticas ou esperar até que você monitore cargas de trabalho para especificá-lo.



A taxa de transferência para workloads pode exceder o limite máximo especificado em até 10%, especialmente se um workload sofrer mudanças rápidas na taxa de transferência. O teto pode ser excedido em até 50% para lidar com explosões. As explosões ocorrem em nós únicos quando os tokens acumulam até 150%



Sobre os andares de taxa de transferência (QoS min)

Um piso de taxa de transferência garante que a taxa de transferência para um workload não fique abaixo de um número mínimo de IOPS ou Mbps, ou IOPS e Mbps. Na figura abaixo, os andares de taxa de transferência para o workload 1 e o workload 3 garantem que eles atendam aos destinos mínimos de taxa de transferência, independentemente da demanda por workload 2.



Como os exemplos sugerem, um teto de throughput limita a taxa de transferência diretamente. Um piso de taxa de transferência mantém a taxa de transferência indiretamente, dando prioridade às cargas de trabalho para as quais o piso foi definido.

Você pode especificar o piso ao criar o grupo de políticas ou esperar até que você monitore cargas de trabalho para especificá-lo.

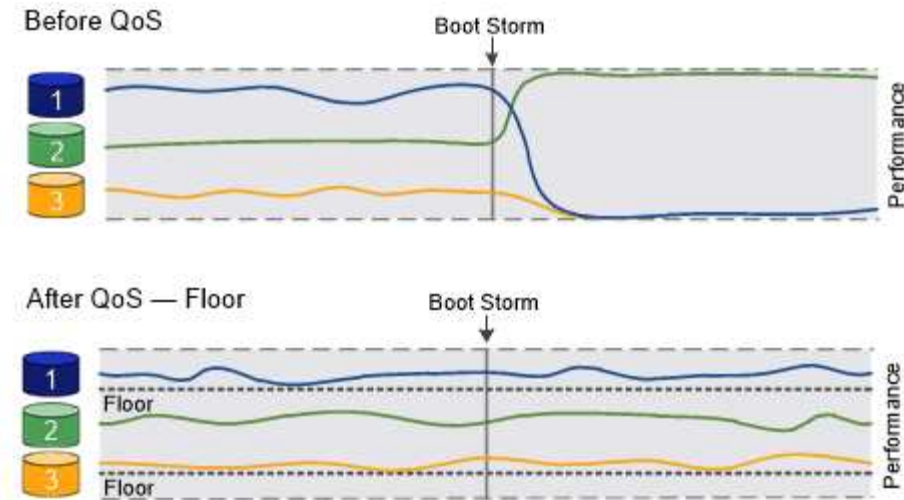
A partir do ONTAP 9.13.1, é possível definir os andares de taxa de transferência no escopo da SVM com [\[adaptive-qos-templates\]](#). Nas versões do ONTAP anteriores a 9.13.1, um grupo de políticas que define uma área de taxa de transferência não pode ser aplicado a um SVM.



Nos lançamentos anteriores ao ONTAP 9.7, os pisos de taxa de transferência são garantidos quando há capacidade de desempenho suficiente disponível.

No ONTAP 9.7 e posterior, os andares de throughput podem ser garantidos mesmo quando há capacidade de desempenho insuficiente disponível. Este novo comportamento do piso é chamado de pisos v2. Para atender às garantias, o piso v2 pode resultar em maior latência em cargas de trabalho sem uma taxa de transferência ou no trabalho que exceda as configurações básicas. Os pisos v2 aplicam-se a QoS e QoS adaptável.

A opção de ativar/desativar o novo comportamento dos pisos v2 está disponível no ONTAP 9.7P6 e posterior. Uma carga de trabalho pode ficar abaixo do nível especificado durante operações críticas como `volume move trigger-cutover`. Mesmo quando a capacidade suficiente está disponível e as operações críticas não estão ocorrendo, a taxa de transferência para uma carga de trabalho pode ficar abaixo do piso especificado em até 5%. Se os andares forem superprovisionados e não houver capacidade de performance, alguns workloads podem ficar abaixo do andar especificado.



Sobre grupos de políticas de QoS compartilhados e não compartilhados

A partir do ONTAP 9.4, você pode usar um grupo de políticas de QoS *não compartilhado* para especificar que o limite ou o piso da taxa de transferência definido se aplica a cada workload de membro individualmente. O comportamento dos grupos de políticas *shared* depende do tipo de política:

- Para limites máximos de taxa de transferência, a taxa de transferência total para as cargas de trabalho atribuídas ao grupo de políticas partilhadas não pode exceder o limite máximo especificado.
- Para andares de taxa de transferência, o grupo de políticas compartilhadas pode ser aplicado somente a um único workload.

