Performance-Management mithilfe von Performance-Kapazität und verfügbaren IOPS-Informationen

Active IQ Unified Manager 9.9

NetApp
January 31, 2025

This PDF was generated from https://docs.netapp.com/de-de/active-iq-unified-manager-99/performance-checker/concept-what-performance-capacity-used-is.html on January 31, 2025. Always check docs.netapp.com for the latest.

Inhalt

Performance-Management mithilfe von Performance-Kapazität und verfügbaren IOPS-Informationen 1
Welche Performance-Kapazität wird verwendet 1
Was der Wert der verwendeten Performance-Kapazität bedeutet 2
Was verfügbar ist, ist IOPS 3
Anzeigen der verwendeten Werte für die Node- und Aggregat-Performance 4
Anzeigen der verfügbaren IOPS-Werte für Node und Aggregat 5
Anzeigen von Zählerdiagrammen zur Performance-Kapazität zur Erkennung von Problemen 6
Performance-Kapazität nutzte Schwellenwertbedingungen für die Performance 8
Verwenden der Performance-Kapazität, die zum Managen der Performance verwendet wird 9

Performance-Management mithilfe von Performance-Kapazität und verfügbaren IOPS-Informationen

Performance Capacity zeigt an, wie viel Durchsatz eine Ressource erhalten kann, ohne die nützliche Performance dieser Ressource zu überschreiten. Wenn Sie die Nutzung vorhandener Performance-Zähler verwenden, ist die Performance-Kapazität der Punkt, an dem Sie die maximale Auslastung eines Node oder Aggregats erhalten, bevor die Latenz zu einem Problem wird.

Unified Manager sammelt Performance-Kapazitätsstatistiken von Nodes und Aggregaten in jedem Cluster. *Performance-Kapazität verwendet* ist der Prozentsatz der derzeit genutzten Performance-Kapazität und *Performance Capacity free* ist der Prozentsatz der noch verfügbaren Performance-Kapazität.

Während die freie Performance-Kapazität einen Prozentsatz der noch verfügbaren Ressource bietet, gibt Ihnen *verfügbare IOPS* die Anzahl an IOPS an, die der Ressource hinzugefügt werden können, bevor sie die maximale Performance-Kapazität erreicht. Mithilfe dieser Kennzahl können Sie sicherstellen, dass Sie Workloads mit einer vorab festgelegten Anzahl von IOPS zu einer Ressource hinzufügen können.

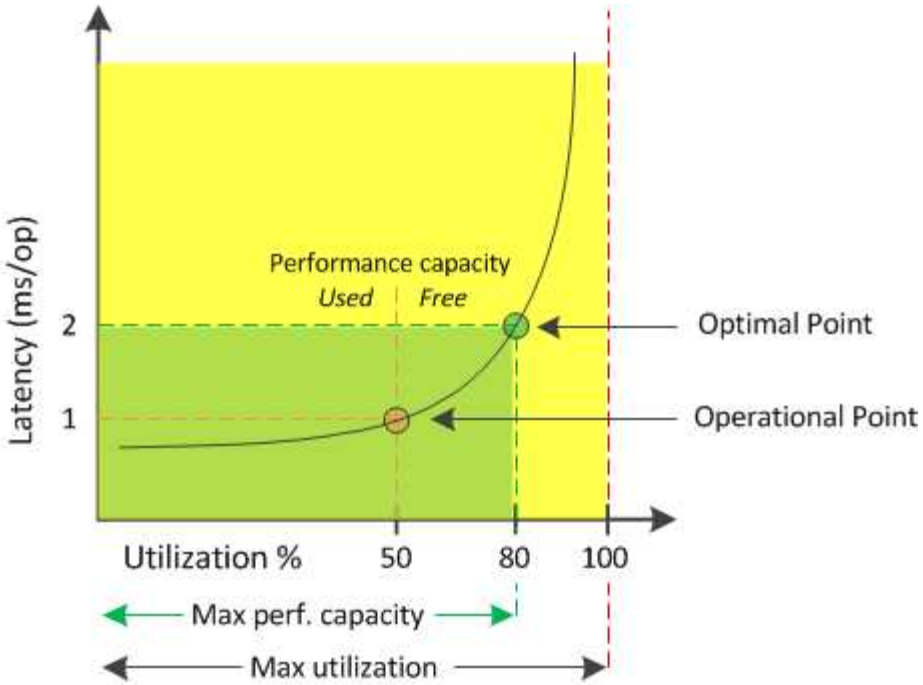
Das Monitoring der Informationen zur Performance-Kapazität bietet folgende Vorteile:

