Présentation des événements de performances et des alertes

Active IQ Unified Manager

NetApp
June 23, 2025

Sommaire

Présentation des événements de performances et des alertes

Les événements de performance sont des incidents liés aux performances des charges de travail sur un cluster. Ils vous aident à identifier les workloads avec des temps de réponse lents. Avec les événements de santé qui se sont produits en même temps, vous pouvez déterminer les problèmes qui pourraient avoir causé, ou contribué à, les délais de réponse lents.

Lorsque Unified Manager détecte plusieurs occurrences de la même condition d'événement pour le même composant de cluster, il traite toutes les occurrences comme un événement unique et non comme des événements distincts.

Vous pouvez configurer des alertes pour envoyer automatiquement une notification par e-mail lorsque des événements de performance de certains types de gravité se produisent.

Sources des événements de performance

Les événements de performance sont des problèmes liés aux performances des charges de travail sur un cluster. Ils vous aident à identifier les objets de stockage avec des temps de réponse lents, également appelés « latence élevée ». Avec d'autres événements de santé qui se sont produits en même temps, vous pouvez déterminer les problèmes qui pourraient avoir causé, ou contribué à, les délais de réponse lents.

Unified Manager reçoit des événements de performance des sources suivantes :

- **Événements de politique de seuil de performances définis par l'utilisateur**

  Problèmes de performances basés sur des valeurs de seuil personnalisées que vous avez définies. Vous configurez des règles de seuil de performances pour les objets de stockage, par exemple des agrégats et des volumes, de sorte que les événements soient générés lorsqu'une valeur de seuil pour un compteur de performances a été atteinte.

  Vous devez définir une règle de seuil de performances et l'affecter à un objet de stockage pour recevoir ces événements.

- **Événements de politique de seuil de performances définis par le système**

  Problèmes de performances basés sur des valeurs seuils définies par le système. Ces règles de seuil sont incluses dans l'installation de Unified Manager afin de couvrir les problèmes de performance les plus courants.

  Ces règles de seuil sont activées par défaut et vous pouvez afficher des événements peu après l'ajout d'un cluster.

- **Événements seuil de performances dynamiques**

  Problèmes de performance dus à des défaillances ou à des erreurs dans une infrastructure IT, ou à la sur-utilisation des ressources du cluster par les charges de travail. La cause de ces événements peut être un simple problème qui se corrige au cours d'un certain temps ou qui peut être résolu par une réparation ou un changement de configuration. Un événement à seuil dynamique indique que les workloads d'un

système ONTAP sont lents en raison d'autres workloads dont l'utilisation des composants du cluster partagé est élevée.

Ces seuils sont activés par défaut et vous pouvez afficher des événements après trois jours de collecte des données d'un nouveau cluster.

Types de sévérité des événements de performance

Chaque événement de performance est associé à un type de gravité pour vous aider à hiérarchiser les événements nécessitant une action corrective immédiate.

- **Critique**

  Un événement sur les performances peut entraîner une interruption des services si des actions correctives ne sont pas prises immédiatement.

  Les événements critiques sont envoyés à partir de seuils définis par l'utilisateur uniquement.

- **Avertissement**

  Un compteur de performances pour un objet de cluster est hors de la plage normale et doit être surveillé pour vérifier qu'il n'atteint pas la gravité critique. Les événements de ce niveau de gravité n'entraînent pas d'interruption des services, mais une action corrective immédiate peut ne pas être nécessaire.

  Les événements d'avertissement sont envoyés à partir de seuils définis par l'utilisateur, définis par le système ou dynamiques.

- **Information**

  L'événement se produit lorsqu'un nouvel objet est découvert ou lorsqu'une action utilisateur est exécutée. Par exemple, lorsqu'un objet de stockage est supprimé ou en cas de modification de la configuration, l'événement contenant des informations de type de gravité est généré.

  Les événements d'informations sont envoyés directement depuis ONTAP lorsqu'il détecte une modification de configuration.

