



Quais são os eventos de desempenho

OnCommand Unified Manager 9.5

NetApp
October 23, 2024

Índice

- Quais são os eventos de desempenho 1
 - Análise e notificação de eventos de performance 1
 - Como o Unified Manager determina o impacto no desempenho de um evento 3
 - Componentes do cluster e por que eles podem estar na contenção..... 4
 - Funções dos workloads envolvidos em um evento de desempenho..... 5

Quais são os eventos de desempenho

Os eventos de desempenho são incidentes relacionados ao desempenho da carga de trabalho em um cluster. Eles ajudam a identificar workloads com tempos de resposta lentos. Juntamente com eventos de saúde que ocorreram ao mesmo tempo, você pode determinar os problemas que podem ter causado ou contribuído para os tempos de resposta lentos.

Quando o Unified Manager detecta várias ocorrências da mesma condição de evento para o mesmo componente de cluster, ele trata todas as ocorrências como um único evento, não como eventos separados.

Análise e notificação de eventos de performance

Os eventos de desempenho notificam você sobre problemas de desempenho de e/S em um workload de volume causado pela contenção em um componente do cluster. O Unified Manager analisa o evento para identificar todos os workloads envolvidos, o componente em contenção e se o evento ainda é um problema que talvez você precise resolver.

O Unified Manager monitora a latência de e/S (tempo de resposta) e IOPS (operações) de volumes em um cluster. Quando outras cargas de trabalho usam excessivamente um componente de cluster, por exemplo, o componente está na contenção e não pode ter desempenho em um nível ideal para atender às demandas de workload. O desempenho de outros workloads que estão usando o mesmo componente pode ser afetado, causando o aumento de suas latências. Se a latência ultrapassar o limite de desempenho, o Unified Manager acionará um evento de desempenho e enviará um alerta por e-mail para notificá-lo.

Análise de eventos

O Unified Manager realiza as seguintes análises, usando as estatísticas de desempenho dos 15 dias anteriores, para identificar os workloads da vítima, os workloads bully e o componente do cluster envolvido em um evento:

- Identifica cargas de trabalho da vítima cuja latência ultrapassou o limite de desempenho, que é o limite superior do intervalo esperado:
 - Para volumes em agregados de HDD ou Flash Pool (híbridos), os eventos são acionados somente quando a latência é maior que 5 milissegundos (ms) e o IOPS é mais de 10 operações por segundo (operações/seg).
 - Para volumes em agregados all-SSD ou agregados FabricPool (compostos), os eventos são acionados apenas quando a latência é superior a 1 ms e o IOPS é superior a 100 operações/segundo
- Identifica o componente do cluster na contenção.



Se a latência das cargas de trabalho da vítima na interconexão de cluster for superior a 1 ms, o Unified Manager tratará isso como significativo e acionará um evento para a interconexão de cluster.

- Identifica as cargas de trabalho bully que estão sobreusando o componente do cluster e fazendo com que ele esteja na contenção.
- Classifica as cargas de trabalho envolvidas, com base em seu desvio na utilização ou atividade de um

componente de cluster, para determinar quais bullies têm a maior alteração no uso do componente de cluster e quais vítimas são as mais impactadas.

Um evento pode ocorrer por apenas um breve momento e depois se corrigir depois que o componente que está usando não está mais em disputa. Um evento contínuo é aquele que ocorre novamente para o mesmo componente do cluster dentro de um intervalo de cinco minutos e permanece no estado ativo. Para eventos contínuos, o Unified Manager aciona um alerta após detectar o mesmo evento durante dois intervalos de análise consecutivos. Os eventos que permanecem não resolvidos, que têm um estado de novo, podem exibir mensagens de descrição diferentes como cargas de trabalho envolvidas na alteração do evento.

Quando um evento é resolvido, ele permanece disponível no Unified Manager como parte do Registro de problemas de desempenho anteriores de um volume. Cada evento tem um ID exclusivo que identifica o tipo de evento e os volumes, o cluster e os componentes do cluster envolvidos.



Um único volume pode ser envolvido em mais de um evento ao mesmo tempo.

