



Leistungsereignisse und Warnungen verstehen

Active IQ Unified Manager

NetApp
October 15, 2025

Inhalt

Leistungsereignisse und Warnungen verstehen	1
Quellen von Leistungsereignissen	1
Schweregradtypen von Leistungsereignissen	2
Von Unified Manager erkannte Konfigurationsänderungen	2
Arten systemdefinierter Richtlinien für Leistungsschwellenwerte	3
Cluster-Schwellenwertrichtlinien	3
Knotenschwellenwertrichtlinien	4
Richtlinien für aggregierte Schwellenwerte	5
Richtlinien für Workload-Latenzschwellenwerte	5
QoS-Schwellenwertrichtlinien	5
Analyse und Benachrichtigung von Leistungsereignissen	6
Ereignisanalyse	6
Ereignisstatus	7
Ereignisbenachrichtigung	7
Ereignisinteraktion	8
So ermittelt Unified Manager die Leistungsauswirkungen eines Ereignisses	8
Clusterkomponenten und warum sie kontrovers diskutiert werden können	9
Rollen der an einem Leistungsereignis beteiligten Workloads	11

Leistungsereignisse und Warnungen verstehen

Leistungsereignisse sind Vorfälle im Zusammenhang mit der Arbeitslastleistung auf einem Cluster. Sie helfen Ihnen, Arbeitslasten mit langsamen Reaktionszeiten zu identifizieren. Zusammen mit den gleichzeitig aufgetretenen Integritätsereignissen können Sie die Probleme ermitteln, die möglicherweise die langsamen Reaktionszeiten verursacht oder dazu beigetragen haben.

Wenn Unified Manager mehrere Vorkommen derselben Ereignisbedingung für dieselbe Clusterkomponente erkennt, behandelt er alle Vorkommen als ein einzelnes Ereignis und nicht als separate Ereignisse.

Sie können Warnmeldungen so konfigurieren, dass automatisch E-Mail-Benachrichtigungen gesendet werden, wenn Leistungsereignisse bestimmter Schweregrade auftreten.

Quellen von Leistungsereignissen

Leistungsereignisse sind Probleme im Zusammenhang mit der Arbeitslastleistung auf einem Cluster. Sie helfen Ihnen, Speicherobjekte mit langsamen Reaktionszeiten (auch als hohe Latenz bezeichnet) zu identifizieren. Zusammen mit anderen gleichzeitig aufgetretenen Integritätsereignissen können Sie die Probleme ermitteln, die möglicherweise die langsamen Reaktionszeiten verursacht oder dazu beigetragen haben.

Unified Manager empfängt Leistungsereignisse aus den folgenden Quellen:

- **Benutzerdefinierte Leistungsschwellenwert-Richtlinienereignisse**

Leistungsprobleme aufgrund von benutzerdefinierten Schwellenwerten, die Sie festgelegt haben. Sie konfigurieren Richtlinien für Leistungsschwellenwerte für Speicherobjekte, beispielsweise Aggregate und Volumes, sodass Ereignisse generiert werden, wenn ein Schwellenwert für einen Leistungsindikator überschritten wurde.

Sie müssen eine Richtlinie für Leistungsschwellenwerte definieren und sie einem Speicherobjekt zuweisen, um diese Ereignisse zu empfangen.

- **Systemdefinierte Richtlinienereignisse für Leistungsschwellenwerte**

Leistungsprobleme basierend auf systemdefinierten Schwellenwerten. Diese Schwellenwertrichtlinien sind in der Installation von Unified Manager enthalten, um häufige Leistungsprobleme abzudecken.

Diese Schwellenwertrichtlinien sind standardmäßig aktiviert und möglicherweise werden kurz nach dem Hinzufügen eines Clusters Ereignisse angezeigt.

- **Dynamische Leistungsschwellenereignisse**

Leistungsprobleme, die auf Ausfälle oder Fehler in einer IT-Infrastruktur oder auf eine Überbeanspruchung der Clusterressourcen durch Workloads zurückzuführen sind. Die Ursache dieser Ereignisse kann ein einfaches Problem sein, das sich im Laufe der Zeit von selbst behebt oder durch eine Reparatur oder Konfigurationsänderung behoben werden kann. Ein dynamisches Schwellenwertereignis zeigt an, dass die Workloads auf einem ONTAP -System aufgrund anderer Workloads mit hoher Nutzung gemeinsam genutzter Clusterkomponenten langsam sind.

Diese Schwellenwerte sind standardmäßig aktiviert und Sie sehen möglicherweise Ereignisse, nachdem drei Tage lang Daten aus einem neuen Cluster erfasst wurden.

Schweregradtypen von Leistungsereignissen

Jedem Leistungsereignis ist ein Schweregrad zugeordnet, der Ihnen dabei hilft, die Ereignisse zu priorisieren, die sofortige Korrekturmaßnahmen erfordern.

- **Kritisch**

Es ist ein Leistungsereignis aufgetreten, das zu einer Dienstunterbrechung führen kann, wenn nicht sofort Abhilfemaßnahmen ergriffen werden.

Kritische Ereignisse werden nur ab benutzerdefinierten Schwellenwerten gesendet.

