



Gerenciamento da performance com a capacidade de performance e as informações de IOPS disponíveis

Active IQ Unified Manager 9.8

NetApp
January 31, 2025

Índice

Gerenciamento da performance com a capacidade de performance e as informações de IOPS disponíveis . .	1
Qual é a capacidade de desempenho utilizada	1
O que significa o valor da capacidade de desempenho utilizada	2
O que é IOPS disponível	3
Exibindo valores de capacidade de desempenho agregado e nó usados	4
Exibindo o nó e agregar valores de IOPS disponíveis	5
Visualização de gráficos de contadores de capacidade de desempenho para identificar problemas	6
Condições de limite de performance usadas pela capacidade de performance	8
Usando o contador de capacidade de desempenho usado para gerenciar o desempenho	9

Gerenciamento da performance com a capacidade de performance e as informações de IOPS disponíveis

Capacidade de desempenho indica quanto throughput você pode obter de um recurso sem superar o desempenho útil desse recurso. Quando vista usando contadores de performance existentes, a capacidade de performance é o ponto em que você obtém a utilização máxima de um nó ou agregado antes que a latência se torne um problema.

O Unified Manager coleta estatísticas de capacidade de performance de nós e agregados em cada cluster. *Capacidade de desempenho usada* é a porcentagem de capacidade de desempenho que está sendo usada atualmente, e *capacidade de desempenho livre* é a porcentagem de capacidade de desempenho que ainda está disponível.

Embora a capacidade de desempenho livre forneça uma porcentagem do recurso que ainda está disponível, *IOPS disponível* informa o número de IOPS que podem ser adicionados ao recurso antes de atingir a capacidade máxima de desempenho. Usando essa métrica, você pode ter certeza de que pode adicionar workloads de um número predeterminado de IOPS a um recurso.

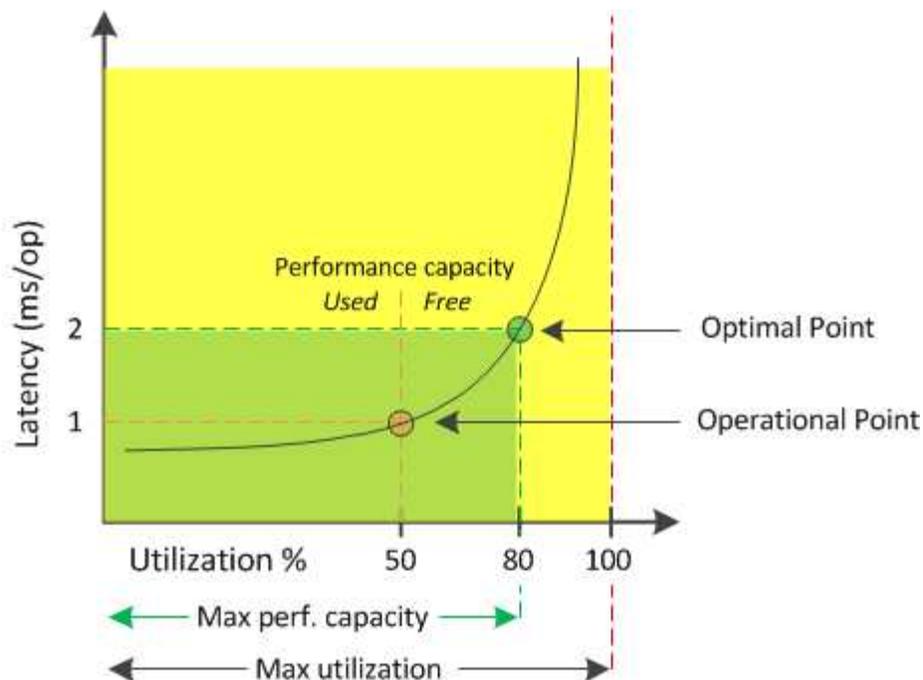
O monitoramento das informações de capacidade de desempenho tem os seguintes benefícios:

- Auxilia no provisionamento e balanceamento de fluxo de trabalho.
- Ajuda a evitar sobrecarregar um nó ou empurrar seus recursos para além do ponto ideal, reduzindo assim a necessidade de solucionar problemas.
- Ajuda você a determinar com maior precisão onde equipamentos de armazenamento adicionais podem ser necessários.