Estado do evento

Os eventos podem estar em um dos seguintes estados:

- **Ativo**

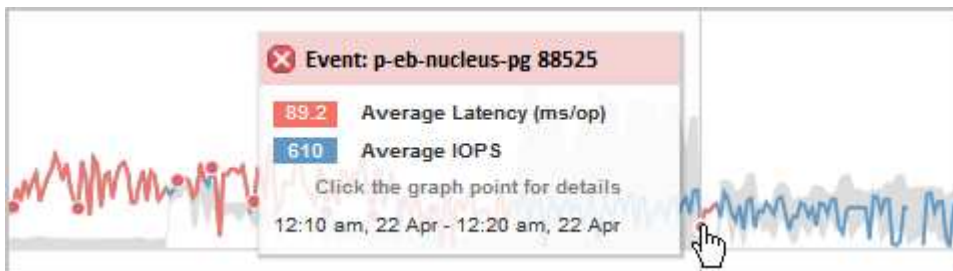
Indica que o evento de desempenho está ativo no momento (novo ou confirmado). O problema que causa o evento não foi corrigido ou não foi resolvido. O contador de performance do objeto de storage permanece acima do limite de performance.

- **Obsoleto**

Indica que o evento já não está ativo. O problema que causa o evento foi corrigido ou foi resolvido. O contador de desempenho do objeto de storage não está mais acima do limite de desempenho.

Notificação de evento

Os alertas de eventos são exibidos na página Dashboards/Overview, Dashboards/Performance, Performance/volume Details e são enviados para endereços de e-mail especificados. Você pode visualizar informações detalhadas de análise sobre um evento e obter sugestões para resolvê-lo na página de detalhes do evento.



Neste exemplo, um evento é indicado por um ponto vermelho (●) no gráfico de latência na página Detalhes de desempenho/volume. Passar o cursor do Mouse sobre o ponto vermelho exibe um popup com mais detalhes sobre o evento e opções para analisá-lo.

Interação de eventos

Na página Detalhes de desempenho/volume, você pode interagir com eventos das seguintes maneiras:

- Mover o ponteiro sobre um ponto vermelho exibe uma mensagem que mostra o ID do evento, juntamente com a latência, o número de operações por segundo e a data e hora em que o evento foi detectado.

Se houver vários eventos para o mesmo período de tempo, a mensagem mostrará o número de eventos, juntamente com a latência média e as operações por segundo para o volume.

- Clicar em um único evento exibe uma caixa de diálogo que mostra informações mais detalhadas sobre o evento, incluindo os componentes do cluster envolvidos, semelhante à seção Resumo na página Detalhes do evento.

O componente em contenção é circulado e realçado a vermelho. Você pode clicar no ID do evento ou em **Exibir análise completa** para visualizar a análise completa na página de detalhes do evento. Se houver vários eventos para o mesmo período de tempo, a caixa de diálogo mostra detalhes sobre os três eventos mais recentes. Você pode clicar em um ID de evento para exibir a análise de evento na página de detalhes do evento. Se houver mais de três eventos para o mesmo período de tempo, clicar no ponto vermelho não exibirá a caixa de diálogo.

Como o Unified Manager determina o impacto no desempenho de um evento

O Unified Manager usa o desvio na atividade, utilização, taxa de transferência de gravação, uso de componentes do cluster ou latência de e/S (tempo de resposta) para um workload a fim de determinar o nível de impacto na performance do workload. Essas informações determinam a função de cada carga de trabalho no evento e como eles são classificados na página de detalhes do evento.

O Unified Manager compara os últimos valores analisados para uma carga de trabalho com o intervalo esperado de valores. A diferença entre os valores analisados pela última vez e o intervalo esperado de valores identifica as cargas de trabalho cujo desempenho foi mais impactado pelo evento.

Por exemplo, suponha que um cluster contenha duas cargas de trabalho: Carga de Trabalho A e carga de trabalho B. o intervalo esperado para carga de trabalho A é de 5-10 milissegundos por operação (ms/op) e sua latência real geralmente é de cerca de 7 ms/op. O intervalo esperado para a carga de trabalho B é de 10-20 ms/op e sua latência real geralmente é de cerca de 15 ms/op. Ambos os workloads estão bem dentro do intervalo esperado para latência. Devido à contenção no cluster, a latência de ambas as cargas de trabalho aumenta para 40 ms/op, cruzando o limite de desempenho, que é os limites superiores do intervalo esperado e acionando eventos. O desvio de latência, dos valores esperados para os valores acima do limite de desempenho, para a carga de trabalho A é de cerca de 33 ms/op, e o desvio para carga de trabalho B é de cerca de 25 ms/op. A latência de ambas as cargas de trabalho aumenta para 40 ms/op, mas o Workload A teve maior impacto no desempenho, pois teve maior desvio de latência em 33 ms/op.