- Hilft bei der Workflow-Bereitstellung und beim Lastausgleich.
- Hilft Ihnen, eine Überlastung eines Knotens zu verhindern oder seine Ressourcen über den optimalen Punkt hinaus zu schieben, wodurch die Fehlerbehebung verringert wird.
- Präzisere Bestimmung, bei denen möglicherweise zusätzliche Storage-Geräte erforderlich sind

Welche Performance-Kapazität wird verwendet

Mit dem verwendeten Zähler für die Performance-Kapazität können Sie ermitteln, ob die Performance eines Node oder Aggregats an einem Punkt anliegt, an dem sich die Performance bei einer Workload-Steigerung verschlechtern kann. Es kann Ihnen auch zeigen, ob ein Node oder Aggregat derzeit zu bestimmten Zeiten überlastet ist. Die Verwendung der Performance-Kapazität ähnelt der Auslastung, allerdings bietet die vorherige Appliance einen besseren Einblick in die verfügbaren Performance-Funktionen in einer physischen Ressource für einen bestimmten Workload.

Die optimale genutzte Performance-Kapazität ist der Punkt, an dem ein Node oder Aggregat optimale Auslastung und Latenz (Reaktionszeit) bietet und effizient eingesetzt wird. In der folgenden Abbildung ist eine Beispiellatenz im Vergleich zur Auslastungskurve gezeigt für ein Aggregat.



In diesem Beispiel weist der_Operationspunkt_ darauf hin, dass das Aggregat derzeit bei einer Auslastung von 50 % und einer Latenz von 1.0 ms/op. Arbeitet Basierend auf den vom Aggregat erfassten Statistiken stellt Unified Manager fest, dass für dieses Aggregat zusätzliche Performance-Kapazität verfügbar ist. In diesem Beispiel wird der *optimale Punkt* als Punkt identifiziert, an dem das Aggregat bei einer Auslastung von 80 % und einer Latenz von 2.0 ms/op. Liegt Somit können Sie diesem Aggregat auch mehr Volumes und LUNs hinzufügen, sodass Ihre Systeme effizienter genutzt werden.

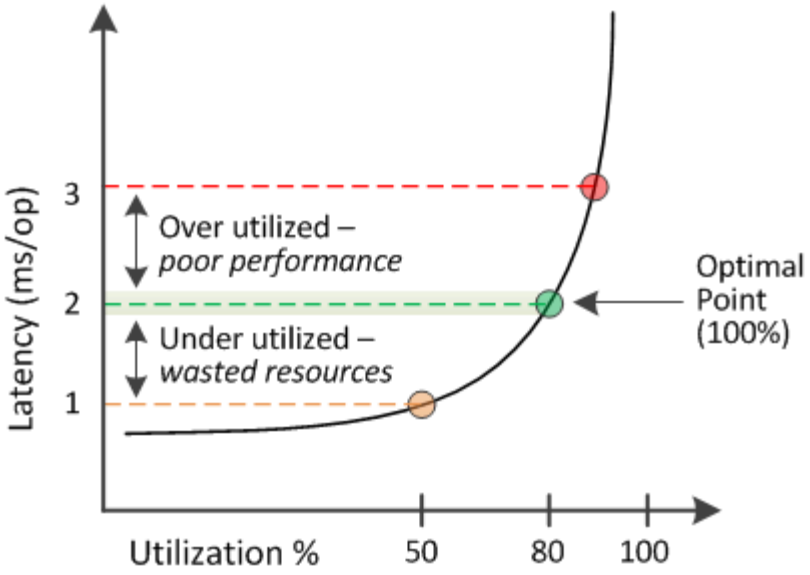
Der verwendete Zähler für die Performance-Kapazität soll eine größere Zahl sein als der Zähler „`Auslastung`“, da die Performance-Kapazität die Auswirkungen auf die Latenz hinzufügt. Wenn beispielsweise ein Node oder Aggregat zu 70 % genutzt wird, kann der Wert für die Performance-Kapazität je nach Latenzwert im Bereich von 80 % bis 100 % liegen.

In einigen Fällen kann der Auslastungszähler jedoch höher auf der Dashboard-Seite sein. Dies ist normal, da das Dashboard die aktuellen Zählerwerte zu jedem Erfassungszeitraum aktualisiert. Es werden die Durchschnittswerte über einen Zeitraum nicht angezeigt, wie die anderen Seiten in der Benutzeroberfläche von Unified Manager. Der verwendete Zähler für die Performance-Kapazität dient am besten als Indikator für die über einen Zeitraum gemittelte Performance, während der Auslastungszähler am besten zur Ermittlung der unmittelbaren Nutzung einer Ressource genutzt wird.