- **Warnung**

Ein Leistungsindikator für ein Clusterobjekt liegt außerhalb des normalen Bereichs und sollte überwacht werden, um sicherzustellen, dass er nicht den kritischen Schweregrad erreicht. Ereignisse dieses Schweregrads verursachen keine Dienstunterbrechung und es sind möglicherweise keine sofortigen Korrekturmaßnahmen erforderlich.

Warnereignisse werden von benutzerdefinierten, systemdefinierten oder dynamischen Schwellenwerten gesendet.

- **Information**

Das Ereignis tritt ein, wenn ein neues Objekt erkannt oder eine Benutzeraktion ausgeführt wird. Wenn beispielsweise ein Speicherobjekt gelöscht wird oder Konfigurationsänderungen vorgenommen werden, wird das Ereignis mit dem Schweregrad „Informationen“ generiert.

Informationsereignisse werden direkt von ONTAP gesendet, wenn eine Konfigurationsänderung erkannt wird.

Weitere Informationen finden Sie unter den folgenden Links:

- ["Was passiert, wenn ein Ereignis empfangen wird?"](#)
- ["Welche Informationen sind in einer Warn-E-Mail enthalten?"](#)
- ["Hinzufügen von Warnungen"](#)
- ["Hinzufügen von Warnungen für Leistungsereignisse"](#)

Von Unified Manager erkannte Konfigurationsänderungen

Unified Manager überwacht Ihre Cluster auf Konfigurationsänderungen, damit Sie feststellen können, ob eine Änderung ein Leistungsereignis verursacht oder dazu beigetragen hat. Auf den Seiten des Performance Explorers wird ein Symbol für Änderungsereignisse angezeigt (●), um das Datum und die Uhrzeit anzugeben, wann die Änderung erkannt wurde.

Sie können die Leistungsdiagramme auf den Seiten des Leistungs-Explorers und auf der Seite

„Arbeitslastanalyse“ überprüfen, um zu sehen, ob sich das Änderungsereignis auf die Leistung des ausgewählten Clusterobjekts ausgewirkt hat. Wenn die Änderung ungefähr zur gleichen Zeit wie ein Leistungsereignis erkannt wurde, hat die Änderung möglicherweise zu dem Problem beigetragen, das zur Auslösung der Ereigniswarnung geführt hat.

Unified Manager kann die folgenden Änderungsereignisse erkennen, die als Informationsereignisse kategorisiert werden:

- Ein Volumen wird zwischen Aggregaten verschoben.

Unified Manager kann erkennen, wann die Verschiebung läuft, abgeschlossen ist oder fehlgeschlagen ist. Wenn Unified Manager während einer Volumeverschiebung ausfällt, erkennt es die Volumeverschiebung, wenn es wieder verfügbar ist, und zeigt ein Änderungsereignis dafür an.

- Die Durchsatzgrenze (MB/s oder IOPS) einer QoS-Richtliniengruppe, die eine oder mehrere überwachte Workloads enthält, ändert sich.

Das Ändern eines Richtliniengruppenlimits kann zu zeitweiligen Spitzen in der Latenz (Reaktionszeit) führen, die auch Ereignisse für die Richtliniengruppe auslösen können. Die Latenz normalisiert sich allmählich und alle durch die Spitzen verursachten Ereignisse werden obsolet.

- Ein Knoten in einem HA-Paar übernimmt den Speicher seines Partnerknotens oder gibt ihn zurück.

Unified Manager kann erkennen, wann die Übernahme-, Teilübernahme- oder Rückgabeoperation abgeschlossen ist. Wenn die Übernahme durch einen in Panik geratenen Knoten verursacht wird, erkennt Unified Manager das Ereignis nicht.

- Ein ONTAP -Upgrade- oder Wiederherstellungsvorgang wurde erfolgreich abgeschlossen.

Es werden die vorherige und die neue Version angezeigt.

Arten systemdefinierter Richtlinien für Leistungsschwellenwerte

Unified Manager bietet einige Standard-Schwellenwertrichtlinien, die die Clusterleistung überwachen und automatisch Ereignisse generieren. Diese Richtlinien sind standardmäßig aktiviert und generieren Warn- oder Informationsereignisse, wenn die überwachten Leistungsschwellenwerte überschritten werden.



Systemdefinierte Richtlinien für Leistungsschwellenwerte sind auf Cloud Volumes ONTAP, ONTAP Edge- oder ONTAP Select Systemen nicht aktiviert.

Wenn Sie unnötige Ereignisse von systemdefinierten Richtlinien für Leistungsschwellenwerte erhalten, können Sie die Ereignisse für einzelne Richtlinien auf der Seite „Ereigniseinrichtung“ deaktivieren.

Cluster-Schwellenwertrichtlinien

Die systemdefinierten Richtlinien für Cluster-Leistungsschwellenwerte werden standardmäßig jedem Cluster zugewiesen, der von Unified Manager überwacht wird:

- **Ungleichgewicht der Clusterlast**

Identifiziert Situationen, in denen ein Knoten mit einer viel höheren Last arbeitet als andere Knoten im Cluster und daher möglicherweise die Arbeitslastlatenzen beeinflusst.