Na página de detalhes do evento, na seção Diagnóstico do sistema, você pode classificar as cargas de trabalho por seu desvio na atividade, utilização ou taxa de transferência de um componente do cluster. Você também pode classificar workloads por latência. Quando você seleciona uma opção de classificação, o Unified Manager analisa o desvio na atividade, utilização, taxa de transferência ou latência desde que o evento foi detectado a partir dos valores esperados para determinar a ordem de classificação da carga de trabalho. Para a latência, os pontos vermelhos (●) indicam um cruzamento de limite de desempenho por uma carga de trabalho da vítima e o impactos subsequente na latência. Cada ponto vermelho indica um nível mais alto de

desvio na latência, o que ajuda a identificar as cargas de trabalho da vítima cuja latência foi mais afetada por um evento.

Componentes do cluster e por que eles podem estar na contenção

Você pode identificar problemas de desempenho do cluster quando um componente do cluster entra em contenção. O desempenho de workloads de volume que usam o componente diminui e seu tempo de resposta (latência) para solicitações de clientes aumenta, o que aciona um evento no Unified Manager.

Um componente que está em disputa não pode funcionar em um nível ideal. Seu desempenho diminuiu e o desempenho de outros componentes e cargas de trabalho do cluster, chamados *vítimas*, pode ter aumentado a latência. Para sair da contenção de um componente, você precisa reduzir o workload ou aumentar a capacidade de lidar com mais trabalho, para que a performance possa retornar aos níveis normais. Como o Unified Manager coleta e analisa a performance do workload em intervalos de cinco minutos, ele detecta quando um componente do cluster é consistentemente sobreusado. Não são detectados picos transitórios de sobreutilização que duram apenas uma curta duração dentro do intervalo de cinco minutos.

Por exemplo, um agregado de storage pode estar sob contenção porque um ou mais workloads nele estão competindo para que suas solicitações de e/S sejam atendidas. Outras cargas de trabalho no agregado podem ser afetadas, fazendo com que seu desempenho diminua. Para reduzir a quantidade de atividade no agregado, há etapas diferentes, como mover uma ou mais cargas de trabalho para um agregado menos ocupado, para diminuir a demanda geral de workload no agregado atual. Para um grupo de políticas de QoS, você pode ajustar o limite de taxa de transferência ou mover workloads para um grupo de políticas diferente, para que os workloads não fiquem mais sendo controlados.

O Unified Manager monitora os seguintes componentes do cluster para alertá-lo quando eles estão na contenção:

- **Rede**

Representa o tempo de espera das solicitações de e/S pelos protocolos iSCSI ou Fibre Channel (FC) no cluster. O tempo de espera é o tempo gasto esperando que as transações iSCSI Ready to Transfer (R2T) ou FCP Transfer Ready (XFER_RDY) sejam concluídas antes que o cluster possa responder a uma solicitação de e/S. Se o componente de rede estiver em contenção, isso significa que o alto tempo de espera na camada de protocolo de bloco está impactando a latência de uma ou mais cargas de trabalho.

- **Processamento de rede**

Representa o componente de software no cluster envolvido com o processamento de e/S entre a camada de protocolo e o cluster. O processamento da rede de tratamento do nó pode ter sido alterado desde que o evento foi detectado. Se o componente de processamento de rede estiver em contenção, isso significa que a alta utilização no nó de processamento de rede está impactando a latência de uma ou mais cargas de trabalho.

- **Política de QoS**

Representa o grupo de políticas de qualidade do serviço (QoS) de storage do qual o workload é membro. Se o componente do grupo de políticas estiver na contenção, isso significa que todas as cargas de trabalho no grupo de políticas estão sendo controladas pelo limite de taxa de transferência definido, o que está impactando a latência de uma ou mais dessas cargas de trabalho.

- **Interconexão de cluster**

Representa os cabos e adaptadores com os quais os nós em cluster estão fisicamente conectados. Se o componente de interconexão de cluster estiver na contenção, isso significa que o tempo de espera alto para solicitações de e/S na interconexão de cluster está impactando a latência de um ou mais workloads.

- **Data Processing**

Representa o componente de software no cluster envolvido com o processamento de e/S entre o cluster e o agregado de storage que contém a carga de trabalho. O Data Processing de tratamento do nó pode ter sido alterado desde que o evento foi detectado. Se o componente Data Processing estiver em contenção, isso significa que a alta utilização no nó Data Processing está impactando a latência de um ou mais workloads.

- **Recursos MetroCluster**

Representa os recursos do MetroCluster, incluindo NVRAM e links interswitches (ISLs), usados para espelhar dados entre clusters em uma configuração do MetroCluster. Se o componente MetroCluster estiver em contenção, isso significa que a alta taxa de transferência de gravação de workloads no cluster local ou um problema de integridade de link está impactando a latência de um ou mais workloads no cluster local. Se o cluster não estiver em uma configuração do MetroCluster, este ícone não será exibido.