Sobre a QoS adaptável

Normalmente, o valor do grupo de políticas que você atribui a um objeto de storage é fixo. Você precisa alterar o valor manualmente quando o tamanho do objeto de armazenamento muda. Um aumento na quantidade de espaço usado em um volume, por exemplo, geralmente requer um aumento correspondente no limite de produtividade especificado para o volume.

O *Adaptive QoS* dimensiona automaticamente o valor do grupo de políticas para o tamanho do workload, mantendo a taxa de IOPS para TBs|GBs conforme o tamanho do workload muda. Essa é uma vantagem significativa quando você gerencia centenas ou milhares de workloads em uma implantação grande.

Normalmente, você usa QoS adaptável para ajustar limites máximos de taxa de transferência, mas também pode usá-la para gerenciar andares de taxa de transferência (quando o tamanho do workload aumenta). O tamanho do workload é expresso como o espaço alocado para o objeto de storage ou o espaço usado pelo objeto de storage.



O espaço usado está disponível para pisos de throughput no ONTAP 9.5 e posterior. Não é suportado para pisos de rendimento no ONTAP 9.4 e anteriores.

- Uma política *allocated space* mantém a relação IOPS/TB|GB de acordo com o tamanho nominal do objeto de armazenamento. Se a taxa for de 100 IOPS/GB, um volume de 150 GB terá um limite máximo de taxa de transferência de 15.000 IOPS enquanto o volume permanecer nesse tamanho. Se o volume for redimensionado para 300 GB, a QoS adaptável ajusta o limite da taxa de transferência para 30.000 IOPS.
- Uma política *used space* (o padrão) mantém a taxa IOPS/TB|GB de acordo com a quantidade de dados reais armazenados antes da eficiência de armazenamento. Se a taxa for de 100 IOPS/GB, um volume de 150 GB que tenha 100 GB de dados armazenados teria um limite máximo de taxa de transferência de 10.000 IOPS. À medida que a quantidade de espaço usado muda, a QoS adaptável ajusta o teto de taxa

de transferência de acordo com a taxa.

A partir do ONTAP 9.5, você pode especificar um tamanho de bloco de e/S para o aplicativo que permite que um limite de taxa de transferência seja expresso em IOPS e Mbps. O limite de Mbps é calculado a partir do tamanho do bloco multiplicado pelo limite de IOPS. Por exemplo, um tamanho de bloco de e/S de 32K MB para um limite de IOPS de 6144IOPS GB/TB produz um limite de Mbps de 192MBps GB.

Você pode esperar o seguinte comportamento para tetos e pisos de rendimento:

- Quando um workload é atribuído a um grupo de políticas de QoS adaptável, o teto ou o piso é atualizado imediatamente.
- Quando um workload em um grupo de políticas de QoS adaptável é redimensionado, o teto ou o piso é atualizado em aproximadamente cinco minutos.

A taxa de transferência deve aumentar em pelo menos 10 IOPS antes que as atualizações ocorram.

Grupos de políticas de QoS adaptáveis sempre não são compartilhados: O limite ou o piso da taxa de transferência definida se aplica a cada workload de membro individualmente.

A partir do ONTAP 9.6, os andares de taxa de transferência são suportados no ONTAP Select premium com SSD.

Modelo de grupo de políticas adaptável

A partir do ONTAP 9.13,1, você pode definir um modelo de QoS adaptável em um SVM. Os modelos de grupo de políticas adaptáveis permitem definir andares e tetos de taxa de transferência para todos os volumes em uma SVM.

Os modelos de grupo de políticas adaptáveis só podem ser definidos após a criação do SVM. Use o `vserver modify` comando com o `-qos-adaptive-policy-group-template` parâmetro para definir a política.

Quando você define um modelo de grupo de políticas adaptativas, os volumes criados ou migrados após a configuração da diretiva herdam automaticamente a política. Quaisquer volumes existentes no SVM não serão afetados quando você atribuir o modelo de política. Se você desativar a política no SVM, qualquer volume posteriormente migrado ou criado no SVM não receberá a política. A desativação do modelo de grupo de políticas adaptativas não afeta os volumes que herdaram o modelo de política à medida que retêm o modelo de política.

Para obter mais informações, [Defina um modelo de grupo de políticas adaptável](#) consulte .