Qual é a capacidade de desempenho utilizada

O contador de capacidade de desempenho usado ajuda a identificar se o desempenho de um nó ou de um agregado está atingindo um ponto em que o desempenho pode degradar se as cargas de trabalho aumentarem. Ele também pode mostrar se um nó ou agregado está sendo usado em excesso durante períodos específicos de tempo. A capacidade de performance usada é semelhante à utilização, mas a primeira fornece mais informações sobre os recursos de performance disponíveis em um recurso físico para uma carga de trabalho específica.

A capacidade de performance usada ideal é o ponto em que um nó ou um agregado tem utilização e latência ideais (tempo de resposta) e está sendo usado com eficiência. Uma curva de latência versus utilização de amostra é mostrada para um agregado na figura a seguir.



Neste exemplo, o *ponto operacional* identifica que o agregado está operando atualmente em 50% de utilização com latência de 1,0 ms/op. Com base nas estatísticas capturadas do agregado, o Unified Manager determina que a capacidade de desempenho adicional está disponível para esse agregado. Neste exemplo, o *ponto ótimo* é identificado como o ponto quando o agregado está em 80% de utilização com latência de 2,0 ms/op. Portanto, é possível adicionar mais volumes e LUNs a esse agregado para que seus sistemas sejam usados com mais eficiência.

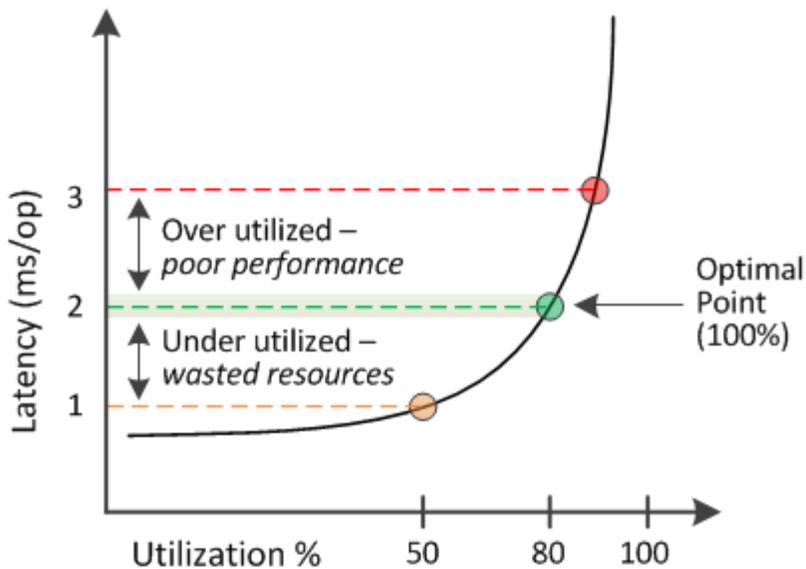
Espera-se que o contador de capacidade de desempenho usado seja um número maior do que o contador de "utilização" porque a capacidade de desempenho aumenta o impacto na latência. Por exemplo, se um nó ou agregado for 70% usado, o valor da capacidade de desempenho pode estar na faixa de 80% a 100%, dependendo do valor de latência.

Em alguns casos, no entanto, o contador de utilização pode ser superior na página Painel de instrumentos. Isso é normal porque o painel atualiza os valores atuais do contador em cada período de coleta; ele não exibe médias em um período de tempo, como as outras páginas na interface de usuário do Unified Manager. O contador de capacidade de desempenho usado é melhor usado como um indicador de desempenho médio ao longo de um período de tempo, enquanto o contador de utilização é melhor usado para determinar o uso instantâneo de um recurso.