- **Operações agregadas ou SSD agregadas**

Representa o agregado de storage no qual os workloads estão sendo executados. Se o componente agregado estiver na contenção, isso significa que a alta utilização no agregado está impactando a latência de um ou mais workloads. Um agregado consiste em todos os HDDs ou uma combinação de HDDs e SSDs (agregado de Flash Pool). Um "agregado SSD" consiste em todos os SSDs (um agregado all-flash) ou uma combinação de SSDs e uma camada de nuvem (agregado FabricPool).

- **Latência da nuvem**

Representa o componente de software no cluster envolvido com o processamento de e/S entre o cluster e a camada de nuvem na qual os dados do usuário são armazenados. Se o componente de latência da nuvem estiver em contenção, isso significa que uma grande quantidade de leituras de volumes hospedados na camada de nuvem está impactando a latência de um ou mais workloads.

- **Sincronizar SnapMirror**

Representa o componente de software no cluster envolvido com a replicação dos dados do usuário do volume primário para o volume secundário em uma relação síncrona do SnapMirror. Se o componente Sync SnapMirror estiver na contenção, isso significa que a atividade das operações síncronas do SnapMirror está impactando a latência de um ou mais workloads.

Funções dos workloads envolvidos em um evento de desempenho

O Unified Manager usa funções para identificar o envolvimento de um workload em um evento de performance. Os papéis incluem vítimas, agressores e tubarões. Uma carga de trabalho definida pelo usuário pode ser uma vítima, um valentão e um tubarão ao mesmo tempo.

Função	Descrição
Vítima	Uma carga de trabalho definida pelo usuário cujo desempenho diminuiu devido a outras cargas de trabalho, chamadas de bullies, que usam excessivamente um componente de cluster. Somente cargas de trabalho definidas pelo usuário são identificadas como vítimas. O Unified Manager identifica os workloads da vítima com base em seu desvio de latência, em que a latência real, durante um evento, aumentou muito em relação ao intervalo esperado de latência.
Bully	Uma carga de trabalho definida pelo usuário ou definida pelo sistema cujo uso excessivo de um componente de cluster causou a diminuição do desempenho de outras cargas de trabalho, chamadas vítimas. O Unified Manager identifica cargas de trabalho bully com base em seu desvio no uso de um componente do cluster, em que o uso real, durante um evento, aumentou muito em relação ao intervalo de uso esperado.
Tubarão	Um workload definido pelo usuário com a maior utilização de um componente de cluster em comparação a todas as cargas de trabalho envolvidas em um evento. O Unified Manager identifica workloads de tubarão com base no uso de um componente de cluster durante um evento.

Os workloads em um cluster podem compartilhar muitos dos componentes do cluster, como agregados de storage e CPU para rede e Data Processing. Quando uma carga de trabalho, como um volume, aumenta o uso de um componente de cluster a ponto de que o componente não pode atender com eficiência às demandas de workload, o componente está em contenção. A carga de trabalho que está usando um componente de cluster é um bully. As outras cargas de trabalho que compartilham esses componentes, e cujo desempenho é afetado pelo agressor, são as vítimas. As atividades de workloads definidos pelo sistema, como deduplicação ou cópias Snapshot, também podem escalar para "bullying".

Quando o Unified Manager detecta um evento, ele identifica todos os workloads e componentes de cluster envolvidos, incluindo os workloads bully que causaram o evento, o componente do cluster que está em contenção e os workloads da vítima cujo desempenho diminuiu devido ao aumento da atividade dos workloads bully.



Se o Unified Manager não conseguir identificar os workloads bully, ele só alertará sobre os workloads da vítima e o componente do cluster envolvido.

O Unified Manager pode identificar workloads vítimas de workloads bully e também identificar quando esses mesmos workloads se tornam workloads bully. Uma carga de trabalho pode ser um bully para si mesma. Por exemplo, uma carga de trabalho de alta performance que está sendo controlada por um limite de grupo de políticas faz com que todas as cargas de trabalho no grupo de políticas sejam limitadas, incluindo a própria. Uma carga de trabalho que é um agressor ou uma vítima em um evento de desempenho contínuo pode mudar sua função ou não ser mais um participante no evento. Na página Detalhes de desempenho/volume,

na tabela Lista de Eventos, quando o volume selecionado altera sua função participante, a data e a hora da alteração de função são exibidas.

Informações sobre direitos autorais

Copyright © 2024 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

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

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

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

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

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

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

Informações sobre marcas comerciais

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