Dies geschieht durch Vergleichen des Werts der genutzten Leistungskapazität für alle Knoten in einem Cluster, um festzustellen, ob ein Knoten den Schwellenwert von 30 % länger als 24 Stunden überschritten hat. Dies ist ein Warnereignis.

- **Ungleichgewicht der Clusterkapazität**

Identifiziert Situationen, in denen ein Aggregat eine viel höhere genutzte Kapazität aufweist als andere Aggregate im Cluster und daher möglicherweise den für Vorgänge erforderlichen Speicherplatz beeinträchtigt.

Dies geschieht durch Vergleichen des verwendeten Kapazitätswerts für alle Aggregate im Cluster, um festzustellen, ob zwischen den Aggregaten ein Unterschied von 70 % besteht. Dies ist ein Warnereignis.

Knotenschwellenwertrichtlinien

Die systemdefinierten Richtlinien für Knotenleistungsschwellenwerte werden standardmäßig jedem Knoten in den Clustern zugewiesen, die von Unified Manager überwacht werden:

- **Schwellenwert für genutzte Leistungskapazität überschritten**

Identifiziert Situationen, in denen ein einzelner Knoten über die Grenzen seiner Betriebseffizienz hinaus arbeitet und daher möglicherweise die Arbeitslastlatenzen beeinträchtigt.

Dies geschieht durch die Suche nach Knoten, die länger als 12 Stunden mehr als 100 % ihrer Leistungskapazität nutzen. Dies ist ein Warnereignis.

- **Knoten-HA-Paar überbeansprucht**

Identifiziert Situationen, in denen Knoten in einem HA-Paar über die Grenzen der Betriebseffizienz des HA-Paares hinaus arbeiten.

Dies geschieht durch Betrachtung des Werts der genutzten Leistungskapazität für die beiden Knoten im HA-Paar. Wenn die kombinierte Leistungskapazität der beiden Knoten länger als 12 Stunden 200 % übersteigt, wirkt sich ein Controller-Failover auf die Arbeitslastlatenzen aus. Dies ist eine Informationsveranstaltung.

- **Knoten-Festplattenfragmentierung**

Identifiziert Situationen, in denen eine oder mehrere Festplatten in einem Aggregat fragmentiert sind, wodurch wichtige Systemdienste verlangsamt werden und möglicherweise die Arbeitslastlatenzen auf einem Knoten beeinträchtigt werden.

Dies geschieht durch die Betrachtung bestimmter Lese- und Schreibvorgangsverhältnisse über alle Aggregate auf einem Knoten hinweg. Diese Richtlinie kann auch während der SyncMirror Neusynchronisierung ausgelöst werden oder wenn bei Festplattenbereinigungsvorgängen Fehler gefunden werden. Dies ist ein Warnereignis.



Die Richtlinie „Knoten-Festplattenfragmentierung“ analysiert nur HDD-Aggregate; Flash Pool-, SSD- und FabricPool -Aggregate werden nicht analysiert.

Richtlinien für aggregierte Schwellenwerte

Die systemdefinierte Richtlinie für aggregierte Leistungsschwellenwerte wird standardmäßig jedem Aggregat in den von Unified Manager überwachten Clustern zugewiesen:

- **Aggregate-Festplatten überlastet**

Identifiziert Situationen, in denen ein Aggregat über den Grenzen seiner Betriebseffizienz arbeitet und dadurch möglicherweise die Arbeitslastlatenzen beeinträchtigt. Diese Situationen werden dadurch identifiziert, dass nach Aggregaten gesucht wird, bei denen die Festplatten im Aggregat länger als 30 Minuten zu mehr als 95 % ausgelastet sind. Diese Multicondition-Richtlinie führt dann die folgende Analyse durch, um die Ursache des Problems zu ermitteln:

- Wird auf einer Festplatte im Aggregat derzeit im Hintergrund eine Wartungsaktivität ausgeführt?

Zu den im Hintergrund ausgeführten Wartungsaktivitäten einer Festplatte zählen unter anderem die Festplattenrekonstruktion, die Festplattenbereinigung, die SyncMirror Neusynchronisierung und die Reparatur.

- Gibt es einen Kommunikationsengpass bei der Fibre-Channel-Verbindung des Festplattenregals?
- Ist im Aggregat zu wenig freier Speicherplatz vorhanden? Ein Warnereignis wird für diese Richtlinie nur dann ausgegeben, wenn eine (oder mehrere) der drei untergeordneten Richtlinien ebenfalls als verletzt gelten. Ein Leistungsereignis wird nicht ausgelöst, wenn nur die Festplatten im Aggregat zu mehr als 95 % ausgelastet sind.



Die Richtlinie „Aggregate disks over-utilized“ analysiert reine HDD-Aggregate und Flash Pool-Aggregate (Hybrid); SSD- und FabricPool -Aggregate werden nicht analysiert.