Was der Wert der verwendeten Performance-Kapazität bedeutet

Der von Performance genutzte Wert hilft Ihnen dabei, die Nodes und Aggregate zu identifizieren, die derzeit zu stark ausgelastet sind oder nicht ausgelastet sind. Auf diese Weise können Sie Workloads neu verteilen, um die Effizienz der Storage-Ressourcen zu steigern.

Die folgende Abbildung zeigt die Latenz im Vergleich zur Auslastungskurve einer Ressource und identifiziert mit farbigen Punkten drei Bereiche, in denen sich der aktuelle betriebliche Punkt befinden könnte.



- Ein Prozentsatz der genutzten Performance-Kapazität gleich 100 ist am optimalen Punkt.

  Die Ressourcen werden jetzt effizient genutzt.

- Ein Prozentsatz an Performance, der höher als 100 ist, bedeutet, dass der Node oder das Aggregat zu stark ausgelastet ist und dass Workloads eine suboptimale Performance erhalten.

  Der Ressource sollten keine neuen Workloads hinzugefügt werden, die bestehende Workloads müssen eventuell neu verteilt werden.

- Ein Prozentsatz an Performance-Kapazität unter 100 zeigt an, dass der Node oder das Aggregat nicht ausgelastet ist und dass die Ressourcen nicht effizient genutzt werden.

  Der Ressource können weitere Workloads hinzugefügt werden.

Im Gegensatz zur Auslastung kann die genutzte Performance-Kapazität über 100 % liegen. Es gibt keinen maximalen Prozentsatz, aber Ressourcen werden in der Regel zwischen 110 % und 140 % liegen, wenn sie überausgelastet sind. Höhere Prozentsätze deuten auf eine Ressource mit schwerwiegenden Problemen hin.

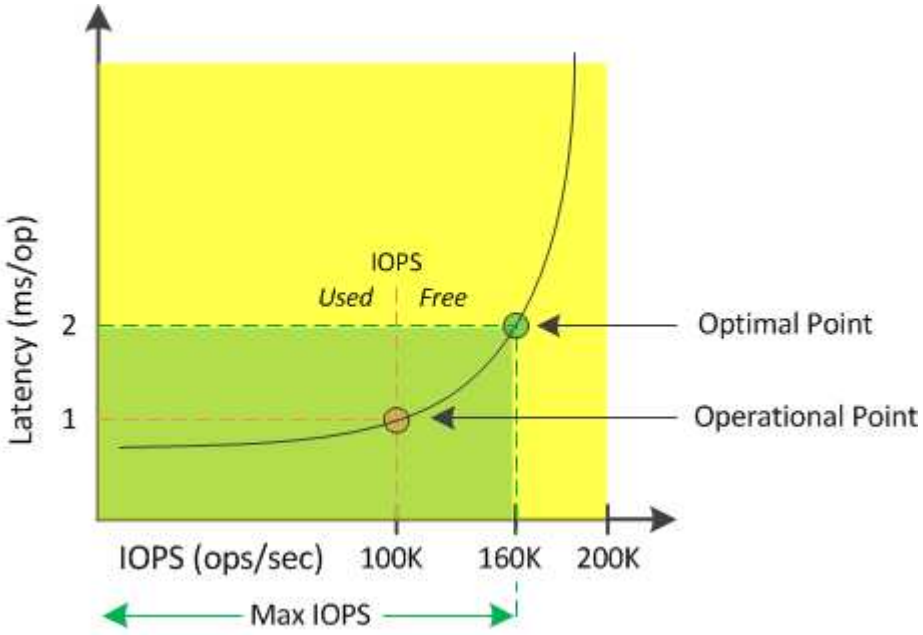
Was verfügbar ist, ist IOPS

Der verfügbare IOPS-Zähler ermittelt die verbleibende Anzahl an IOPS, die einem Node oder Aggregat hinzugefügt werden kann, bevor die Ressource ihr Limit erreicht. Die gesamten IOPS, die ein Node bereitstellen kann, basieren auf den physischen Eigenschaften des Node, beispielsweise auf der Anzahl der CPUs, der CPU-Geschwindigkeit und dem RAM-Umfang. Die gesamten IOPS, die ein Aggregat bereitstellen kann, basieren auf den physischen Eigenschaften der Festplatten, zum Beispiel auf einer SATA-, SAS- oder SSD-Festplatte.

