Gestire le performance utilizzando la capacità delle performance e le informazioni IOPS disponibili

Active IQ Unified Manager 9.8

NetApp
January 31, 2025

This PDF was generated from https://docs.netapp.com/it-it/active-iq-unified-manager-98/performance-checker/concept-what-performance-capacity-used-is.html on January 31, 2025. Always check docs.netapp.com for the latest.

Sommario

Gestire le performance utilizzando la capacità delle performance e le informazioni IOPS disponibili 1
Qual è la capacità di performance utilizzata 1
Cosa significa il valore utilizzato dalla capacità delle performance 2
Quali IOPS sono disponibili 3
Visualizzazione dei valori utilizzati per la capacità di nodo e le performance aggregate 4
Visualizzazione dei valori IOPS disponibili di nodo e aggregazione 5
Visualizzazione dei grafici dei contatori di capacità delle performance per identificare i problemi 6
La capacità di performance ha utilizzato le condizioni di soglia delle performance 8
Utilizzo della capacità di performance utilizzata per gestire le performance 9

Gestire le performance utilizzando la capacità delle performance e le informazioni IOPS disponibili

Performance Capacity indica la quantità di throughput che è possibile ottenere da una risorsa senza superare le utili performance di tale risorsa. Quando viene visualizzata utilizzando i contatori delle performance esistenti, la capacità delle performance è il punto in cui si ottiene il massimo utilizzo da un nodo o aggregato prima che la latenza diventi un problema.

Unified Manager raccoglie le statistiche sulla capacità delle performance dai nodi e dagli aggregati di ciascun cluster. *Capacità di performance utilizzata* è la percentuale di capacità di performance attualmente utilizzata e *capacità di performance libera* è la percentuale di capacità di performance ancora disponibile.

Mentre la capacità delle performance libera fornisce una percentuale della risorsa ancora disponibile, *IOPS disponibili* indica il numero di IOPS che possono essere aggiunti alla risorsa prima di raggiungere la capacità di performance massima. Utilizzando questa metrica, puoi essere sicuro di poter aggiungere carichi di lavoro di un numero predeterminato di IOPS a una risorsa.

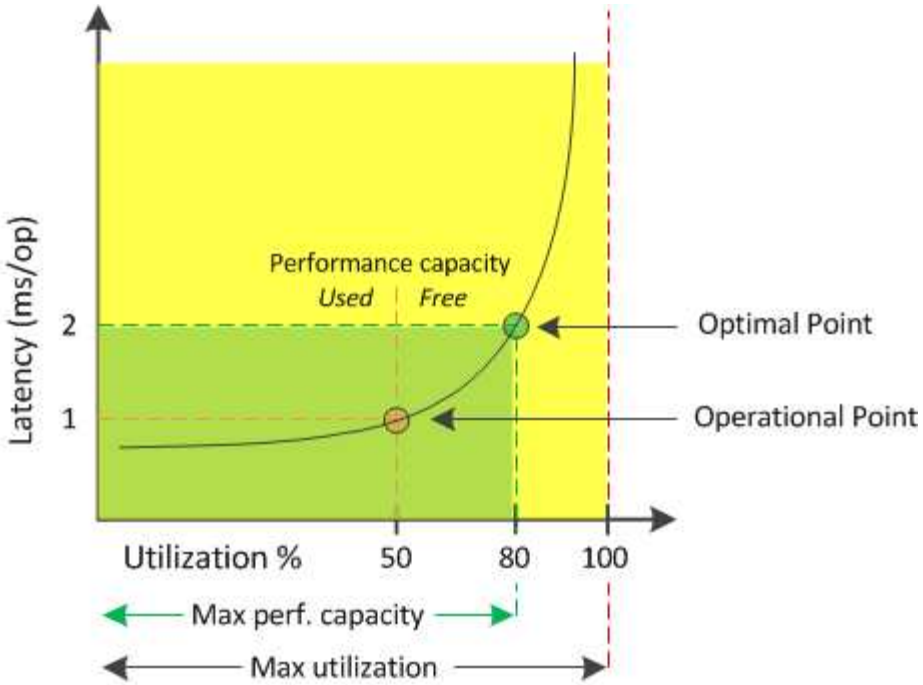
Il monitoraggio delle informazioni sulla capacità delle performance offre i seguenti vantaggi:

- Fornisce assistenza per il provisioning e il bilanciamento del workflow.
- Consente di evitare di sovraccaricare un nodo o di spingerne le risorse oltre il punto ottimale, riducendo così la necessità di eseguire il troubleshooting.
- Consente di determinare con maggiore precisione dove potrebbero essere necessarie apparecchiature di storage aggiuntive.