Richtlinien für Workload-Latenzschwellenwerte

Die systemdefinierten Richtlinien für die Schwellenwerte für die Arbeitslastlatenz werden jeder Arbeitslast zugewiesen, für die eine konfigurierte Leistungsservicelevelrichtlinie mit einem definierten Wert für die „erwartete Latenz“ vorliegt:

- **Workload-Volumen/LUN-Latenzschwelle überschritten, wie durch Leistungs-Servicelevel definiert**

Identifiziert Volumes (Dateifreigaben) und LUNs, die ihr Limit für die „erwartete Latenz“ überschritten haben und die die Workload-Leistung beeinträchtigen. Dies ist ein Warnereignis.

Dies geschieht durch die Suche nach Arbeitslasten, die in der letzten Stunde 30 % der Zeit den erwarteten Latenzwert überschritten haben.

QoS-Schwellenwertrichtlinien

Die systemdefinierten QoS-Leistungsschwellenwertrichtlinien werden allen Workloads zugewiesen, für die eine ONTAP QoS-Richtlinie für den maximalen Durchsatz (IOPS, IOPS/TB oder MB/s) konfiguriert ist. Unified Manager löst ein Ereignis aus, wenn der Workload-Durchsatzwert 15 % unter dem konfigurierten QoS-Wert liegt:

- **QoS Max IOPS oder MB/s-Schwellenwert**

Identifiziert Volumes und LUNs, die ihr QoS-Maximum an IOPS oder MB/s-Durchsatzlimit überschritten haben und die die Arbeitslastlatenz beeinträchtigen. Dies ist ein Warnereignis.

Wenn einer Richtliniengruppe eine einzelne Arbeitslast zugewiesen wird, wird nach Arbeitslasten gesucht, die den in der zugewiesenen QoS-Richtliniengruppe definierten maximalen Durchsatzschwellenwert während jedes Erfassungszeitraums der vorherigen Stunde überschritten haben.

Wenn mehrere Workloads eine einzige QoS-Richtlinie gemeinsam nutzen, geschieht dies, indem die IOPS oder MB/s aller Workloads in der Richtlinie addiert und diese Summe mit dem Schwellenwert verglichen wird.

- **QoS Peak IOPS/TB oder IOPS/TB mit Blockgrößenschwelle**

Identifiziert Volumes, die ihr adaptives QoS-Spitzen-IOPS/TB-Durchsatzlimit (oder IOPS/TB mit Blockgrößenlimit) überschritten haben und die die Arbeitslastlatenz beeinträchtigen. Dies ist ein Warnereignis.

Dies geschieht durch die Umwandlung des in der adaptiven QoS-Richtlinie definierten IOPS/TB-Spitzenschwellenwerts in einen QoS-Maximal-IOPS-Wert basierend auf der Größe jedes Volumes. Anschließend wird nach Volumes gesucht, die den QoS-Maximal-IOPS-Wert während jedes Leistungserfassungszeitraums der vorherigen Stunde überschritten haben.



Diese Richtlinie wird nur auf Volumes angewendet, wenn der Cluster mit ONTAP 9.3 oder einer späteren Software installiert ist.

Wenn das Element „Blockgröße“ in der adaptiven QoS-Richtlinie definiert wurde, wird der Schwellenwert basierend auf der Größe jedes Volumes in einen maximalen QoS-MB/s-Wert umgewandelt. Anschließend wird nach Volumes gesucht, die während jedes Leistungserfassungszeitraums der vorherigen Stunde den QoS-Maximum-MB/s-Wert überschritten haben.



Diese Richtlinie wird nur auf Volumes angewendet, wenn der Cluster mit ONTAP 9.5 oder einer späteren Software installiert ist.

Analyse und Benachrichtigung von Leistungsereignissen

Leistungsereignisse benachrichtigen Sie über E/A-Leistungsprobleme bei einer Arbeitslast, die durch Konflikte bei einer Clusterkomponente verursacht werden. Unified Manager analysiert das Ereignis, um alle beteiligten Workloads und die umstrittene Komponente zu identifizieren und festzustellen, ob das Ereignis immer noch ein Problem darstellt, das Sie möglicherweise lösen müssen.

Unified Manager überwacht die E/A-Latenz (Reaktionszeit) und IOPS (Operationen) für Volumes in einem Cluster. Wenn beispielsweise andere Workloads eine Clusterkomponente übermäßig beanspruchen, ist die Komponente überlastet und kann die optimale Leistung nicht erbringen, um die Workload-Anforderungen zu erfüllen. Die Leistung anderer Workloads, die dieselbe Komponente verwenden, kann beeinträchtigt werden, was zu einer Erhöhung ihrer Latenzen führt. Wenn die Latenz den dynamischen Leistungsschwellenwert überschreitet, löst Unified Manager ein Leistungsereignis aus, um Sie zu benachrichtigen.

Ereignisanalyse

Unified Manager führt die folgenden Analysen durch und verwendet dabei die Leistungsstatistiken der letzten 15 Tage, um die Opfer-Workloads, die Bully-Workloads und die an einem Ereignis beteiligte Clusterkomponente zu identifizieren:

- Identifiziert betroffene Workloads, deren Latenz den dynamischen Leistungsschwellenwert überschritten hat, der die Obergrenze der Latenzprognose darstellt:
 - Bei Volumes auf HDD- oder Flash Pool-Hybridaggregaten (lokale Ebene) werden Ereignisse nur ausgelöst, wenn die Latenz größer als 5 Millisekunden (ms) ist und die IOPS mehr als 10 Vorgänge pro Sekunde (Ops/Sek.) betragen.
 - Bei Volumes auf reinen SSD-Aggregaten oder FabricPool Aggregaten (Cloud-Ebene) werden Ereignisse nur ausgelöst, wenn die Latenz größer als 1 ms und die IOPS über 100 Ops/Sek. liegen.
- Identifiziert die konkurrierende Clusterkomponente.