Während der freie Zähler für Performance-Kapazität den Prozentsatz einer noch verfügbaren Ressource angibt, liefert der verfügbare IOPS-Zähler genau die Anzahl der IOPS (Workloads) für eine Ressource, bevor sie die maximale Performance-Kapazität erreicht.

Wenn Sie beispielsweise zwei Storage-Systeme FAS2520 und FAS8060 nutzen, erhalten Sie bei einem Performance-Kapazitätswert von 30 % auch freie Performance-Kapazität. Dieser Wert bietet jedoch keine Übersicht darüber, wie viele Workloads auf diesen Nodes implementiert werden können. Der verfügbare IOPS-Zähler zeigt möglicherweise, dass bei der FAS8060 500 verfügbare IOPS verfügbar sind, bei der FAS2520 jedoch nur 100 verfügbare IOPS.

Die folgende Abbildung zeigt eine Beispiellatenz im Vergleich zur IOPS-Kurve für einen Node.



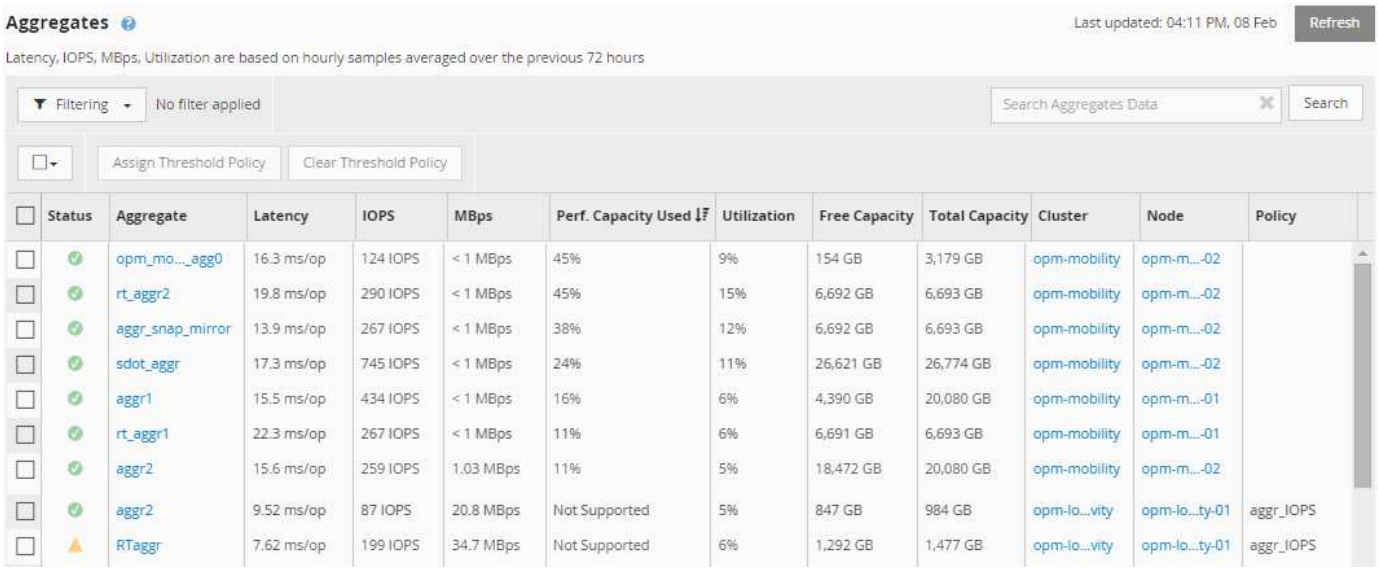
Die maximale Anzahl an IOPS, die eine Ressource liefern kann, ist die Anzahl der IOPS, wenn der Zähler für die Performance-Kapazität bei 100 % liegt (der optimale Punkt). Der Betriebspunkt identifiziert, dass der Node derzeit bei 100.000 IOPS mit einer Latenz von 1.0 ms/op arbeitet Auf der Grundlage der vom Node erfassten Statistiken ermittelt Unified Manager, dass die maximalen IOPS für den Node 160.000 beträgt, was bedeutet, dass 60.000 freie oder verfügbare IOPS vorhanden sind. Daher können Sie diesem Node weitere Workloads hinzufügen, sodass die Systeme effizienter genutzt werden.