Qual è la capacità di performance utilizzata

Il contatore delle performance utilizzate consente di identificare se le performance di un nodo o di un aggregato stanno raggiungendo un punto in cui le performance potrebbero degradarsi se i carichi di lavoro aumentano. Può anche mostrare se un nodo o un aggregato è attualmente in uso in eccesso durante periodi di tempo specifici. La capacità di performance utilizzata è simile all'utilizzo, ma la prima fornisce maggiori informazioni sulle capacità di performance disponibili in una risorsa fisica per un carico di lavoro specifico.

La capacità di performance ottimale utilizzata è il punto in cui un nodo o un aggregato ha un utilizzo e una latenza ottimali (tempo di risposta) e viene utilizzato in modo efficiente. Nella figura seguente viene mostrata una curva di latenza rispetto all'utilizzo di esempio per un aggregato.



In questo esempio, il *punto operativo* indica che l'aggregato sta attualmente operando al 50% di utilizzo con una latenza di 1.0 ms/op. In base alle statistiche acquisite dall'aggregato, Unified Manager determina che è disponibile una capacità di performance aggiuntiva per questo aggregato. In questo esempio, il *punto ottimale* viene identificato come il punto in cui l'aggregato è al 80% di utilizzo con latenza di 2.0 ms/op. Pertanto, è possibile aggiungere più volumi e LUN a questo aggregato in modo che i sistemi vengano utilizzati in modo più efficiente.

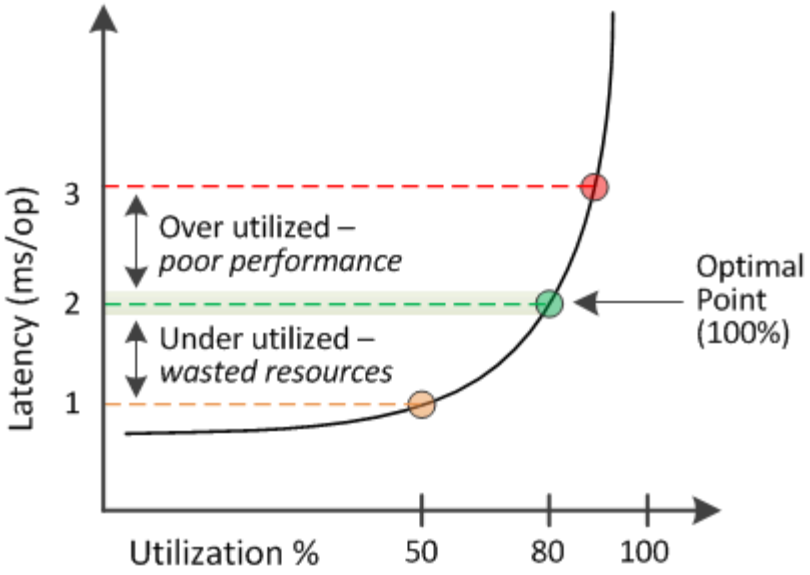
Si prevede che il contatore della capacità di performance utilizzata sia un numero maggiore del contatore "Utilization", in quanto la capacità di performance aumenta l'impatto sulla latenza. Ad esempio, se si utilizza un nodo o un aggregato al 70%, il valore della capacità delle performance può essere compreso tra il 80% e il 100%, a seconda del valore di latenza.

In alcuni casi, tuttavia, il contatore di utilizzo potrebbe essere più alto nella pagina Dashboard. Questo è normale perché il dashboard aggiorna i valori correnti del contatore in ogni periodo di raccolta; non visualizza le medie in un periodo di tempo come le altre pagine nell'interfaccia utente di Unified Manager. Il contatore della capacità di performance utilizzata viene utilizzato al meglio come indicatore delle performance medie in un periodo di tempo, mentre il contatore di utilizzo viene utilizzato al meglio per determinare l'utilizzo istantaneo di una risorsa.