Wenn die Latenz der Opfer-Workloads an der Cluster-Verbindung größer als 1 ms ist, behandelt Unified Manager dies als signifikant und löst ein Ereignis für die Cluster-Verbindung aus.

- Identifiziert die Bully-Workloads, die die Clusterkomponente überbeanspruchen und zu Konflikten führen.
- Ordnet die beteiligten Arbeitslasten anhand ihrer Abweichung bei der Nutzung oder Aktivität einer Clusterkomponente ein, um zu ermitteln, bei welchen Tätern die größten Änderungen bei der Nutzung der Clusterkomponente auftreten und welche Opfer am stärksten betroffen sind.

Ein Ereignis kann nur für einen kurzen Moment auftreten und sich dann selbst korrigieren, nachdem die verwendete Komponente nicht mehr im Konflikt steht. Ein kontinuierliches Ereignis ist ein Ereignis, das für dieselbe Clusterkomponente innerhalb eines Fünf-Minuten-Intervalls erneut auftritt und im aktiven Zustand bleibt. Bei kontinuierlichen Ereignissen löst Unified Manager einen Alarm aus, nachdem das gleiche Ereignis während zwei aufeinanderfolgender Analyseintervalle erkannt wurde.

Wenn ein Ereignis behoben ist, bleibt es im Unified Manager als Teil der Aufzeichnung früherer Leistungsprobleme für ein Volume verfügbar. Jedes Ereignis verfügt über eine eindeutige ID, die den Ereignistyp und die beteiligten Volumes, Cluster und Clusterkomponenten identifiziert.



Ein einzelnes Volume kann gleichzeitig an mehreren Ereignissen beteiligt sein.

Ereignisstatus

Ereignisse können sich in einem der folgenden Zustände befinden:

• Aktiv

Zeigt an, dass das Leistungsereignis derzeit aktiv ist (neu oder bestätigt). Das Problem, das das Ereignis verursacht hat, hat sich nicht von selbst behoben oder wurde nicht gelöst. Der Leistungsindikator für das Speicherobjekt bleibt über dem Leistungsschwellenwert.

• Veraltet

Zeigt an, dass das Ereignis nicht mehr aktiv ist. Das Problem, das das Ereignis verursacht hat, hat sich von selbst behoben oder wurde gelöst. Der Leistungsindikator für das Speicherobjekt liegt nicht mehr über dem Leistungsschwellenwert.

Ereignisbenachrichtigung

Die Ereignisse werden auf der Dashboard-Seite und auf vielen anderen Seiten der Benutzeroberfläche angezeigt und Benachrichtigungen zu diesen Ereignissen werden an angegebene E-Mail-Adressen gesendet. Auf der Seite „Ereignisdetails“ und auf der Seite „Arbeitslastanalyse“ können Sie detaillierte

Analyseinformationen zu einem Ereignis anzeigen und Vorschläge zur Lösung des Ereignisses erhalten.

Ereignisinteraktion

Auf der Seite „Ereignisdetails“ und auf der Seite „Arbeitslastanalyse“ können Sie auf folgende Weise mit Ereignissen interagieren:

- Wenn Sie die Maus über ein Ereignis bewegen, wird eine Meldung mit Datum und Uhrzeit der Ereigniserkennung angezeigt.

Wenn für denselben Zeitraum mehrere Ereignisse vorliegen, zeigt die Meldung die Anzahl der Ereignisse an.

- Wenn Sie auf ein einzelnes Ereignis klicken, wird ein Dialogfeld mit ausführlicheren Informationen zu dem Ereignis angezeigt, einschließlich der beteiligten Clusterkomponenten.

Die umstrittene Komponente ist eingekreist und rot hervorgehoben. Sie können auf **Vollständige Analyse anzeigen** klicken, um die vollständige Analyse auf der Seite mit den Ereignisdetails anzuzeigen. Wenn für denselben Zeitraum mehrere Ereignisse vorliegen, werden im Dialogfeld Details zu den drei aktuellsten Ereignissen angezeigt. Sie können auf ein Ereignis klicken, um die Ereignisanalyse auf der Seite „Ereignisdetails“ anzuzeigen.

So ermittelt Unified Manager die Leistungsauswirkungen eines Ereignisses

Unified Manager verwendet die Abweichung bei Aktivität, Auslastung, Schreibdurchsatz, Clusterkomponentennutzung oder E/A-Latenz (Reaktionszeit) für eine Arbeitslast, um das Ausmaß der Auswirkungen auf die Arbeitslastleistung zu bestimmen. Diese Informationen bestimmen die Rolle der einzelnen Arbeitslasten im Ereignis und ihre Rangfolge auf der Seite mit den Ereignisdetails.