Suporte geral

A tabela a seguir mostra as diferenças no suporte para limites máximos de taxa de transferência, andares de taxa de transferência e QoS adaptável.

Recurso ou recurso	Teto com taxa de transferência	Piso de taxa de transferência	Piso de taxa de transferência v2	QoS adaptável
Versão ONTAP 9	Tudo	9,2 e mais tarde	9,7 e mais tarde	9,3 e mais tarde

Recurso ou recurso	Teto com taxa de transferência	Piso de taxa de transferência	Piso de taxa de transferência v2	QoS adaptável
Plataformas	Tudo	<ul style="list-style-type: none"> AFF C190 * ONTAP Select premium com SSD * 	<ul style="list-style-type: none"> AFF C190 ONTAP Select premium com SSD 	Tudo
Protocolos	Tudo	Tudo	Tudo	Tudo
FabricPool	Sim	Sim, se a política de disposição em categorias estiver definida como "nenhum" e não houver blocos na nuvem.	Sim, se a política de disposição em categorias estiver definida como "nenhum" e não houver blocos na nuvem.	Não
SnapMirror síncrono	Sim	Não	Não	Sim

O suporte ao C190 e ao ONTAP Select começou com o lançamento do ONTAP 9.6.

Workloads compatíveis com limites máximos de taxa de transferência

A tabela a seguir mostra o suporte do workload para limites máximos de taxa de transferência pela versão do ONTAP 9. Volumes raiz, espelhos de compartilhamento de carga e espelhos de proteção de dados não são compatíveis.

Suporte à carga de trabalho - limite máximo	ONTAP 9,0	ONTAP 9,1	ONTAP 9,2	ONTAP 9,3	ONTAP 9.4 - 9,7	ONTAP 9 F.8 e mais tarde
Volume	sim	sim	sim	sim	sim	sim
Ficheiro	sim	sim	sim	sim	sim	sim
LUN	sim	sim	sim	sim	sim	sim
SVM	sim	sim	sim	sim	sim	sim
Volume FlexGroup	não	não	não	sim	sim	sim
qtrees*	não	não	não	não	não	sim

Suporte à carga de trabalho - limite máximo	ONTAP 9,0	ONTAP 9,1	ONTAP 9,2	ONTAP 9,3	ONTAP 9.4 - 9,7	ONTAP 9 F.8 e mais tarde
Vários workloads por grupo de políticas	sim	sim	sim	sim	sim	sim
Grupos de políticas não compartilhados	não	não	não	não	sim	sim

A partir do ONTAP 9.8, o acesso NFS é compatível com qtrees nos volumes FlexVol e FlexGroup com NFS habilitado. A partir do ONTAP 9.9,1, o acesso SMB também é suportado em qtrees nos volumes FlexVol e FlexGroup com SMB ativado.

Workloads compatíveis em pisos de taxa de transferência

A tabela a seguir mostra o suporte do workload para andares de taxa de transferência pela versão do ONTAP 9. Volumes raiz, espelhos de compartilhamento de carga e espelhos de proteção de dados não são compatíveis.

Suporte de carga de trabalho - básico	ONTAP 9,2	ONTAP 9,3	ONTAP 9.4 - 9,7	ONTAP 9.8 - 9.13.0	ONTAP 9.13,1 e posterior
Volume	sim	sim	sim	sim	sim
Ficheiro	não	sim	sim	sim	sim
LUN	sim	sim	sim	sim	sim
SVM	não	não	não	não	sim
Volume FlexGroup	não	não	sim	sim	sim
qtrees *	não	não	não	sim	sim
Vários workloads por grupo de políticas	não	não	sim	sim	sim
Grupos de políticas não compartilhados	não	não	sim	sim	sim

A partir do ONTAP 9.8, o acesso NFS é suportado em qtrees nos volumes FlexVol e FlexGroup com NFS ativado. A partir do ONTAP 9.9,1, o acesso SMB também é suportado em qtrees nos volumes FlexVol e FlexGroup com SMB ativado.

Workloads compatíveis com QoS adaptável

A tabela a seguir mostra o suporte do workload para QoS adaptável pela versão do ONTAP 9. Volumes raiz, espelhos de compartilhamento de carga e espelhos de proteção de dados não são compatíveis.