O que significa o valor da capacidade de desempenho utilizada

O valor da capacidade de desempenho usado ajuda a identificar os nós e agregados que estão sendo sobreutilizados ou subutilizados atualmente. Isso permite redistribuir workloads para aumentar a eficiência dos recursos de storage.

A figura a seguir mostra a curva de latência versus utilização de um recurso e identifica, com pontos coloridos, três áreas onde o ponto operacional atual poderia estar localizado.



- Uma porcentagem de capacidade de desempenho usada igual a 100 está no ponto ideal.

Os recursos estão sendo usados de forma eficiente neste momento.

- Uma porcentagem de capacidade de desempenho usada acima de 100 indica que o nó ou o agregado está sobreutilizado e que as cargas de trabalho estão recebendo desempenho abaixo do ideal.

Nenhuma nova carga de trabalho deve ser adicionada ao recurso, e as cargas de trabalho existentes podem precisar ser redistribuídas.

- Uma porcentagem de capacidade de desempenho usada abaixo de 100 indica que o nó ou o agregado está subutilizado e que os recursos não estão sendo usados com eficiência.

Mais workloads podem ser adicionados ao recurso.



Diferentemente da utilização, a porcentagem de capacidade de performance usada pode estar acima de 100%. Não há porcentagem máxima, mas os recursos geralmente estarão na faixa de 110% a 140% quando estiverem sendo sobreutilizados. Percentagens mais elevadas indicariam um recurso com problemas graves.

O que é IOPS disponível

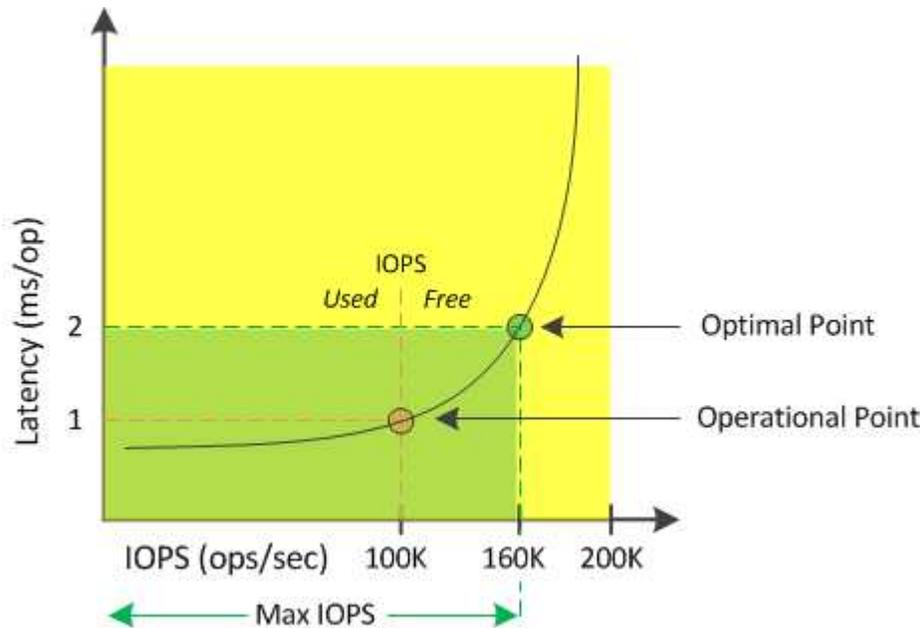
O contador de IOPS disponível identifica o número restante de IOPS que pode ser adicionado a um nó ou a um agregado antes que o recurso atinja seu limite. O total de IOPS que um nó pode fornecer é baseado nas características físicas do nó - por exemplo, o número de CPUs, a velocidade da CPU e a quantidade de RAM. O total de IOPS que um agregado pode fornecer é baseado nas propriedades físicas dos discos - por exemplo, um disco SATA, SAS ou SSD.