Unified Manager vergleicht die zuletzt analysierten Werte für eine Arbeitslast mit dem erwarteten Wertebereich (Latenzprognose). Die Differenz zwischen den zuletzt analysierten Werten und dem erwarteten Wertebereich identifiziert die Workloads, deren Leistung durch das Ereignis am stärksten beeinträchtigt wurde.

Angenommen, ein Cluster enthält zwei Workloads: Workload A und Workload B. Die Latenzprognose für Workload A beträgt 5–10 Millisekunden pro Vorgang (ms/Operation) und die tatsächliche Latenz liegt normalerweise bei etwa 7 ms/Operation. Die Latenzprognose für Workload B beträgt 10–20 ms/Operation und die tatsächliche Latenz liegt normalerweise bei etwa 15 ms/Operation. Beide Workloads liegen deutlich innerhalb ihrer Latenzprognose. Aufgrund von Konflikten im Cluster erhöht sich die Latenz beider Workloads auf 40 ms/Operation, wodurch der dynamische Leistungsschwellenwert, der die Obergrenze der Latenzprognose darstellt, überschritten wird und Ereignisse ausgelöst werden. Die Abweichung der Latenz von den erwarteten Werten zu den Werten über dem Leistungsschwellenwert beträgt für Workload A etwa 33 ms/Op, und die Abweichung für Workload B beträgt etwa 25 ms/Op. Die Latenz beider Workloads steigt auf 40 ms/op, Workload A hatte jedoch die größeren Auswirkungen auf die Leistung, da sie mit 33 ms/op die höhere Latenzabweichung aufwies.

Auf der Seite „Ereignisdetails“ im Abschnitt „Systemdiagnose“ können Sie Workloads nach ihrer Abweichung bei Aktivität, Auslastung oder Durchsatz für eine Clusterkomponente sortieren. Sie können Workloads auch nach Latenz sortieren. Wenn Sie eine Sortieroption auswählen, analysiert Unified Manager die Abweichung bei Aktivität, Auslastung, Durchsatz oder Latenz seit der Erkennung des Ereignisses von den erwarteten Werten, um die Sortierreihenfolge der Arbeitslast zu bestimmen. Für die Latenz stehen die roten Punkte (●) weisen

auf das Überschreiten einer Leistungsschwelle durch eine betroffene Arbeitslast und die daraus resultierenden Auswirkungen auf die Latenz hin. Jeder rote Punkt zeigt eine höhere Abweichung bei der Latenz an, wodurch Sie die betroffenen Workloads identifizieren können, deren Latenz durch ein Ereignis am stärksten beeinträchtigt wurde.

Clusterkomponenten und warum sie kontrovers diskutiert werden können

Sie können Leistungsprobleme im Cluster erkennen, wenn eine Clusterkomponente in einen Konflikt gerät. Die Leistung von Workloads, die die Komponente verwenden, verlangsamt sich und ihre Antwortzeit (Latenz) für Clientanforderungen erhöht sich, was ein Ereignis in Unified Manager auslöst.

Eine Komponente, die im Streit liegt, kann nicht die optimale Leistung erbringen. Seine Leistung hat nachgelassen und die Leistung anderer Clusterkomponenten und Workloads, sogenannter *Opfer*, hat möglicherweise die Latenz erhöht. Um eine Komponente aus dem Wettbewerb zu nehmen, müssen Sie ihre Arbeitslast reduzieren oder ihre Fähigkeit erhöhen, mehr Arbeit zu bewältigen, damit die Leistung wieder auf ein normales Niveau zurückkehren kann. Da Unified Manager die Workload-Leistung in Fünf-Minuten-Intervallen erfasst und analysiert, erkennt er nur, wenn eine Clusterkomponente dauerhaft überbeansprucht wird. Vorübergehende Überlastungsspitzen, die innerhalb des Fünf-Minuten-Intervalls nur von kurzer Dauer sind, werden nicht erkannt.

Beispielsweise kann es zu einem Konflikt um ein Speicheraggregat kommen, weil eine oder mehrere Workloads darauf um die Erfüllung ihrer E/A-Anforderungen konkurrieren. Andere Arbeitslasten im Aggregat können beeinträchtigt werden, was zu einer Leistungsminderung führt. Um die Aktivität auf dem Aggregat zu reduzieren, können Sie verschiedene Schritte unternehmen, z. B. eine oder mehrere Arbeitslasten auf ein weniger ausgelastetes Aggregat oder einen weniger ausgelasteten Knoten verschieben, um die Gesamtarbeitslastanforderung an das aktuelle Aggregat zu verringern. Für eine QoS-Richtliniengruppe können Sie die Durchsatzgrenze anpassen oder Workloads in eine andere Richtliniengruppe verschieben, sodass die Workloads nicht mehr gedrosselt werden.

Unified Manager überwacht die folgenden Clusterkomponenten, um Sie zu warnen, wenn es zu Konflikten kommt:

- **Netzwerk**

Stellt die Wartezeit von E/A-Anfragen durch die externen Netzwerkprotokolle im Cluster dar. Die Wartezeit ist die Zeit, die darauf gewartet wird, dass die „Transfer Ready“-Transaktionen abgeschlossen werden, bevor der Cluster auf eine E/A-Anforderung antworten kann. Wenn es zu Konflikten mit der Netzwerkkomponente kommt, bedeutet dies, dass eine hohe Wartezeit auf der Protokollebene die Latenz einer oder mehrerer Workloads beeinträchtigt.