Cosa significa il valore utilizzato dalla capacità delle performance

Il valore utilizzato per la capacità delle performance ti aiuta a identificare i nodi e gli aggregati che sono attualmente sovrautilizzati o sottoutilizzati. Ciò consente di ridistribuire i carichi di lavoro per rendere le risorse di storage più efficienti.

La figura seguente mostra la curva di latenza rispetto all'utilizzo di una risorsa e identifica, con punti colorati, tre aree in cui è possibile individuare il punto operativo corrente.



- Una percentuale di performance utilizzata pari a 100 è al punto ottimale.

  A questo punto, le risorse vengono utilizzate in modo efficiente.

- Una percentuale di performance utilizzata superiore a 100 indica che il nodo o l'aggregato è sovrautilizzato e che i carichi di lavoro ricevono performance non ottimali.

  Non aggiungere nuovi workload alla risorsa e potrebbe essere necessario ridistribuire i workload esistenti.

- Una percentuale di performance utilizzata inferiore a 100 indica che il nodo o l'aggregato è sottoutilizzato e che le risorse non vengono utilizzate in modo efficace.

  È possibile aggiungere più carichi di lavoro alla risorsa.

A differenza dell'utilizzo, la percentuale di performance della capacità utilizzata può essere superiore al 100%. Non esiste una percentuale massima, ma le risorse in genere rientrano nell'intervallo compreso tra il 110% e il 140% quando vengono utilizzate in eccesso. Percentuali più elevate indicano una risorsa con problemi gravi.

Quali IOPS sono disponibili

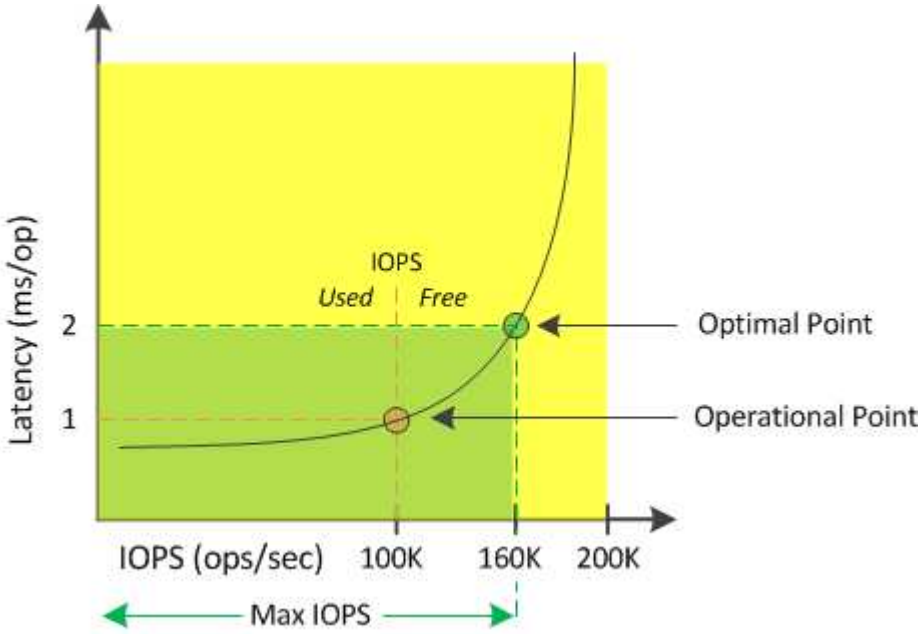
Il contatore IOPS disponibile identifica il numero rimanente di IOPS che è possibile aggiungere a un nodo o a un aggregato prima che la risorsa raggiunga il limite. Gli IOPS totali che un nodo può fornire si basano sulle caratteristiche fisiche del nodo, ad esempio il numero di CPU, la velocità della CPU e la quantità di RAM. Gli IOPS totali che un aggregato può fornire si basano sulle proprietà fisiche dei dischi, ad esempio un disco SATA, SAS o SSD.