Embora o contador de capacidade de performance forneça a porcentagem de um recurso que ainda está disponível, o contador de IOPS disponível informa o número exato de IOPS (cargas de trabalho) pode ser adicionado a um recurso antes de atingir a capacidade máxima de performance.

Por exemplo, se você estiver usando um par de sistemas de storage FAS2520 e FAS8060, um valor sem

capacidade de performance de 30% significa que você tem alguma capacidade de performance gratuita. No entanto, esse valor não fornece visibilidade sobre o número de workloads que você pode implantar nesses nós. O contador de IOPS disponível pode mostrar que você tem 500 IOPS disponíveis no FAS8060, mas apenas 100 IOPS disponíveis no FAS2520.

Uma curva de latência de exemplo versus IOPS para um nó é mostrada na figura a seguir.



O número máximo de IOPS que um recurso pode fornecer é o número de IOPS quando o contador de capacidade de desempenho usado é de 100% (o ponto ideal). O ponto operacional identifica que o nó está operando atualmente em 100K IOPS com latência de 1,0 ms/op. Com base nas estatísticas capturadas a partir do nó, o Unified Manager determina que o máximo de IOPS para o nó é 160K, o que significa que há 60K IOPS livres ou disponíveis. Portanto, você pode adicionar mais workloads a esse nó para que seus sistemas sejam usados com mais eficiência.



Quando há uma atividade mínima do usuário no recurso, o valor de IOPS disponível é calculado assumindo uma carga de trabalho genérica com base em aproximadamente 4.500 IOPS por núcleo da CPU. Isso ocorre porque o Unified Manager não tem os dados para estimar com precisão as características da carga de trabalho que está sendo atendida.

Exibindo valores de capacidade de desempenho agregado e nó usados

Você pode monitorar os valores de capacidade de performance usados para todos os nós ou para todos os agregados em um cluster, ou visualizar detalhes de um único nó ou agregado.

Os valores de capacidade de desempenho usados são exibidos no Painel, nas páginas de inventário de desempenho, na página de principais artistas, na página criar política de limite, nas páginas do Performance Explorer e nos gráficos detalhados. Por exemplo, a página desempenho: Todos os agregados fornece uma coluna capacidade de desempenho usada para exibir o valor da capacidade de desempenho usada para todos os agregados.

Latency, IOPS, MBps, Utilization are based on hourly samples averaged over the previous 72 hours

Filtering		No filter applied		Search Aggregates Data		Search						
Assign Threshold Policy		Clear Threshold Policy										
Status	Aggregate	Latency	IOPS	MBps	Perf. Capacity Used ↓	Utilization	Free Capacity	Total Capacity	Cluster	Node	Policy	
<input type="checkbox"/>		opm_mo..._agg0	16.3 ms/op	124 IOPS	< 1 MBps	45%	9%	154 GB	3,179 GB	opm-mobility	opm-m...-02	
<input type="checkbox"/>		rt_aggr2	19.8 ms/op	290 IOPS	< 1 MBps	45%	15%	6,692 GB	6,693 GB	opm-mobility	opm-m...-02	
<input type="checkbox"/>		aggr_snap_mirror	13.9 ms/op	267 IOPS	< 1 MBps	38%	12%	6,692 GB	6,693 GB	opm-mobility	opm-m...-02	
<input type="checkbox"/>		sdot_aggr	17.3 ms/op	745 IOPS	< 1 MBps	24%	11%	26,621 GB	26,774 GB	opm-mobility	opm-m...-02	
<input type="checkbox"/>		aggr1	15.5 ms/op	434 IOPS	< 1 MBps	16%	6%	4,390 GB	20,080 GB	opm-mobility	opm-m...-01	
<input type="checkbox"/>		rt_aggr1	22.3 ms/op	267 IOPS	< 1 MBps	11%	6%	6,691 GB	6,693 GB	opm-mobility	opm-m...-01	
<input type="checkbox"/>		aggr2	15.6 ms/op	259 IOPS	1.03 MBps	11%	5%	18,472 GB	20,080 GB	opm-mobility	opm-m...-02	
<input type="checkbox"/>		aggr2	9.52 ms/op	87 IOPS	20.8 MBps	Not Supported	5%	847 GB	984 GB	opm-lo...vity	opm-lo...ty-01	aggr_IOPS
<input type="checkbox"/>		RTaggr	7.62 ms/op	199 IOPS	34.7 MBps	Not Supported	6%	1,292 GB	1,477 GB	opm-lo...vity	opm-lo...ty-01	aggr_IOPS