Pour plus d'informations, consultez les liens suivants :

- "Que se passe-t-il lorsqu'un événement est reçu"
- "Les informations contenues dans un e-mail d'alerte"
- "Ajout d'alertes"
- "Ajout d'alertes en cas d'événements de performances"

Modifications de configuration détectées par Unified Manager

Unified Manager surveille vos clusters pour modifier la configuration, ce qui vous permet de déterminer si une modification a pu être causée ou contribué à un événement de performances. Les pages de l'Explorateur de performances affichent une icône d'événement de changement () pour indiquer la date et l'heure de détection de la

modification.

Vous pouvez consulter les graphiques de performances dans les pages de l'explorateur de performances et dans la page analyse de la charge de travail pour voir si l'événement de modification a affecté les performances de l'objet de cluster sélectionné. Si la modification a été détectée en même temps qu'un événement de performance ou à peu près, la modification peut avoir contribué au problème, qui a déclenché l'alerte d'événement.

Unified Manager peut détecter les événements de modification suivants, classés dans la catégorie « événements d'information » :

- Un volume est déplacé entre agrégats.

  Unified Manager peut détecter lorsque le déplacement est en cours, terminé ou échoué. Lorsqu'Unified Manager est inactif pendant le déplacement d'un volume, lors de sa sauvegarde, il détecte le déplacement de volume et affiche un événement de modification pour celui-ci.

- Le débit (Mbit/s ou IOPS) d'un groupe de règles de QoS contenant un ou plusieurs changements de charge de travail surveillés.

  La modification de la limite d'un groupe de règles peut entraîner des pics intermittents de latence (temps de réponse), qui peuvent également déclencher des événements pour le groupe de règles. La latence revient progressivement à la normale et tous les événements provoqués par les pics deviennent obsolètes.

- Un nœud d'une paire haute disponibilité prend le relais ou renvoie le stockage de son nœud partenaire.

  Unified Manager peut détecter la fin de l'opération de basculement, de basculement partiel ou de rétablissement. Si le basculement est causé par un nœud paniqué, Unified Manager ne détecte pas l'événement.

- Une opération de mise à niveau ou de restauration de ONTAP a été effectuée correctement.

  La version précédente et la nouvelle version sont affichées.

Types de règles de seuils de performance définies par le système

Unified Manager fournit des règles de seuil standard qui contrôlent les performances du cluster et génèrent automatiquement des événements. Ces règles sont activées par défaut et génèrent des événements d'avertissement ou d'information lorsque les seuils de performances surveillés sont enfreintes.

Les règles de seuil de performance définies par le système ne sont pas activées sur les systèmes Cloud Volumes ONTAP, ONTAP Edge ou ONTAP Select.

Si vous recevez des événements inutiles provenant de règles de seuils de performance définies par le système, vous pouvez désactiver les événements de règles individuelles à partir de la page de configuration des événements.

Règles de seuil du cluster

Les règles de seuil des performances du cluster définies par le système sont attribuées, par défaut, à chaque cluster contrôlé par Unified Manager :

- **Déséquilibre de charge du groupe**

  Identifie les situations où un nœud fonctionne à une charge bien plus élevée que les autres nœuds du cluster et peut donc affecter les latences des charges de travail.

  Pour ce faire, il compare la valeur de capacité en termes de performances utilisée pour tous les nœuds d'un cluster afin de voir si un nœud a dépassé la valeur seuil de 30 % pendant plus de 24 heures. Il s'agit d'un incident d'avertissement.

- **Déséquilibre de capacité du groupe**

  Identifie les situations où la capacité utilisée d'un agrégat est bien plus élevée que celle des autres agrégats du cluster et affecte donc potentiellement l'espace requis pour les opérations.

  Pour ce faire, elle compare la valeur de capacité utilisée de tous les agrégats du cluster afin de voir si la différence entre 70 % d'un agrégat. Il s'agit d'un incident d'avertissement.

Règles de seuil des nœuds

Les règles de seuil de performance des nœuds définies par le système sont attribuées par défaut à chaque nœud des clusters contrôlé par Unified Manager :

- **Seuil de capacité utilisée de performances dépassé**

  Identifie les situations dans lesquelles un nœud fonctionne au-delà des limites de son efficacité opérationnelle et risque par conséquent d'affecter la latence des charges de travail.