Suporte a workload - QoS adaptável	ONTAP 9,3	ONTAP 9.4 - 9.13.0	ONTAP 9.13,1 e posterior
Volume	sim	sim	sim
Ficheiro	não	sim	sim
LUN	não	sim	sim
SVM	não	não	sim
Volume FlexGroup	não	sim	sim
Vários workloads por grupo de políticas	sim	sim	sim
Grupos de políticas não compartilhados	sim	sim	sim

Número máximo de cargas de trabalho e grupos de políticas

A tabela a seguir mostra o número máximo de cargas de trabalho e grupos de políticas por versão do ONTAP 9.

Suporte a workload	ONTAP 9 .3 e anteriores	ONTAP 9 .4 e mais tarde
Máximo de workloads por cluster	12.000	40.000
Máximo de workloads por nó	12.000	40.000
Máximo de grupos de políticas	12.000	12.000

Ativar ou desativar os pisos de rendimento v2

Você pode ativar ou desativar os andares de taxa de transferência v2 no AFF. A predefinição é Enabled (ativado). Com os andares v2 ativados, os andares de taxa de transferência podem ser atendidos quando os controladores são muito usados em detrimento da latência mais alta em outros workloads. Os pisos v2 aplicam-se a QoS e QoS adaptável.

Passos

1. Alterar para nível de privilégio avançado:

```
set -privilege advanced
```

2. Introduza um dos seguintes comandos:

Se você quiser...	Use este comando:
Desativar pisos v2	<code>qos settings throughput-floors-v2 -enable false</code>
Ativar os pisos v2	<code>qos settings throughput-floors-v2 -enable true</code>



Para desativar os pisos de taxa de transferência v2 num cluster MetroCluster, tem de executar o.

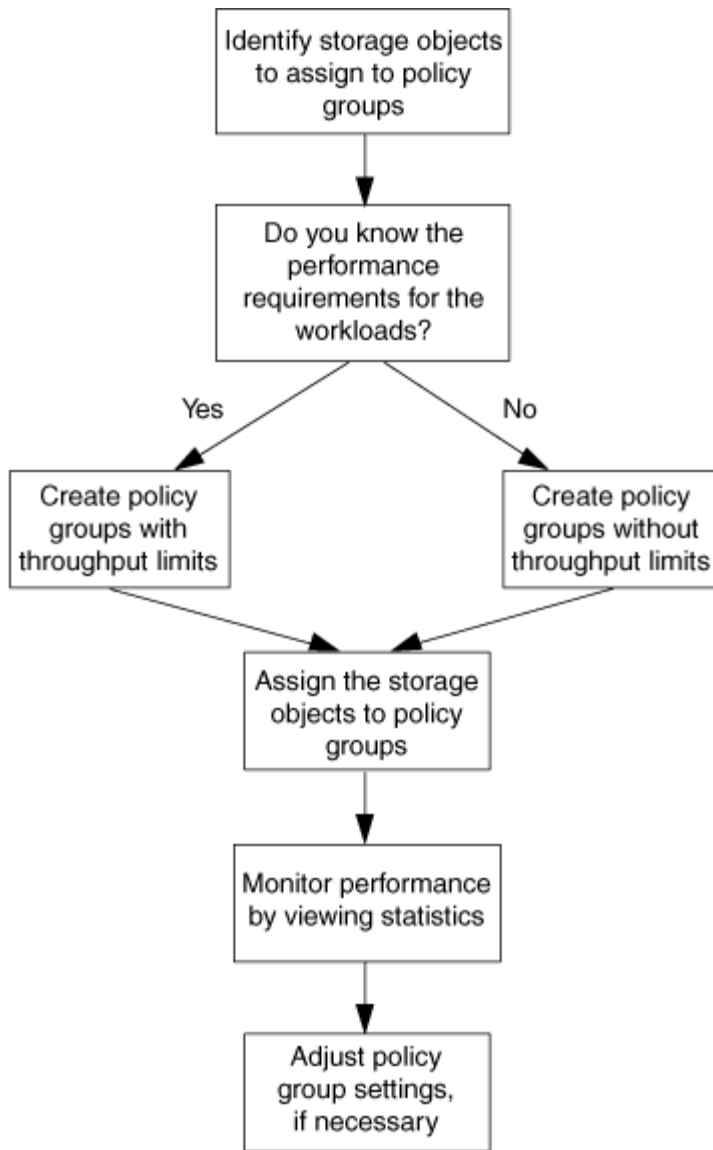
```
qos settings throughput-floors-v2 -enable false
```

comando nos clusters de origem e destino.

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

Fluxo de trabalho de QoS do storage

Se você já conhece os requisitos de desempenho para os workloads que deseja gerenciar com QoS, poderá especificar o limite de taxa de transferência ao criar o grupo de políticas. Caso contrário, você pode esperar até que você monitore as cargas de trabalho para especificar o limite.