Mentre il contatore di performance free fornisce la percentuale di una risorsa ancora disponibile, il contatore IOPS disponibile indica che è possibile aggiungere un numero esatto di IOPS (carichi di lavoro) a una risorsa prima di raggiungere la capacità di performance massima.

Ad esempio, se si utilizza una coppia di sistemi storage FAS2520 e FAS8060, un valore del 30% senza capacità di performance significa che si dispone di una certa capacità di performance libera. Tuttavia, questo

valore non fornisce visibilità sul numero di workload che è possibile implementare in tali nodi. Il contatore IOPS disponibile potrebbe indicare che sono disponibili 500 IOPS su FAS8060, ma solo 100 IOPS su FAS2520.

Nella figura seguente viene mostrato un esempio di latenza rispetto alla curva IOPS per un nodo.



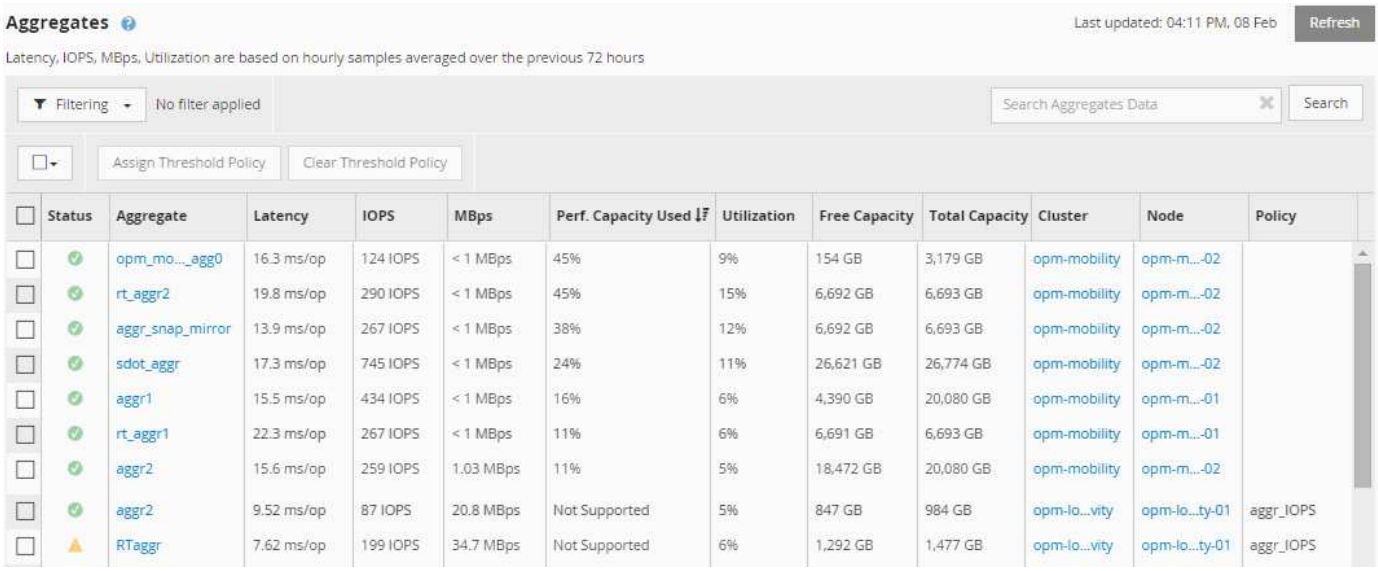
Il numero massimo di IOPS che una risorsa può fornire è il numero di IOPS quando il contatore della capacità di performance utilizzata è al 100% (il punto ottimale). Il punto operativo indica che il nodo sta attualmente operando a 100.000 IOPS con latenza di 1.0 ms/op. In base alle statistiche acquisite dal nodo, Unified Manager determina che il numero massimo di IOPS per il nodo è 160K, il che significa che ci sono 60K IOPS liberi o disponibili. Pertanto, è possibile aggiungere più carichi di lavoro a questo nodo in modo che i sistemi vengano utilizzati in modo più efficiente.