  Pour ce faire, il recherche des nœuds qui utilisent plus de 100 % de leur capacité en performance pendant plus de 12 heures. Il s'agit d'un incident d'avertissement.

- **Surutilisation de la paire HA de nœuds**

  Identifie les situations dans lesquelles les nœuds d'une paire haute disponibilité fonctionnent au-dessus des limites de l'efficacité opérationnelle de la paire haute disponibilité.

  Pour ce faire, le système étudie la valeur de la capacité en termes de performances utilisée pour les deux nœuds de la paire haute disponibilité. Si la capacité de performance combinée des deux nœuds dépasse 200 % pendant plus de 12 heures, un basculement de contrôleur affecte les latences des charges de travail. Il s'agit d'un événement informatif.

- **Fragmentation de disque de nœud**

  Identifie les situations où un ou plusieurs disques d'un agrégat sont fragmentés, ralentissant les principaux services système et potentiellement affecter les latences des charges de travail sur un nœud.

  Pour ce faire, il s'agit de certains ratios d'opération de lecture et d'écriture sur tous les agrégats d'un nœud. Cette règle peut également être déclenchée lors de la resynchronisation SyncMirror ou lorsque des erreurs sont détectées lors des opérations de nettoyage du disque. Il s'agit d'un incident d'avertissement.

La règle de « fragmentation des disques des nœuds » analyse les agrégats uniquement composés de disques durs ; les agrégats Flash Pool, SSD et FabricPool ne sont pas analysés.

Règles de seuil agrégées

La règle de seuil de performance des agrégats définis par le système est attribuée par défaut à chaque agrégat des clusters contrôlé par Unified Manager :

- **Disques agrégés sur-utilisés**

  Identifie les situations dans lesquelles un agrégat fonctionne au-delà des limites de son efficacité opérationnelle et peut ainsi affecter le latence des charges de travail. Ce cas est identifié par la recherche d'agrégats où les disques de l'agrégat sont utilisés à plus de 95 % pendant plus de 30 minutes. Cette règle multicondition effectue alors l'analyse suivante pour déterminer la cause du problème :

  - Un disque de l'agrégat est-il actuellement en cours d'opération de maintenance en arrière-plan ?

    Certaines activités de maintenance en arrière-plan qu'un disque peut être en cours de reconstruction sont : disque, nettoyage de disque, resynchronisation SyncMirror et reparité.

  - Existe-t-il un goulet d'étranglement au niveau des communications dans l'interconnexion Fibre Channel du tiroir disque ?
  - L'agrégat dispose-t-il trop peu d'espace libre ? Un événement d'avertissement est émis pour cette politique uniquement si une ou plusieurs des trois politiques subordonnées sont également considérées comme enfreintes. Un événement de performances n'est pas déclenché si seuls les disques de l'agrégat sont utilisés à plus de 95 %.

La politique « d'agrégation de disques sur-utilisés » analyse les agrégats de disques durs uniquement et les agrégats Flash Pool (hybrides) ; les agrégats SSD et FabricPool ne sont pas analysés.

Règles de seuil de latence des workloads

Les règles de seuil de latence de la charge de travail définies par le système sont attribuées à toute charge de travail dont la règle de niveau de service de performance est configurée et dont la valeur de « latence attendue » est définie :

- **Seuil de latence de volume de charge de travail/LUN dépassé tel que défini par le niveau de service de performances**

  Identifie les volumes (partages de fichiers) et les LUN qui ont dépassé leur limite de « latence attendue » et qui ont un impact sur les performances des charges de travail. Il s'agit d'un incident d'avertissement.

  Pour ce faire, il recherche des charges de travail qui ont dépassé la valeur de latence prévue pour 30 % de l'heure précédente.