- **Netzwerkverarbeitung**

Stellt die Softwarekomponente im Cluster dar, die an der E/A-Verarbeitung zwischen der Protokollschicht und dem Cluster beteiligt ist. Der Knoten, der die Netzwerkverarbeitung durchführt, hat sich möglicherweise seit der Erkennung des Ereignisses geändert. Wenn es bei der Netzwerkverarbeitungskomponente zu Konflikten kommt, bedeutet dies, dass eine hohe Auslastung des Netzwerkverarbeitungsknotens die Latenz einer oder mehrerer Arbeitslasten beeinträchtigt.

Wenn Sie einen All-SAN-Array-Cluster in einer Aktiv-Aktiv-Konfiguration verwenden, wird der Latenzwert der Netzwerkverarbeitung für beide Knoten angezeigt, sodass Sie überprüfen können, ob die Knoten die Last gleichmäßig teilen.

- **QoS-Limit Max**

Stellt die maximale Durchsatzeinstellung (Spitzenwert) der der Arbeitslast zugewiesenen Speicher-QoS-Richtliniengruppe (Quality of Service) dar. Wenn die Richtliniengruppenkomponente im Konflikt steht, bedeutet dies, dass alle Workloads in der Richtliniengruppe durch die festgelegte Durchsatzgrenze gedrosselt werden, was sich auf die Latenz eines oder mehrerer dieser Workloads auswirkt.

- **Mindest-QoS-Limit**

Stellt die Latenz einer Arbeitslast dar, die durch die anderen Arbeitslasten zugewiesene Einstellung für den minimalen (erwarteten) QoS-Durchsatz verursacht wird. Wenn das für bestimmte Workloads festgelegte QoS-Minimum den Großteil der Bandbreite nutzt, um den versprochenen Durchsatz zu garantieren, werden andere Workloads gedrosselt und weisen eine höhere Latenz auf.

- **Cluster-Verbindung**

Stellt die Kabel und Adapter dar, mit denen Clusterknoten physisch verbunden sind. Wenn es bei der Cluster-Interconnect-Komponente zu Konflikten kommt, bedeutet dies, dass die lange Wartezeit für E/A-Anfragen bei der Cluster-Interconnect-Komponente die Latenz einer oder mehrerer Workloads beeinträchtigt.

- * Data Processing*

Stellt die Softwarekomponente im Cluster dar, die an der E/A-Verarbeitung zwischen dem Cluster und dem Speicheraggregat beteiligt ist, das die Arbeitslast enthält. Der Knoten, der die Datenverarbeitung durchführt, hat sich möglicherweise seit der Erkennung des Ereignisses geändert. Wenn es bei der Datenverarbeitungskomponente zu Konflikten kommt, bedeutet dies, dass eine hohe Auslastung des Datenverarbeitungsknotens die Latenz einer oder mehrerer Arbeitslasten beeinträchtigt.

- **Lautstärkeaktivierung**

Stellt den Prozess dar, der die Nutzung aller aktiven Volumes verfolgt. In großen Umgebungen mit mehr als 1.000 aktiven Volumes verfolgt dieser Prozess, wie viele kritische Volumes gleichzeitig über den Knoten auf Ressourcen zugreifen müssen. Wenn die Anzahl gleichzeitig aktiver Volumes den empfohlenen Höchstschwellenwert überschreitet, kommt es bei einigen der nicht kritischen Volumes zu Latenzen, wie hier angegeben.

- * MetroCluster -Ressourcen*

Stellt die MetroCluster Ressourcen dar, einschließlich NVRAM und Interswitch-Links (ISLs), die zum Spiegeln von Daten zwischen Clustern in einer MetroCluster -Konfiguration verwendet werden. Wenn die MetroCluster Komponente im Konflikt steht, bedeutet dies, dass ein hoher Schreibdurchsatz von Workloads auf dem lokalen Cluster oder ein Link-Health-Problem die Latenz einer oder mehrerer Workloads auf dem lokalen Cluster beeinträchtigt. Wenn sich der Cluster nicht in einer MetroCluster -Konfiguration befindet, wird dieses Symbol nicht angezeigt.

- **Aggregate- oder SSD-Aggregate-Operationen**

Stellt das Speicheraggregat dar, auf dem die Workloads ausgeführt werden. Wenn es zu Konflikten bei der Aggregatkomponente kommt, bedeutet dies, dass eine hohe Auslastung des Aggregats die Latenz einer oder mehrerer Workloads beeinträchtigt. Ein Aggregat besteht aus allen HDDs oder einer Mischung aus HDDs und SSDs (ein Flash Pool-Aggregat) oder einer Mischung aus HDDs und einer Cloud-Ebene (ein FabricPool -Aggregat). Ein „SSD-Aggregat“ besteht aus allen SSDs (einem All-Flash-Aggregat) oder einer Mischung aus SSDs und einer Cloud-Ebene (einem FabricPool -Aggregat).