Wenn die Ressource minimale Benutzeraktivitäten hat, wird der verfügbare IOPS-Wert unter Annahme eines allgemeinen Workloads auf Grundlage von etwa 4,500 IOPS pro CPU-Kern berechnet. Das liegt daran, dass Unified Manager nicht über die Daten verfügt, um die Eigenschaften des Workloads, der bereitgestellt werden soll, exakt abzuschätzen.

Anzeigen der verwendeten Werte für die Node- und Aggregat-Performance

Sie können die verwendeten Werte für die Performance-Kapazität für alle Nodes oder für alle Aggregate in einem Cluster überwachen oder Details für einen einzelnen Node oder Aggregat anzeigen.

Die verwendeten Werte für die Performance-Kapazität werden auf der Seite Dashboard, auf den Seiten Performance Inventory, auf der Seite Top Performers, Create Threshold Policy, auf den Seiten Performance Explorer und in Detaildiagrammen angezeigt. Beispielsweise bietet die Seite Performance: Alle Aggregate eine Spalte. Die Performance-Kapazität wird verwendet, um den verwendeten Wert für die Performance-Kapazität aller Aggregate anzuzeigen.



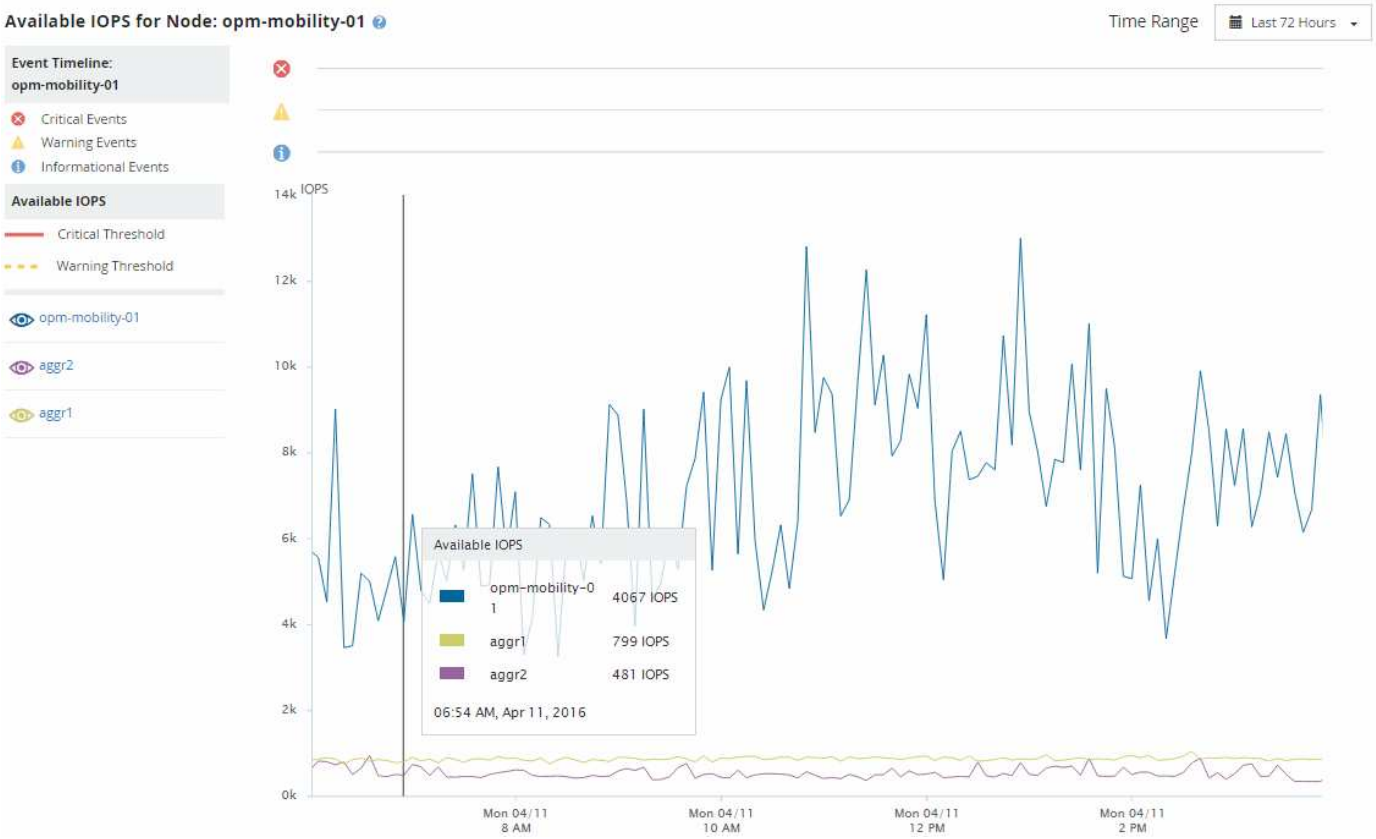
Durch das Monitoring des verwendeten Zählers für die Performance-Kapazität können Sie Folgendes identifizieren:

- Unabhängig davon, ob Nodes oder Aggregate auf jedem Cluster einen hohen Wert an Performance-Kapazität aufweisen
- Gibt an, ob Nodes oder Aggregate auf beliebigen Clustern über Ereignisse mit aktiver Performance-Kapazität verfügen
- Die Nodes und Aggregate verfügen über die in einem Cluster genutzte Kapazität mit der höchsten und niedrigsten Performance
- Zählerwerte im Bereich Latenz und Auslastung in Verbindung mit Nodes oder Aggregaten mit hohen Werten im Bereich der Performance-Kapazität
- Auswirkungen auf die verwendete Performance-Kapazität für Nodes in einem HA-Paar auf einen Node, wenn einer der Nodes ausfällt
- Die am stärksten ausgelasteten Volumes und LUNs auf einem Aggregat mit hoher Performance-Kapazität

Anzeigen der verfügbaren IOPS-Werte für Node und Aggregat

Sie können die verfügbaren IOPS-Werte für alle Nodes und alle Aggregate in einem Cluster überwachen. Alternativ können Sie Details zu einem einzelnen Node oder Aggregat anzeigen.

Verfügbare IOPS-Werte werden auf den Seiten „Performance Inventory“ und in den Seitendiagrammen des Performance Explorers für Nodes und Aggregate angezeigt. Wenn Sie beispielsweise einen Node auf der Seite „Node/Performance Explorer“ anzeigen, können Sie das Zählerdiagramm „`verfügbare IOPS`“ in der Liste auswählen, damit Sie die verfügbaren IOPS-Werte für den Node und mehrere Aggregate auf diesem Node vergleichen können.



Durch das Monitoring des verfügbaren IOPS-Zählers können Sie Folgendes identifizieren:

- Die Nodes oder Aggregate mit den höchsten verfügbaren IOPS-Werten unterstützen Sie bei der Entscheidung, wo zukünftige Workloads implementiert werden können.
- Die Nodes oder Aggregate mit den kleinsten verfügbaren IOPS-Werten, um die zu überwachenden Ressourcen auf potenzielle künftige Performance-Probleme zu identifizieren.
- Die am stärksten ausgelasteten Volumes und LUNs auf einem Aggregat mit kleinem verfügbaren IOPS-Wert.

Anzeigen von Zählerdiagrammen zur Performance-Kapazität zur Erkennung von Problemen

Auf der Seite „Performance Explorer“ können Sie die verwendeten Performance-Kapazitäten für Nodes und Aggregate anzeigen. Damit können Sie detaillierte Performance-Kapazitätsdaten für die ausgewählten Nodes und Aggregate für einen bestimmten Zeitraum anzeigen.