Defina um limite de taxa de transferência com QoS

Você pode usar o `max-throughput` campo de um grupo de políticas para definir um limite máximo de taxa de transferência para workloads de objetos de storage (QoS Max). Você pode aplicar o grupo de políticas ao criar ou modificar o objeto de armazenamento.

O que você vai precisar

- Você deve ser um administrador de cluster para criar um grupo de políticas.
- Você deve ser um administrador de cluster para aplicar um grupo de políticas a um SVM.

Sobre esta tarefa

- A partir do ONTAP 9.4, você pode usar um grupo de políticas de QoS *não compartilhado* para especificar que o limite de taxa de transferência definido se aplica a cada workload de membro individualmente. Caso contrário, o grupo de políticas é *compartilhado*: a taxa de transferência total para as cargas de trabalho atribuídas ao grupo de políticas não pode exceder o limite máximo especificado.

Defina `-is-shared=false` para que o `qos policy-group create` comando especifique um grupo de políticas não compartilhado.

- Você pode especificar o limite de taxa de transferência para o limite máximo em IOPS, MB/s ou IOPS, MB/s. Se você especificar IOPS e MB/s, qualquer limite atingido primeiro será aplicado.



Se você definir um teto e um piso para a mesma carga de trabalho, poderá especificar o limite de taxa de transferência para o limite máximo apenas em IOPS.

- Um objeto de storage que esteja sujeito a um limite de QoS precisa estar contido pelo SVM a que o grupo de políticas pertence. Vários grupos de políticas podem pertencer ao mesmo SVM.
- Não é possível atribuir um objeto de armazenamento a um grupo de políticas se o objeto que contém ou os objetos filho pertencerem ao grupo de políticas.
- É uma prática recomendada de QoS aplicar um grupo de políticas ao mesmo tipo de objetos de storage.

Passos

1. Criar um grupo de políticas:

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

Para obter a sintaxe completa do comando, consulte a página man. Você pode usar o `qos policy-group modify` comando para ajustar os tetos de taxa de transferência.

O comando a seguir cria o grupo de políticas compartilhadas `pg-vs1` com uma taxa de transferência máxima de 5.000 IOPS:

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

O comando a seguir cria o grupo de políticas não compartilhadas `pg-vs3` com uma taxa de transferência máxima de 100 IOPS e 400 Kb/S:

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

O comando a seguir cria o grupo de políticas não compartilhadas `pg-vs4` sem um limite de taxa de transferência:

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

2. Aplique um grupo de políticas a um SVM, arquivo, volume ou LUN:

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

Para obter a sintaxe completa do comando, consulte as páginas man. Você pode usar o `storage_object modify` comando para aplicar um grupo de políticas diferente ao objeto de armazenamento.

O comando a seguir aplica o grupo de políticas `pg-vs1` ao SVM `vs1`:

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

Os comandos a seguir aplicam o grupo de políticas `pg-app` aos volumes `app1` e `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. Monitorar o desempenho do grupo de políticas:

```
qos statistics performance show
```

Para obter a sintaxe completa do comando, consulte a página `man`.



Monitore o desempenho do cluster. Não use uma ferramenta no host para monitorar o desempenho.

O comando a seguir mostra o desempenho do grupo de políticas:

```
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. Monitorar a performance do workload:

```
qos statistics workload performance show
```

Para obter a sintaxe completa do comando, consulte a página `man`.



Monitore o desempenho do cluster. Não use uma ferramenta no host para monitorar o desempenho.

O comando a seguir mostra o desempenho da carga de trabalho:


```
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
```



Use o `qos statistics workload latency show` comando para visualizar estatísticas detalhadas de latência para workloads de QoS.

Defina um piso de taxa de transferência com QoS

Você pode usar o `min-throughput` campo de um grupo de políticas para definir um piso de taxa de transferência para workloads de objetos de storage (QoS min). Você pode aplicar o grupo de políticas ao criar ou modificar o objeto de armazenamento. A partir do ONTAP 9.8, você pode especificar o piso da taxa de transferência em IOPS ou Mbps, ou IOPS e Mbps.

Antes de começar

- Você deve estar executando o ONTAP 9.2 ou posterior. Os pisos de taxa de transferência estão disponíveis a partir do ONTAP 9.2.
- Você deve ser um administrador de cluster para criar um grupo de políticas.
- A partir do ONTAP 9.13,1, você pode aplicar pisos de taxa de transferência no nível SVM usando um [modelo de grupo de políticas adaptável](#). Não é possível definir um modelo de grupo de políticas adaptável em um SVM com um grupo de políticas de QoS.

Sobre esta tarefa

- A partir do ONTAP 9.4, você pode usar um grupo de políticas de QoS *não compartilhado* para especificar que o piso da taxa de transferência definido seja aplicado individualmente a cada workload de membro. Essa é a única condição em que um grupo de políticas para uma área de transferência pode ser aplicado a várias cargas de trabalho.

Defina `-is-shared=false` para que o `qos policy-group create` comando especifique um grupo de políticas não compartilhado.

- A taxa de transferência para uma carga de trabalho pode ficar abaixo do nível especificado se houver capacidade de desempenho (espaço livre) insuficiente no nó ou no agregado.
- Um objeto de storage que esteja sujeito a um limite de QoS precisa estar contido pelo SVM a que o grupo de políticas pertence. Vários grupos de políticas podem pertencer ao mesmo SVM.
- É uma prática recomendada de QoS aplicar um grupo de políticas ao mesmo tipo de objetos de storage.
- Um grupo de políticas que define um piso de taxa de transferência não pode ser aplicado a um SVM.

Passos

1. Verifique se há capacidade de desempenho adequada no nó ou no agregado, conforme descrito ["Identificação da capacidade de performance restante"](#) em .
2. Criar um grupo de políticas:

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

Para obter a sintaxe de comando completa, consulte a página man para sua versão do ONTAP. Você pode usar o `qos policy-group modify` comando para ajustar os andares de taxa de transferência.

O comando a seguir cria o grupo de políticas compartilhadas `pg-vs2` com uma taxa de transferência mínima de 1.000 IOPS:

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

O comando a seguir cria o grupo de políticas não compartilhadas `pg-vs4` sem um limite de taxa de transferência:

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

3. Aplicar um grupo de políticas a um volume ou LUN:

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

Para obter a sintaxe completa do comando, consulte as páginas man. Você pode usar o `_storage_object_modify` comando para aplicar um grupo de políticas diferente ao objeto de armazenamento.

O comando a seguir aplica o grupo de políticas `pg-app2` ao volume `app2`:

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

4. Monitorar o desempenho do grupo de políticas:

```
qos statistics performance show
```

Para obter a sintaxe completa do comando, consulte a página man.



Monitore o desempenho do cluster. Não use uma ferramenta no host para monitorar o desempenho.

O comando a seguir mostra o desempenho do grupo de políticas:

```
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
```

5. Monitorar a performance do workload:

```
qos statistics workload performance show
```

Para obter a sintaxe completa do comando, consulte a página `man`.



Monitore o desempenho do cluster. Não use uma ferramenta no host para monitorar o desempenho.

O comando a seguir mostra o desempenho da carga de trabalho:

```
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
```



Use o `qos statistics workload latency show` comando para visualizar estatísticas detalhadas de latência para workloads de QoS.

Use grupos de políticas de QoS adaptáveis

Você pode usar um grupo de políticas *Adaptive QoS* para escalar automaticamente um limite de taxa de transferência ou um tamanho de chão para volume, mantendo a taxa de IOPS para TBs|GBs conforme o tamanho do volume muda. Essa é uma vantagem significativa quando você gerencia centenas ou milhares de workloads em uma implantação grande.

Antes de começar

- Você deve estar executando o ONTAP 9.3 ou posterior. Os grupos de políticas de QoS adaptáveis estão disponíveis a partir do ONTAP 9.3.
- Você deve ser um administrador de cluster para criar um grupo de políticas.

Sobre esta tarefa

Um objeto de storage pode ser membro de um grupo de políticas adaptáveis ou de um grupo de políticas não adaptáveis, mas não ambos. O SVM do objeto de storage e a política devem ser os mesmos. O objeto de storage deve estar on-line.

Grupos de políticas de QoS adaptáveis sempre não são compartilhados: O limite ou o piso da taxa de transferência definida se aplica a cada workload de membro individualmente.

A proporção de limites de taxa de transferência para o tamanho do objeto de armazenamento é determinada pela interação dos seguintes campos:

- `expected-iops` É o mínimo esperado de IOPS por TB|GB alocado.



`expected-iops` É garantido apenas nas plataformas AFF.
`expected-iops` Será garantido para o FabricPool somente se a política de disposição em categorias estiver definida como "nenhuma" e não houver blocos na nuvem. `expected-iops` É garantido para volumes que não estão em uma relação síncrona SnapMirror.

- `peak-iops` É o máximo de IOPS possível por TB|GB alocado ou usado.
- `expected-iops-allocation` especifica se o espaço alocado (o padrão) ou o espaço usado é usado para iops-esperado.



`expected-iops-allocation` Está disponível no ONTAP 9.5 e posterior. Ele não é suportado no ONTAP 9.4 e anterior.

- `peak-iops-allocation` especifica se o espaço alocado ou o espaço usado (o padrão) é usado para `peak-iops`.
- `absolute-min-iops` É o número mínimo absoluto de IOPS. Você pode usar este campo com objetos de armazenamento muito pequenos. Substitui ambos `peak-iops` e/ `expected-iops` ou quando `absolute-min-iops` é maior do que o `expected-iops` calculado .

Por exemplo, se você definir `expected-iops` como 1.000 IOPS/TB e o tamanho do volume for inferior a 1 GB, o calculado `expected-iops` será uma IOP fracionária. O calculado `peak-iops` será uma fração ainda menor. Você pode evitar isso definindo `absolute-min-iops` um valor realista.

- `block-size` Especifica o tamanho do bloco de e/S da aplicação. A predefinição é 32K. Os valores válidos são 8K, 16K, 32K, 64K, QUALQUER. QUALQUER significa que o tamanho do bloco não é imposto.

Três grupos de políticas de QoS adaptáveis padrão estão disponíveis, como mostrado na tabela a seguir. Você pode aplicar esses grupos de políticas diretamente a um volume.

Grupo de políticas padrão	IOPS/TB esperados	IOPS/TB de pico	IOPS mín. Absoluto
extreme	6.144	12.288	1000

performance	2.048	4.096	500
value	128	512	75

Não é possível atribuir um objeto de armazenamento a um grupo de políticas se o objeto que contém ou os objetos filho pertencerem a um grupo de políticas. A tabela a seguir lista as restrições.

Se você atribuir...	Então você não pode atribuir...
SVM em um grupo de políticas	Quaisquer objetos de storage contidos pelo SVM em um grupo de políticas
Volume para um grupo de políticas	Volume contendo SVM ou LUNs filho, em um grupo de políticas
LUN para um grupo de políticas	LUN que contém volume ou SVM em um grupo de políticas
Arquivo para um grupo de políticas	Os arquivos contêm volume ou SVM em um grupo de políticas

Passos

1. Criar um grupo de políticas de QoS adaptável:

```
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
```

Para obter a sintaxe completa do comando, consulte a página man.



-expected-iops-allocation E -block-size está disponível em ONTAP 9.5 e posterior. Essas opções não são suportadas no ONTAP 9.4 e versões anteriores.

O comando a seguir cria um grupo de políticas de QoS adaptável `adpg-app1` -expected-iops definido como 300 IOPS/TB, -peak-iops definido como 1.000 IOPS/TB, -peak-iops-allocation definido como used-space e -absolute-min-iops definido como 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. Aplicar um grupo de políticas de QoS adaptável a um volume:

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

Para obter a sintaxe completa do comando, consulte as páginas man.

O comando a seguir aplica o grupo de políticas de QoS adaptável `adpg-app1` ao volume `app1`:

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

Os comandos a seguir aplicam o grupo de políticas de QoS adaptável padrão `extreme` ao novo volume `app4` e ao volume existente `app5`. O limite máximo de taxa de transferência definido para o grupo de políticas aplica-se a volumes `app4` e `app5` individualmente:

```
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
```

Defina um modelo de grupo de políticas adaptável

A partir do ONTAP 9.13,1, você pode aplicar pisos e tetos de taxa de transferência no nível SVM usando um modelo de grupo de políticas adaptável.

Sobre esta tarefa

- O modelo de grupo de políticas adaptativas é uma política `apg1` padrão. A política pode ser modificada a qualquer momento. Ela só pode ser definida com a API REST CLI ou ONTAP e só pode ser aplicada a SVMs existentes.
- O modelo de grupo de políticas adaptável afeta apenas os volumes criados ou migrados para o SVM após você definir a política. Os volumes existentes no SVM mantêm seu status atual.

Se você desabilitar o modelo de grupo de políticas adaptáveis, os volumes no SVM manterão suas políticas existentes. Somente os volumes posteriormente criados ou migrados para o SVM serão afetados pelo desfalecimento.

- Não é possível definir um modelo de grupo de políticas adaptável em um SVM com um grupo de políticas de QoS.
- Os modelos de grupo de políticas adaptáveis são projetados para plataformas AFF. Um modelo de grupo de políticas adaptável pode ser definido em outras plataformas, mas a política pode não impor uma taxa de transferência mínima. Da mesma forma, você pode adicionar um modelo de grupo de políticas adaptável a um SVM em um agregado do FabricPool ou em um agregado que não ofereça suporte a taxa de transferência mínima. No entanto, o nível de taxa de transferência não será imposto.
- Se o SVM estiver em uma configuração do MetroCluster ou em uma relação do SnapMirror, o modelo de grupo de políticas adaptável será aplicado no SVM espelhado.

Passos

1. Modifique o SVM para aplicar o modelo de grupo de políticas adaptável: `vserver modify -qos`