- **Cloud-Latenz**

Stellt die Softwarekomponente im Cluster dar, die an der E/A-Verarbeitung zwischen dem Cluster und der Cloud-Ebene beteiligt ist, auf der Benutzerdaten gespeichert sind. Wenn die Cloud-Latenzkomponente in Konflikt steht, bedeutet dies, dass eine große Anzahl von Lesevorgängen von Volumes, die auf der Cloud-Ebene gehostet werden, die Latenz einer oder mehrerer Workloads beeinträchtigt.

- * SnapMirror synchronisieren*

Stellt die Softwarekomponente im Cluster dar, die in einer synchronen SnapMirror -Beziehung an der Replikation von Benutzerdaten vom primären Volume auf das sekundäre Volume beteiligt ist. Wenn es bei der Synchronisierungskomponente von SnapMirror zu Konflikten kommt, bedeutet dies, dass die Aktivität der synchronen SnapMirror -Vorgänge die Latenz einer oder mehrerer Arbeitslasten beeinträchtigt.

Rollen der an einem Leistungsereignis beteiligten Workloads

Unified Manager verwendet Rollen, um die Beteiligung einer Arbeitslast an einem Leistungsereignis zu identifizieren. Zu den Rollen gehören Opfer, Tyrannen und Haie. Eine benutzerdefinierte Arbeitslast kann gleichzeitig Opfer, Tyrann und Hai sein.

Rolle	Beschreibung
Opfer	Eine benutzerdefinierte Arbeitslast, deren Leistung aufgrund anderer Arbeitslasten, sogenannter „Bullies“, die eine Clusterkomponente übermäßig beanspruchen, abgenommen hat. Als Opfer werden nur benutzerdefinierte Workloads identifiziert. Unified Manager identifiziert die Arbeitslast der Opfer anhand ihrer Latenzabweichung, wobei die tatsächliche Latenz während eines Ereignisses im Vergleich zur Latenzprognose (erwarteter Bereich) stark angestiegen ist.
Schikanieren	Eine benutzerdefinierte oder systemdefinierte Arbeitslast, deren übermäßige Nutzung einer Clusterkomponente zu einer Leistungsminderung anderer Arbeitslasten, sogenannter Opfer, geführt hat. Unified Manager identifiziert problematische Workloads anhand ihrer Abweichung bei der Nutzung einer Clusterkomponente, wobei die tatsächliche Nutzung während eines Ereignisses stark über den erwarteten Nutzungsbereich hinausgeht.
Hai	Eine benutzerdefinierte Arbeitslast mit der höchsten Nutzung einer Clusterkomponente im Vergleich zu allen an einem Ereignis beteiligten Arbeitslasten. Unified Manager identifiziert Shark-Workloads basierend auf ihrer Nutzung einer Clusterkomponente während eines Ereignisses.

Workloads auf einem Cluster können viele Clusterkomponenten gemeinsam nutzen, beispielsweise Aggregate und die CPU für die Netzwerk- und Datenverarbeitung. Wenn eine Arbeitslast, beispielsweise ein Volume, die Nutzung einer Clusterkomponente so weit erhöht, dass die Komponente die Arbeitslastanforderungen nicht mehr effizient erfüllen kann, liegt ein Konflikt mit der Komponente vor. Die Arbeitslast, die eine Clusterkomponente überbeansprucht, ist ein Tyrann. Die anderen Workloads, die diese Komponenten gemeinsam nutzen und deren Leistung durch den Mobber beeinträchtigt wird, sind die Opfer. Auch Aktivitäten von systemdefinierten Workloads, wie etwa Deduplizierung oder Snapshot-Kopien, können zu „Mobbing“ eskalieren.

Wenn Unified Manager ein Ereignis erkennt, identifiziert es alle beteiligten Workloads und Clusterkomponenten, einschließlich der Bully-Workloads, die das Ereignis verursacht haben, der Clusterkomponente, um die es geht, und der Opfer-Workloads, deren Leistung aufgrund der erhöhten Aktivität der Bully-Workloads abgenommen hat.



Wenn Unified Manager die schädlichen Workloads nicht identifizieren kann, gibt er nur Warnmeldungen zu den betroffenen Workloads und der betroffenen Clusterkomponente aus.

Unified Manager kann Arbeitslasten identifizieren, die Opfer von Mobbing-Arbeitslasten sind, und auch erkennen, wann diese Arbeitslasten zu Mobbing-Arbeitslasten werden. Eine Arbeitsbelastung kann sich selbst zum Tyrannen machen. Beispielsweise führt eine Hochleistungs-Workload, die durch eine Richtliniengruppenbeschränkung gedrosselt wird, dazu, dass alle Workloads in der Richtliniengruppe gedrosselt werden, einschließlich der eigenen Workloads. Eine Arbeitsbelastung, die in einem laufenden Leistungsereignis als Tyrann oder Opfer auftritt, kann ihre Rolle ändern oder nicht länger an dem Ereignis teilnehmen.

Copyright-Informationen

Copyright © 2025 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtinhabers 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, BEISPIELHAFT 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 IRGEND EINER 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.