Über diese Aufgabe

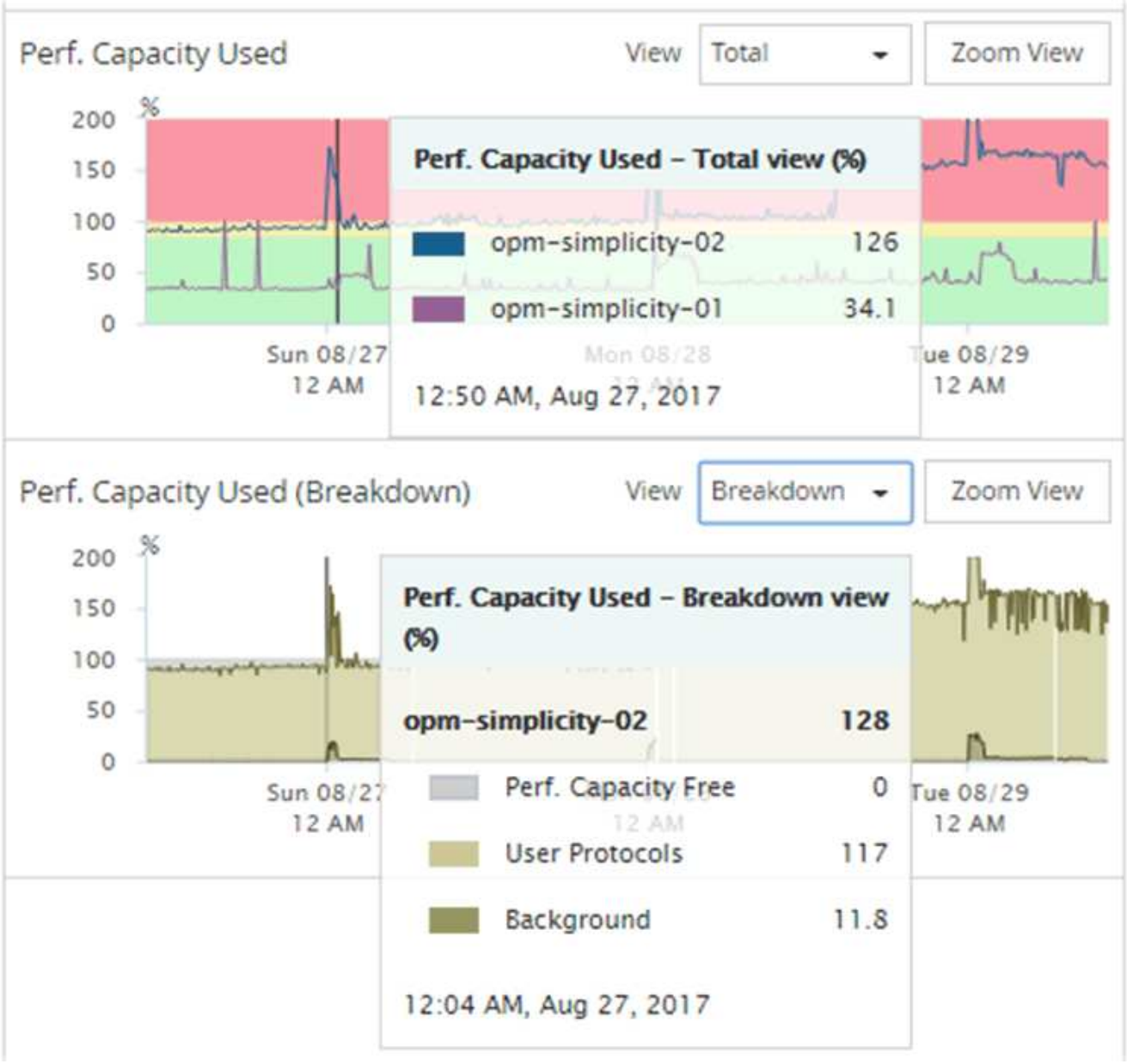
Das Standard-Zählerdiagramm zeigt die Werte der verwendeten Performance-Kapazität für die ausgewählten Nodes oder Aggregate an. Das Counter Chart Breakdown zeigt die Werte für die Gesamtkapazität des Root-Objekts an, das basierend auf Benutzerprotokollen und Hintergrundsystemprozessen in die Nutzung unterteilt ist. Darüber hinaus wird die Menge der freien Performance-Kapazität dargestellt.

Da einige Hintergrundaktivitäten zu System- und Datenmanagement als Benutzer-Workloads identifiziert und als Benutzerprotokolle kategorisiert werden, erscheint der Prozentsatz der Benutzerprotokolle künstlich hoch, wenn diese Prozesse ausgeführt werden. Diese Prozesse laufen normalerweise um Mitternacht, wenn die Cluster-Nutzung gering ist. Wenn bei der Durchführung von Benutzerprotokollen um Mitternacht eine Spitze sichtbar ist, überprüfen Sie, ob Cluster-Backup-Jobs oder andere Hintergrundaktivitäten zu diesem Zeitpunkt konfiguriert wurden.

Schritte

1. Wählen Sie die **Explorer** Registerkarte von einem Knoten oder Aggregat * Landing* Seite.
2. Klicken Sie im Fenster **Counter Charts** auf **Choose Charts** und wählen Sie dann die Option *Perf. Diagramm „verwendete Kapazität“.
3. Blättern Sie nach unten, bis Sie das Diagramm anzeigen können.

   Die Farben des Standarddiagramms zeigen an, wenn sich das Objekt im optimalen Bereich (gelb) befindet, wenn das Objekt nicht ausgelastet ist (grün) und wenn das Objekt überausgelastet ist (rot). Das Diagramm zeigt detaillierte Performance-Kapazitätsdetails nur für das Root-Objekt.