```
-adaptive-policy-group-template apg1
```

2. Confirme se a política foi definida: `vserver show -fields qos-adaptive-policy-group`

Informações sobre direitos autorais

Copyright © 2025 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

O software derivado do material da NetApp protegido por direitos autorais está sujeito à seguinte licença e isenção de responsabilidade:

ESTE SOFTWARE É FORNECIDO PELA NETAPP "NO PRESENTE ESTADO" E SEM QUAISQUER GARANTIAS EXPRESSAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO E ADEQUAÇÃO A UM DETERMINADO PROPÓSITO, CONFORME A ISENÇÃO DE RESPONSABILIDADE DESTES DOCUMENTOS. EM HIPÓTESE ALGUMA A NETAPP SERÁ RESPONSÁVEL POR QUALQUER DANO DIRETO, INDIRETO, INCIDENTAL, ESPECIAL, EXEMPLAR OU CONSEQUENCIAL (INCLUINDO, SEM LIMITAÇÕES, AQUISIÇÃO DE PRODUTOS OU SERVIÇOS SOBRESSALIENTES; PERDA DE USO, DADOS OU LUCROS; OU INTERRUPÇÃO DOS NEGÓCIOS), INDEPENDENTEMENTE DA CAUSA E DO PRINCÍPIO DE RESPONSABILIDADE, SEJA EM CONTRATO, POR RESPONSABILIDADE OBJETIVA OU PREJUÍZO (INCLUINDO NEGLIGÊNCIA OU DE OUTRO MODO), RESULTANTE DO USO DESTES SOFTWARES, MESMO SE ADVERTIDA DA RESPONSABILIDADE DE TAL DANO.

A NetApp reserva-se o direito de alterar quaisquer produtos descritos neste documento, a qualquer momento e sem aviso. A NetApp não assume nenhuma responsabilidade nem obrigação decorrentes do uso dos produtos descritos neste documento, exceto conforme expressamente acordado por escrito pela NetApp. O uso ou a compra deste produto não representam uma licença sob quaisquer direitos de patente, direitos de marca comercial ou quaisquer outros direitos de propriedade intelectual da NetApp.

O produto descrito neste manual pode estar protegido por uma ou mais patentes dos EUA, patentes estrangeiras ou pedidos pendentes.

LEGENDA DE DIREITOS LIMITADOS: o uso, a duplicação ou a divulgação pelo governo estão sujeitos a restrições conforme estabelecido no subparágrafo (b)(3) dos Direitos em Dados Técnicos - Itens Não Comerciais no DFARS 252.227-7013 (fevereiro de 2014) e no FAR 52.227- 19 (dezembro de 2007).

Os dados aqui contidos pertencem a um produto comercial e/ou serviço comercial (conforme definido no FAR 2.101) e são de propriedade da NetApp, Inc. Todos os dados técnicos e software de computador da NetApp fornecidos sob este Contrato são de natureza comercial e desenvolvidos exclusivamente com despesas privadas. O Governo dos EUA tem uma licença mundial limitada, irrevogável, não exclusiva, intransferível e não sublicenciável para usar os Dados que estão relacionados apenas com o suporte e para cumprir os contratos governamentais desse país que determinam o fornecimento de tais Dados. Salvo disposição em contrário no presente documento, não é permitido usar, divulgar, reproduzir, modificar, executar ou exibir os dados sem a aprovação prévia por escrito da NetApp, Inc. Os direitos de licença pertencentes ao governo dos Estados Unidos para o Departamento de Defesa estão limitados aos direitos identificados na cláusula 252.227-7015(b) (fevereiro de 2014) do DFARS.

Informações sobre marcas comerciais

NETAPP, o logotipo NETAPP e as marcas listadas em <http://www.netapp.com/TM> são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.