Quando l'attività dell'utente nella risorsa è minima, il valore IOPS disponibile viene calcolato ipotizzando un carico di lavoro generico basato su circa 4,500 IOPS per core della CPU. Ciò è dovuto al fatto che Unified Manager non dispone dei dati necessari per stimare con precisione le caratteristiche del carico di lavoro che viene servito.

Visualizzazione dei valori utilizzati per la capacità di nodo e le performance aggregate

È possibile monitorare i valori della capacità di performance utilizzata per tutti i nodi o per tutti gli aggregati di un cluster oppure visualizzare i dettagli di un singolo nodo o aggregato.

I valori utilizzati per la capacità delle performance vengono visualizzati nella dashboard, nelle pagine Performance Inventory, nella pagina Top Performer, nella pagina Create Threshold Policy, nelle pagine Performance Explorer e nei grafici dettagliati. Ad esempio, la pagina Performance: All aggregates (prestazioni: Tutti gli aggregati) fornisce una colonna Performance Capacity (capacità di performance) utilizzata per visualizzare il valore della capacità di performance utilizzata per tutti gli aggregati.



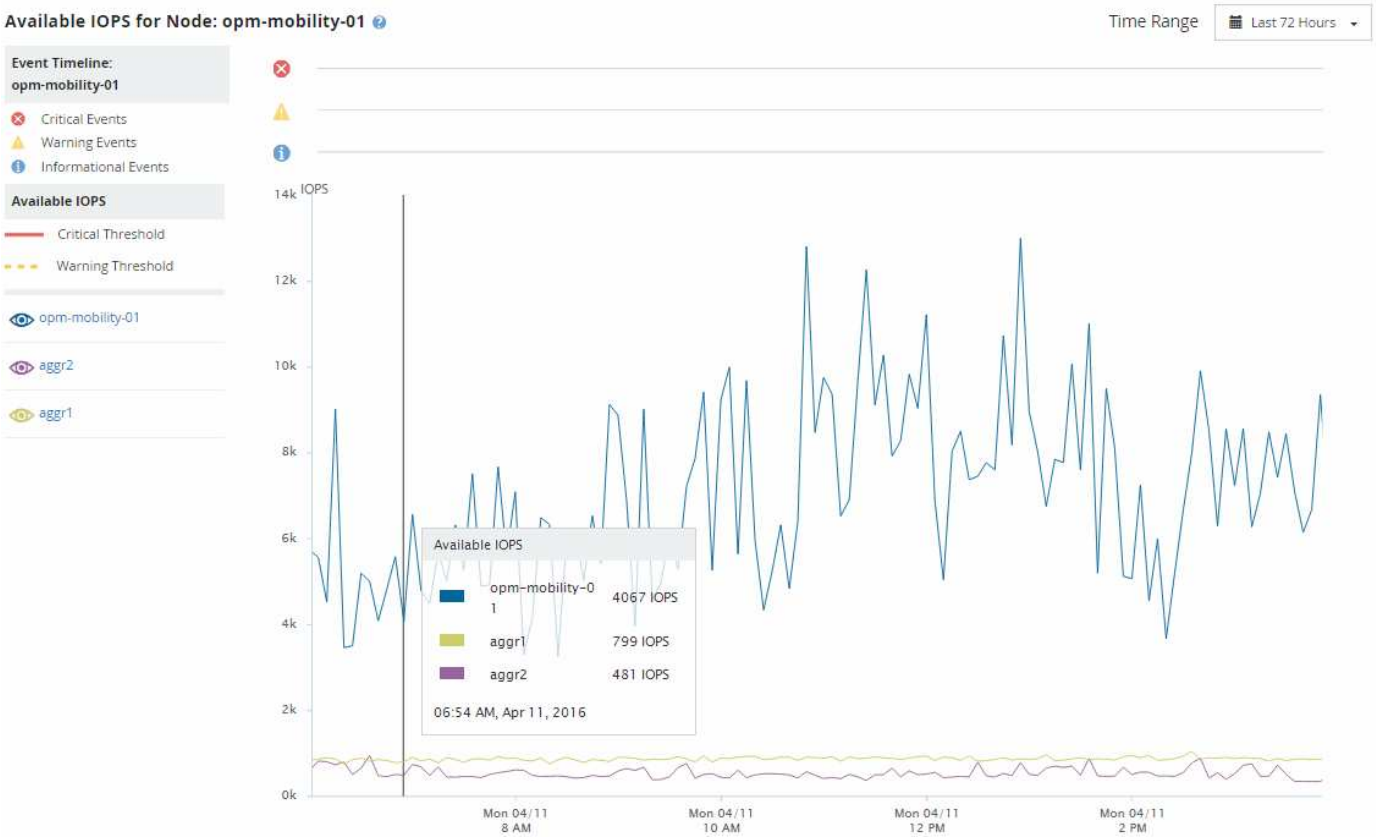
Il monitoraggio del contatore della capacità di performance utilizzata consente di identificare quanto segue:

- Sia che i nodi o gli aggregati di qualsiasi cluster abbiano un valore di utilizzo della capacità dalle performance elevate
- Sia che i nodi o gli aggregati di qualsiasi cluster abbiano eventi attivi di utilizzo della capacità delle performance
- I nodi e gli aggregati che hanno il valore più alto e più basso utilizzato per le performance in un cluster
- Valori dei contatori di latenza e utilizzo in combinazione con nodi o aggregati che hanno valori elevati di utilizzo della capacità delle performance
- In che modo la capacità di performance utilizzata per i nodi di una coppia ha sarà influenzata in caso di guasto di uno dei nodi
- I volumi e le LUN più impegnativi su un aggregato che ha un valore di utilizzo della capacità dalle performance elevate

Visualizzazione dei valori IOPS disponibili di nodo e aggregazione

È possibile monitorare i valori IOPS disponibili per tutti i nodi o per tutti gli aggregati di un cluster oppure visualizzare i dettagli di un singolo nodo o aggregato.

I valori IOPS disponibili vengono visualizzati nelle pagine Performance Inventory e nei grafici della pagina Performance Explorer per nodi e aggregati. Ad esempio, quando si visualizza un nodo nella pagina Node/Performance Explorer (Esplora nodi/prestazioni), è possibile selezionare il grafico del contatore “Available IOPS” (IOPS disponibili) dall’elenco in modo da poter confrontare i valori IOPS disponibili per il nodo e gli aggregati multipli su quel nodo.



Il monitoraggio del contatore IOPS disponibile consente di identificare:

- I nodi o gli aggregati che hanno i valori IOPS più elevati disponibili per determinare dove è possibile implementare i carichi di lavoro futuri.
- I nodi o gli aggregati che hanno i valori IOPS più piccoli disponibili per identificare le risorse da monitorare per potenziali problemi di performance futuri.
- I volumi e le LUN più impegnativi su un aggregato con un valore IOPS ridotto.

Visualizzazione dei grafici dei contatori di capacità delle performance per identificare i problemi

È possibile visualizzare i grafici relativi alla capacità di performance utilizzata per nodi e aggregati nella pagina Performance Explorer (Esplora prestazioni). In questo modo è possibile visualizzare dati dettagliati sulla capacità delle performance per i nodi e gli aggregati selezionati per un periodo di tempo specifico.