O monitoramento do contador de capacidade de desempenho usado permite identificar o seguinte:

- Independentemente de nós ou agregados em qualquer cluster terem um valor usado de alta capacidade de performance
- Mesmo que nós ou agregados em qualquer cluster tenham eventos usados de capacidade de performance ativa
- Os nós e agregados que têm o valor mais alto e mais baixo da capacidade de performance usada em um cluster
- Valores de contador de latência e utilização em conjunto com nós ou agregados que têm valores de capacidade de alta performance usados
- Como os valores da capacidade de performance usada para nós em um par de HA serão afetados se um dos nós falhar
- Os volumes e LUNs mais ativos em um agregado que tenha um valor de capacidade de alta performance usada

Exibindo o nó e agregar valores de IOPS disponíveis

Você pode monitorar os valores de IOPS disponíveis para todos os nós ou para todos os agregados em um cluster, ou visualizar detalhes de um único nó ou agregado.

Os valores de IOPS disponíveis aparecem nas páginas Inventário de desempenho e nos gráficos de páginas do Performance Explorer para nós e agregados. Por exemplo, ao visualizar um nó na página do Explorador de nó/desempenho, você pode selecionar o gráfico de contador "Available IOPS" na lista para comparar os valores de IOPS disponíveis para o nó e vários agregados nesse nó.



O monitoramento do contador de IOPS disponível permite identificar:

- Nós ou agregados que têm os maiores valores de IOPS disponíveis para ajudar a determinar onde workloads futuros podem ser implantados.
- Os nós ou agregados que têm os menores valores de IOPS disponíveis para identificar os recursos que você deve monitorar para possíveis problemas de desempenho futuros.
- Os volumes e LUNs mais ativos em um agregado que tenha um pequeno valor de IOPS disponível.

Visualização de gráficos de contadores de capacidade de desempenho para identificar problemas

Você pode visualizar gráficos usados de capacidade de performance para nós e agregados na página Performance Explorer. Isso permite visualizar dados detalhados de capacidade de performance para os nós selecionados e agregados para um período de tempo específico.

Sobre esta tarefa

O gráfico de contador padrão exibe os valores de capacidade de desempenho usados para os nós ou agregados selecionados. O gráfico de contador de avarias exibe os valores de capacidade de desempenho total para o objeto raiz separado em uso com base em protocolos de usuário versus processos de sistema em segundo plano. Além disso, a quantidade de capacidade de desempenho livre também é mostrada.