Règles de seuil de QoS

Les règles de seuil de performances de QoS définies par le système sont attribuées à toute charge de travail dont la règle de débit maximal est la QoS ONTAP configurée (IOPS, IOPS/To ou Mo/s). Unified Manager déclenche un événement lorsque la valeur du débit des workloads est inférieure de 15 % à la valeur de la QoS

configurée :

- **QoS Max IOPS ou seuil MB/s**

  Identifie les volumes et les LUN qui ont dépassé leur limite maximale en termes d'IOPS ou de débit en Mo/s de qualité de service, et qui affectent la latence des charges de travail. Il s'agit d'un incident d'avertissement.

  Lorsqu'une seule charge de travail est attribuée à un groupe de règles, elle recherche les charges de travail qui ont dépassé le seuil de débit maximal défini dans le groupe de règles QoS attribué au cours de chaque période de collecte pendant l'heure précédente.

  Lorsque plusieurs charges de travail partagent une seule règle de QoS, celle-ci est ajoutée en ajoutant les IOPS ou les Mo/s de tous les workloads de la règle et en vérifiant le total dans la limite.

- **QoS Peak IOPS/To ou IOPS/To avec seuil de taille de bloc**

  Identifie les volumes qui ont dépassé la limite de débit en IOPS/To adaptative pour la qualité de service (ou IOPS/To avec limite de taille de bloc), tout en affectant la latence de la charge de travail. Il s'agit d'un incident d'avertissement.

  Pour ce faire, la conversion du seuil maximal d'IOPS/To défini dans la règle de QoS adaptative en une valeur maximale d'IOPS basée sur la taille de chaque volume. Elle recherche les volumes qui ont dépassé la limite d'IOPS maximale de QoS au cours de chaque période de collecte de performances pendant l'heure précédente.

  > Cette règle s'applique aux volumes uniquement lorsque le cluster est installé avec ONTAP 9.3 et les versions ultérieures.

  Lorsque l'élément « taille de bloc » a été défini dans la règle de QoS adaptative, le seuil est converti en valeur MB/s maximale basée sur la taille de chaque volume. Ensuite, il recherche les volumes qui ont dépassé la limite de qualité de service en Mo/s au cours de chaque période de collecte des performances pour l'heure précédente.

  > Cette règle s'applique aux volumes uniquement lorsque le cluster est installé avec ONTAP 9.5 et les versions ultérieures.

Analyse et notification des événements de performance

Les événements de performance vous signalent les problèmes de performances d'E/S concernant une charge de travail générée par des conflits sur un composant de cluster. Unified Manager analyse l'événement pour identifier toutes les charges de travail impliquées, le composant dans les conflits et si l'événement reste un problème à résoudre.

Unified Manager surveille la latence (temps de réponse) et les IOPS (opérations) des volumes d'un cluster. Lorsque d'autres charges de travail surfont un composant de cluster, par exemple, les conflits sont possibles et le composant ne peut pas fonctionner à un niveau optimal pour répondre aux demandes de charge de travail. Les performances des autres charges de travail qui utilisent le même composant peuvent être affectées, ce qui entraîne une augmentation des latences. Si la latence franchit le seuil de performance dynamique, Unified Manager déclenche un événement de performance afin de vous en avertir.

Analyse des événements

Unified Manager effectue les analyses suivantes, en s'appuyant sur les statistiques de performance des 15 derniers jours, pour identifier les workloads victime, les workloads dominants et le composant de cluster impliqué dans un événement :

- Identifie les charges de travail victimes dont la latence a dépassé le seuil de performance dynamique, qui est la limite supérieure de la prévision de latence :
  - Pour les volumes des agrégats hybrides HDD ou Flash Pool (niveau local), les événements sont déclenchés uniquement lorsque la latence est supérieure à 5 millisecondes (ms) et que les IOPS représentent plus de 10 opérations par seconde (OPS/s).
  - Pour les volumes situés sur des agrégats 100 % SSD ou des agrégats FabricPool (niveau cloud), les événements sont déclenchés uniquement lorsque la latence est supérieure à 1 ms et que les IOPS sont plus de 100 OPS/s.
- Identifie le composant de cluster dans les conflits.