4. Wenn Sie eine Karte in einem vollen Format anzeigen möchten, klicken Sie auf **Zoom View**.

   Auf diese Weise können Sie mehrere Zählerdiagramme in einem separaten Fenster öffnen, um die genutzte Performance-Kapazität mit IOPS- oder MB/s-Werten im gleichen Zeitraum zu vergleichen.

Performance-Kapazität nutzte Schwellenwertbedingungen für die Performance

Sie können benutzerdefinierte Performance-Schwellenwertrichtlinien erstellen, damit Ereignisse ausgelöst werden, wenn der für einen Node oder ein Aggregat genutzte Performance-Wert die festgelegte Einstellung für den verwendeten Schwellenwert für die Performance-Kapazität überschreitet.

Außerdem können Nodes mit einem Schwellenwert von „`Performance Capacity used Takeover`“ konfiguriert werden. Diese Schwellenwertrichtlinie gibt die für beide Nodes verwendeten Performance-

Statistiken in einem HA-Paar an und ermittelt, ob einem der beiden Nodes genügend Kapazität fehlen würde, wenn der andere Node ausfällt. Da der Workload während des Failover` Kombination der Workloads der beiden Partner-Nodes ist, kann die gleiche Performance-Kapazität, die für eine Übernahme verwendet wird, auf beide Nodes angewendet werden.

> Die genutzte Performance-Kapazität entspricht der Leistung im Allgemeinen den Nodes. Ist der Node-übergreifende Datenverkehr jedoch über seinen Failover-Partner für einen der Nodes vorgesehen, kann die Gesamt-Performance, die bei der Ausführung aller Workloads auf einem Partner-Node verwendet wird, auf einem anderen Partner-Node geringfügig anders sein – je nachdem, welcher Node ausgefallen ist.

Die verwendeten Performance-Kapazitäten können auch als sekundäre Performance-Schwellenwerteinstellungen verwendet werden, um bei der Definition von Schwellenwerten für LUNs und Volumes eine kombinierte Schwellenwertrichtlinie zu erstellen. Die verwendete Performance-Kapazität wird auf das Aggregat oder den Node angewendet, auf dem sich das Volume oder die LUN befindet. Sie können beispielsweise anhand der folgenden Kriterien eine kombinierte Schwellenwertrichtlinie erstellen:

Storage Objekt	Performance-Zähler	Warnschwellenwert	Kritischer Schwellenwert	Dauer
Datenmenge	Latenz	15 ms/OP	25 ms/OP	20 Minuten

Aufgrund von Grenzwertrichtlinien wird ein Ereignis nur erzeugt, wenn beide Bedingungen während der gesamten Dauer nicht erfüllt werden.

Verwenden der Performance-Kapazität, die zum Managen der Performance verwendet wird

In der Regel möchten Unternehmen mit einer Performance-Kapazität von unter 100 Prozent betreiben, um die Ressourcen effizient zu nutzen und gleichzeitig zusätzliche Performance-Kapazitäten zu reservieren, um Spitzenlasten zu erzielen. Anhand von Schwellenwertrichtlinien kann angepasst werden, wenn Warnmeldungen gesendet werden, um hohe Werte für die verwendete Kapazität zu erreichen.

Sie können bestimmte Ziele basierend auf Ihren Performance-Anforderungen festlegen. So könnten Finanzdienstleister mehr Performance-Kapazität reservieren, um eine zeitnahe Ausführung von Transaktionen zu gewährleisten. Diese Unternehmen möchten möglicherweise Schwellenwerte für die verwendete Performance-Kapazität im Bereich von 70-80 Prozent festlegen. Produzierende Unternehmen mit geringeren Margen können sich für weniger Performance-Kapazität entscheiden, wenn sie bereit sind, Performance zu riskieren, um DIE IT-Kosten besser zu managen. Diese Unternehmen können Grenzwerte für die verwendete Performance-Kapazität im Bereich von 85-95 Prozent festlegen.

Wenn der verwendete Wert für die Performance-Kapazität den in einer benutzerdefinierten Schwellenwertrichtlinie festgelegten Prozentsatz überschreitet, sendet Unified Manager eine Alarm-E-Mail und fügt das Ereignis zur Seite „Ereignisbestand“ hinzu. Auf diese Weise lassen sich potenzielle Probleme verwalten, bevor sie die Performance beeinträchtigen. Diese Ereignisse können auch als Indikatoren verwendet werden, an denen Sie Workloads und Änderungen innerhalb Ihrer Nodes und Aggregate vornehmen müssen.

Copyright-Informationen

Copyright © 2025 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGENDEINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter http://www.netapp.com/TM aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.