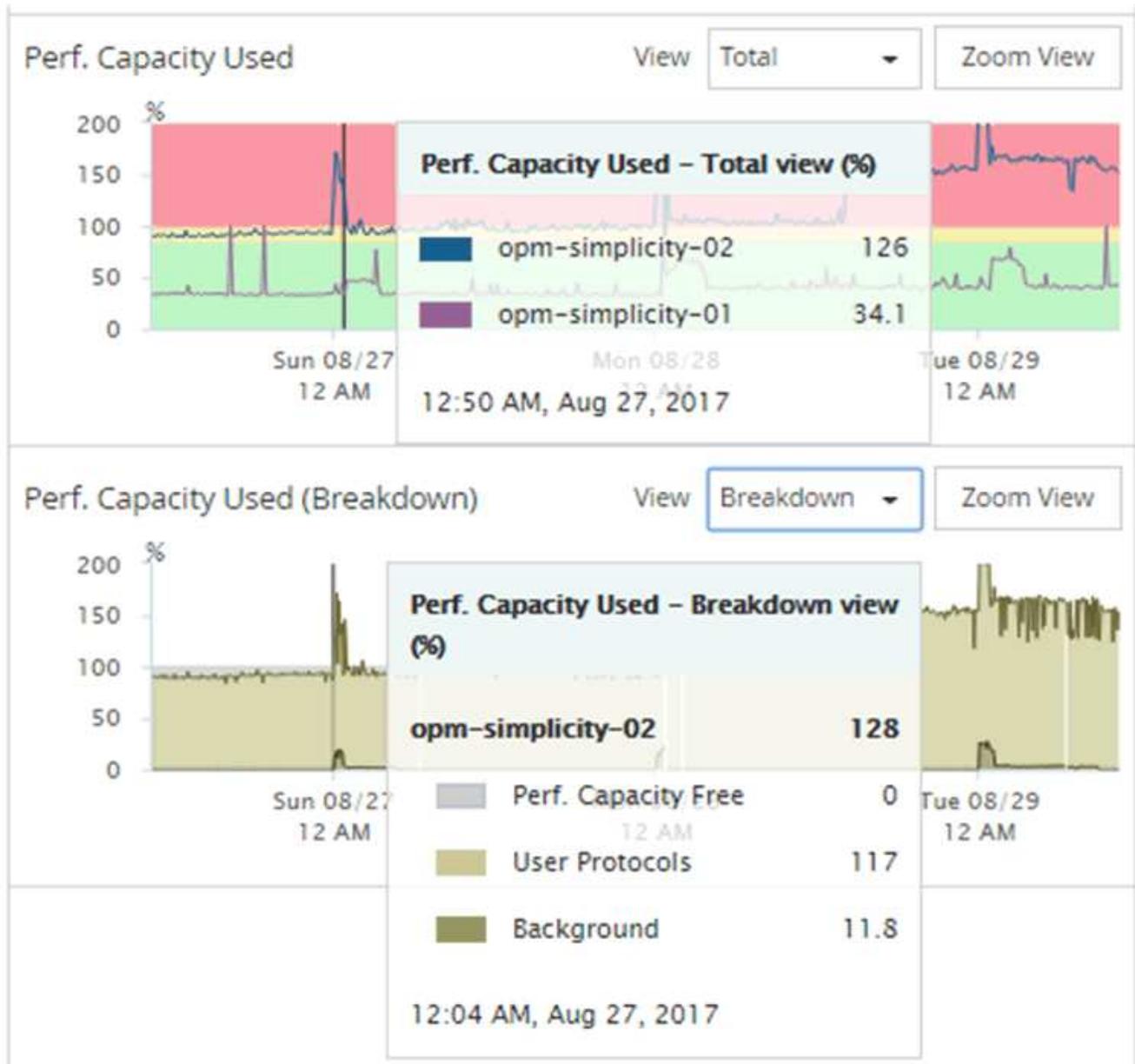


Como algumas atividades em segundo plano associadas ao gerenciamento de dados e sistema são identificadas como cargas de trabalho do usuário e categorizadas como protocolos de usuário, a porcentagem de protocolos do usuário pode parecer artificialmente alta quando esses processos são executados. Esses processos normalmente são executados por volta da meia-noite quando o uso do cluster é baixo. Se você vir um pico na atividade do protocolo do usuário por volta da meia-noite, verifique se os trabalhos de backup do cluster ou outras atividades em segundo plano estão configurados para serem executados nesse momento.