A proposito di questa attività

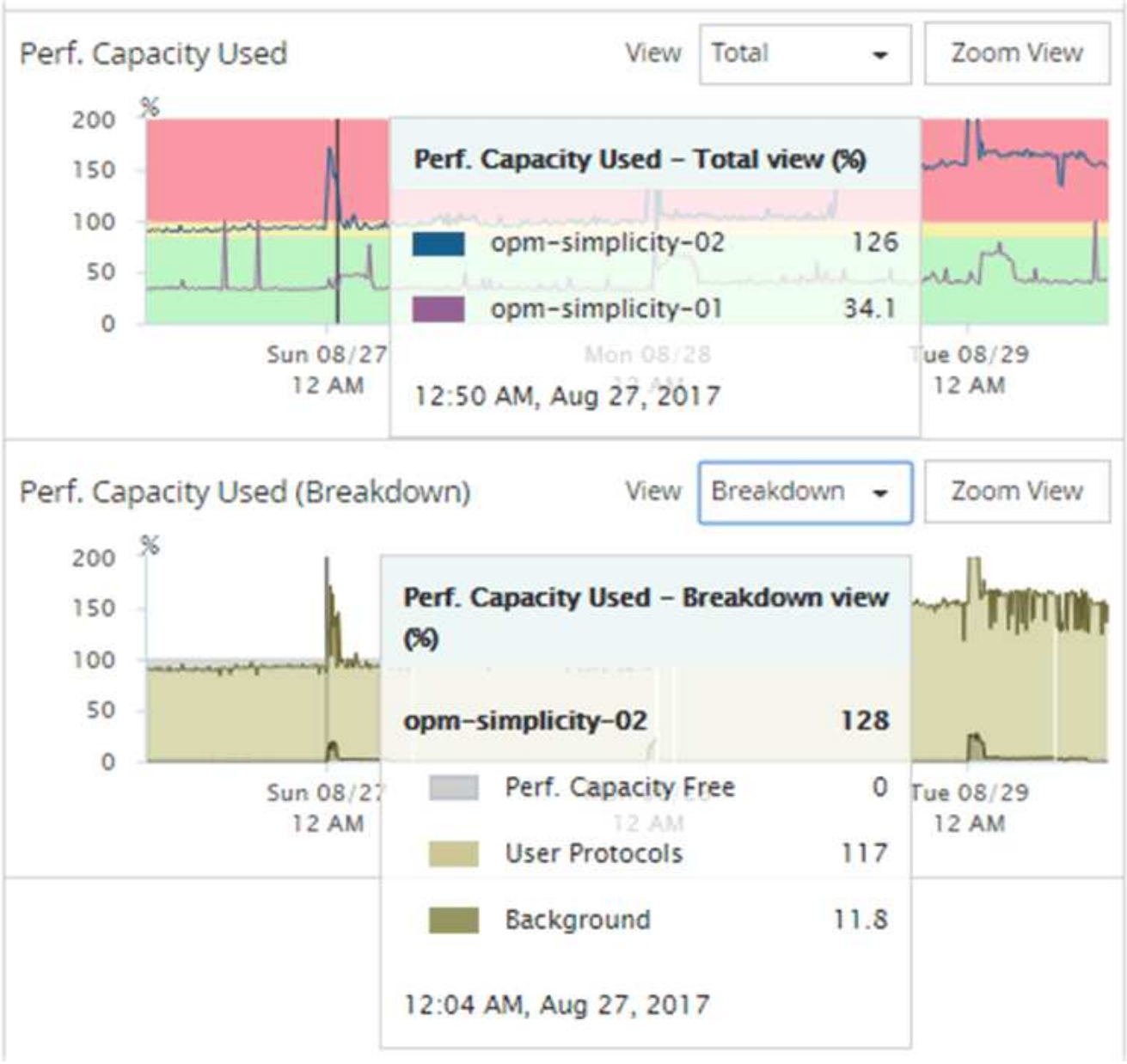
Il grafico standard del contatore visualizza i valori della capacità di performance utilizzata per i nodi o gli aggregati selezionati. Il grafico del contatore dei guasti visualizza i valori di capacità delle performance totali per l'oggetto root separati in base all'utilizzo in base ai protocolli utente rispetto ai processi di sistema in background. Inoltre, viene mostrata anche la quantità di capacità di performance libera.

Poiché alcune attività in background associate alla gestione del sistema e dei dati sono identificate come carichi di lavoro degli utenti e classificate come protocolli utente, la percentuale dei protocolli utente potrebbe apparire artificialmente elevata quando tali processi vengono eseguiti. In genere, questi processi vengono eseguiti intorno alla mezzanotte quando l'utilizzo del cluster è basso. Se si rileva un picco nell'attività del protocollo utente intorno alla mezzanotte, verificare se i processi di backup del cluster o altre attività in background sono configurati per l'esecuzione in quel momento.

Fasi

1. Selezionare la scheda **Explorer** da un nodo o da una pagina **Landing** aggregata.
2. Nel riquadro **Counter Chart**, fare clic su **Choose Chart**, quindi selezionare **Perf. Grafico della capacità utilizzata**.
3. Scorrere verso il basso fino a visualizzare la mappa.