  Si la latence des charges de travail victimes au niveau de l'interconnexion de cluster est supérieure à 1 ms, Unified Manager le traite comme important et déclenche un événement pour l'interconnexion de cluster.

- Identifie les charges de travail dominantes qui font l'objet d'une surutilisation du composant de cluster et qui l'entraînent des conflits.
- Classe les charges de travail impliquées, en fonction de leur déviation de l'utilisation ou de l'activité d'un composant du cluster, afin de déterminer les principaux changements d'utilisation du composant du cluster et les victimes les plus affectées.

Un événement peut se produire brièvement et se corriger après le composant qu'il utilise n'est plus en conflit. Un événement continu est un événement qui se produit de nouveau pour le même composant de cluster au cours d'un intervalle de cinq minutes et qui reste à l'état actif. Pour les événements continus, Unified Manager déclenche une alerte après avoir détecté le même événement à deux intervalles d'analyse consécutifs.

Lorsqu'un événement est résolu, il reste disponible dans Unified Manager dans le cadre de l'enregistrement des anciens problèmes de performances d'un volume. Chaque événement possède un ID unique qui identifie le type d'événement et les volumes, le cluster et les composants de cluster impliqués.

Un seul volume peut être impliqué dans plusieurs événements simultanément.

État de l'événement

Les événements peuvent être dans l'un des États suivants :

- **Actif**

  Indique que l'événement de performance est actuellement actif (nouveau ou reconnu). Le problème à l'origine de l'incident n'a pas été corrigé lui-même ou n'a pas été résolu. Le compteur de performances de l'objet de stockage reste au-dessus du seuil de performance.

- **Obsolète**

  Indique que l'incident n'est plus actif. Le problème à l'origine de l'incident s'est corrigé ou a été résolu. Le compteur de performance de l'objet de stockage n'est plus au-dessus du seuil de performance.

Notification d’événement

Les événements sont affichés sur la page Tableau de bord et sur de nombreuses autres pages de l’interface utilisateur, et les alertes pour ces événements sont envoyées à des adresses e-mail spécifiées. Vous pouvez afficher des informations d’analyse détaillées sur un événement et obtenir des suggestions de résolution de cet événement sur la page Détails de l’événement et sur la page analyse des charges de travail.

Interaction d’événement

Sur la page Détails de l’événement et sur la page analyse de la charge de travail, vous pouvez interagir avec les événements de la manière suivante :

- Le déplacement de la souris sur un événement affiche un message indiquant la date et l’heure de détection de l’événement.

  S’il y a plusieurs événements pour la même période, le message indique le nombre d’événements.

- Lorsque vous cliquez sur un seul événement, une boîte de dialogue affiche des informations plus détaillées sur l’événement, notamment les composants de cluster impliqués.

  Le composant en conflit est entouré et mis en évidence en rouge. Vous pouvez cliquer sur **Afficher l’analyse complète** pour afficher l’analyse complète sur la page Détails de l’événement. S’il existe plusieurs événements pour la même période, la boîte de dialogue affiche des détails sur les trois événements les plus récents. Vous pouvez cliquer sur un événement pour afficher l’analyse des événements sur la page Détails de l’événement.

Comment Unified Manager détermine l’impact sur les performances d’un événement

Unified Manager utilise l’écart d’activité, d’utilisation, de débit d’écriture, de l’utilisation d’un composant du cluster ou de latence d’E/S (temps de réponse) pour une charge de travail afin de déterminer le niveau d’impact sur les performances d’une charge de travail. Ces informations déterminent le rôle de chaque charge de travail dans l’événement et leur classement sur la page Détails de l’événement.

Unified Manager compare les dernières valeurs analysées pour une charge de travail à la plage de valeurs attendue (prévision de latence). La différence entre les valeurs analysées pour la dernière fois et la plage de valeurs attendue identifie les workloads pour lesquels les performances ont le plus été affectées par l’événement.