Passos

1. Selecione a guia **Explorer** de um nó ou agregar a página **Landing**.
2. No painel **Counter Charts**, clique em **Choose charts** e selecione **Perf. Gráfico de capacidade utilizada**.
3. Role para baixo até que você possa visualizar o gráfico.

As cores do gráfico padrão mostram quando o objeto está no intervalo ideal (amarelo), quando o objeto está subutilizado (verde) e quando o objeto é sobreutilizado (vermelho). O gráfico de detalhamento mostra detalhes detalhados da capacidade de desempenho somente para o objeto raiz.



4. Se você quiser exibir qualquer gráfico em um formato completo, clique em **Zoom View**.

Dessa forma, você pode abrir vários gráficos de contador em uma janela separada para comparar os valores de capacidade de desempenho usados com valores de IOPS ou Mbps no mesmo período de tempo.

Condições de limite de performance usadas pela capacidade de performance

Você pode criar políticas de limite de performance definidas pelo usuário para que os eventos sejam acionados quando o valor de capacidade de performance usada para um nó ou agregado exceder a configuração de limite de capacidade de performance definida.

Além disso, os nós podem ser configurados com uma política de limite de capacidade de desempenho usada

de aquisição. Essa política de limite totaliza as estatísticas de capacidade de performance usadas para ambos os nós em um par de HA para determinar se um dos nós não teria capacidade suficiente se o outro nó falhar. Como a carga de trabalho durante o failover é a combinação das cargas de trabalho dos dois nós parceiros, a mesma capacidade de performance usada política de takeover pode ser aplicada a ambos os nós.



Essa capacidade de desempenho utilizada é geralmente verdadeira entre nós. No entanto, se houver significativamente mais tráfego entre nós destinado a um dos nós por meio de seu parceiro de failover, a capacidade total de performance usada ao executar todos os workloads em um nó de parceiro em vez do outro nó de parceiro poderá ser um pouco diferente, dependendo de qual nó tiver falhado.

As condições de capacidade de performance usadas também podem ser usadas como configurações de limite de performance secundário para criar uma política de limite de combinação ao definir limites para LUNs e volumes. A condição de capacidade de desempenho usada é aplicada ao agregado ou nó no qual reside o volume ou LUN. Por exemplo, você pode criar uma política de limite de combinação usando os seguintes critérios:

Objeto de storage	Contador de desempenho	Limite de aviso	Limite crítico	Duração
Volume	Latência	15 ms/op	25 ms/op	20 minutos

As políticas de limite de combinação fazem com que um evento seja gerado somente quando ambas as condições forem violadas durante toda a duração.

Usando o contador de capacidade de desempenho usado para gerenciar o desempenho

Em geral, as organizações querem operar com uma porcentagem de capacidade de performance usada abaixo de 100 para que os recursos estejam sendo usados com eficiência, reservando alguma capacidade de performance adicional para atender às demandas do período de pico. Você pode usar políticas de limite para personalizar quando os alertas são enviados para valores usados de capacidade de alta performance.

Você pode estabelecer metas específicas com base em seus requisitos de performance. Por exemplo, as empresas de serviços financeiros podem reservar mais capacidade de desempenho para garantir a execução oportuna de negociações. Essas empresas podem querer definir limites de capacidade de desempenho usados na faixa de 70-80%. Empresas de manufatura com margens menores podem optar por reservar menos capacidade de desempenho se estiverem dispostas a arriscar o desempenho para gerenciar melhor os custos DE TI. Essas empresas podem definir limites de capacidade de desempenho usados na faixa de 85-95%.

Quando o valor da capacidade de performance usada excede a porcentagem definida em uma política de limite definida pelo usuário, o Unified Manager envia um e-mail de alerta e adiciona o evento à página Inventário de Eventos. Isso permite que você gerencie possíveis problemas antes que eles afetem o desempenho. Esses eventos também podem ser usados como indicadores de que você precisa para fazer mudanças e mudanças de workload em seus nós e agregados.

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 INTERRUPTÃO 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.