   I colori del grafico standard mostrano quando l'oggetto si trova nell'intervallo ottimale (giallo), quando l'oggetto è sottoutilizzato (verde) e quando l'oggetto è sovrautilizzato (rosso). Il grafico dettagliato mostra i dettagli della capacità delle performance solo per l'oggetto root.



4. Se si desidera visualizzare uno dei grafici in un formato a dimensione intera, fare clic su **Zoom View** (Vista zoom).

   In questo modo è possibile aprire più diagrammi di contatore in una finestra separata per confrontare i valori di capacità delle performance utilizzati con i valori IOPS o Mbps nello stesso intervallo di tempo.

La capacità di performance ha utilizzato le condizioni di soglia delle performance

È possibile creare criteri di soglia delle performance definiti dall'utente in modo che gli eventi vengano attivati quando il valore della capacità di performance utilizzata per un nodo o aggregato supera l'impostazione di soglia definita per la capacità di performance utilizzata.

Inoltre, i nodi possono essere configurati con un criterio di soglia "Performance Capacity used Takeover".

Questo criterio di soglia totalizza la capacità di performance utilizzata dalle statistiche per entrambi i nodi di una coppia ha per determinare se uno dei due nodi non dispone di capacità sufficiente in caso di guasto dell'altro nodo. Poiché il carico di lavoro durante il failover è la combinazione dei carichi di lavoro dei due nodi partner`, la stessa capacità di performance utilizzata per la policy di takeover può essere applicata a entrambi i nodi.

> Questa equivalenza di capacità di performance utilizzata è generalmente vera tra i nodi. Tuttavia, se il traffico tra nodi è significativamente maggiore per uno dei nodi attraverso il partner di failover, la capacità di performance totale utilizzata per l'esecuzione di tutti i carichi di lavoro su un nodo partner rispetto all'altro nodo partner potrebbe essere leggermente diversa a seconda del nodo guasto.

Le condizioni di utilizzo della capacità di performance possono anche essere utilizzate come impostazioni di soglia delle performance secondarie per creare una policy di soglia combinata quando si definiscono le soglie per LUN e volumi. La condizione di capacità di performance utilizzata viene applicata all'aggregato o al nodo su cui risiede il volume o il LUN. Ad esempio, è possibile creare una combinazione di criteri di soglia utilizzando i seguenti criteri:

Oggetto di storage	Contatore delle performance	Soglia di avviso	Soglia critica	Durata
Volume	Latenza	15 ms/op	25 ms/op	20 minuti

I criteri di soglia combinati causano la generazione di un evento solo quando entrambe le condizioni vengono violate per l'intera durata.

Utilizzo della capacità di performance utilizzata per gestire le performance

In genere, le organizzazioni desiderano operare con una percentuale di capacità utilizzata per le performance inferiore a 100, in modo che le risorse vengano utilizzate in modo efficiente, riservando al tempo stesso una capacità di performance aggiuntiva per supportare le richieste di periodi di picco. È possibile utilizzare i criteri di soglia per personalizzare l'invio di avvisi per i valori di capacità utilizzata dalle performance elevate.

Puoi stabilire obiettivi specifici in base ai tuoi requisiti di performance. Ad esempio, le società di servizi finanziari potrebbero riservare una maggiore capacità di performance per garantire la tempestiva esecuzione delle negoziazioni. Queste aziende potrebbero voler impostare le soglie di utilizzo della capacità di performance nell'intervallo del 70-80%. Le aziende manifatturiere con margini inferiori potrebbero scegliere di riservare una capacità di performance inferiore se sono disposte a rischiare le performance per gestire meglio i costi IT. Queste aziende potrebbero impostare le soglie di utilizzo della capacità di performance nell'intervallo del 85-95%.

Quando il valore della capacità di performance utilizzata supera la percentuale impostata in un criterio di soglia definito dall'utente, Unified Manager invia un'email di avviso e aggiunge l'evento alla pagina Event Inventory. Ciò consente di gestire i potenziali problemi prima che influiscano sulle performance. Questi eventi possono anche essere utilizzati come indicatori necessari per spostare i carichi di lavoro e apportare modifiche all'interno dei nodi e degli aggregati.

Informazioni sul copyright

Copyright © 2025 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEQUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina http://www.netapp.com/TM sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.