Supposons par exemple qu’un cluster contienne deux charges de travail : la charge De travail A et la charge de travail B. Les prévisions de latence pour la charge de travail A sont de 5-10 millisecondes par opération (ms/op) et sa latence réelle est généralement d’environ 7 ms/op. La prévision de latence pour la charge de travail B est de 10-20 ms/op et sa latence réelle est généralement d’environ 15 ms/op. La latence prévue pour les deux charges de travail est très bonne. En raison de conflits sur le cluster, la latence des deux charges de travail augmente à 40 ms/opération, franchissement du seuil de performance dynamique, qui correspond aux limites supérieures des prévisions de latence et au déclenchement d’événements. L’écart de latence, entre les valeurs attendues et les valeurs supérieures au seuil de performances, pour la charge de travail A est d’environ 33 ms/op, et l’écart pour la charge de travail B est d’environ 25 ms/op. La latence des deux charges de travail atteint 40 ms/activité, mais la charge de travail A avait l’impact le plus important sur les performances, car elle avait l’écart de latence le plus élevé à 33 ms/opération.

Sur la page Détails de l'événement, dans la section diagnostic système, vous pouvez trier les charges de travail par variation de l'activité, de l'utilisation ou du débit d'un composant de cluster. Vous pouvez également trier les charges de travail par latence. Lorsque vous sélectionnez une option de tri, Unified Manager analyse l'écart en termes d'activité, d'utilisation, de débit ou de latence depuis que l'événement a été détecté à partir des valeurs attendues pour déterminer l'ordre de tri de la charge de travail. Pour la latence, les points rouges (🔴) indiquez un seuil de performances franchissement par une charge de travail victime et l'impact qui en découle sur la latence. Chaque point rouge indique un niveau d'écart plus élevé de latence, ce qui vous aide à identifier les workloads victimes dont la latence a le plus été affectée par un événement.

Les composants du cluster et les conflits

Vous pouvez identifier les problèmes de performance du cluster lorsqu'un composant du cluster entre en conflit. Les performances des charges de travail qui utilisent le ralentissement du composant et leur temps de réponse (latence) augmente pour les requêtes client, ce qui déclenche un événement dans Unified Manager.

Un composant en conflit ne peut pas se faire à un niveau optimal. Ses performances ont diminué, et la performance des autres composants et charges de travail du cluster, appelés *victimes*, peut avoir augmenté la latence. Pour mettre un composant à l'extérieur des conflits, vous devez réduire sa charge de travail ou augmenter sa capacité à gérer davantage de travail, de sorte que les performances puissent revenir à des niveaux normaux. Unified Manager collecte et analyse les performances des charges de travail toutes les cinq minutes. En effet, il ne détecte que lorsqu'un composant du cluster est constamment sur-utilisé. Les pics transitoires de surutilisation qui durent pendant une courte durée dans l'intervalle de cinq minutes ne sont pas détectés.

Par exemple, un agrégat de stockage peut être soumis à des conflits car une ou plusieurs charges de travail y sont en concurrence pour que leurs demandes d'E/S soient traitées. Des charges de travail peuvent être affectées sur l'agrégat, ce qui entraîne une baisse des performances. Pour réduire la quantité d'activité sur l'agrégat, différentes étapes sont possibles : déplacer une ou plusieurs charges de travail vers un agrégat ou un nœud moins occupé, par exemple, afin de réduire les besoins globaux de la charge de travail sur l'agrégat en cours. Pour un groupe de règles de qualité de service, vous pouvez ajuster la limite de débit ou déplacer les workloads vers un autre groupe de règles, de sorte que les charges de travail ne soient plus restreintes.

Unified Manager contrôle les composants de cluster suivants pour vous alerter en cas de conflit :

- **Réseau**

  Représente le temps d'attente des demandes d'E/S par les protocoles réseau externes sur le cluster. Le temps d'attente est le temps passé à attendre la fin des transactions « de transfert prêt » avant que le cluster puisse répondre à une demande d'E/S. Si le composant réseau constitue un conflit, cela signifie qu'un temps d'attente élevé au niveau de la couche de protocole a un impact sur la latence d'une ou de plusieurs charges de travail.

- **Traitement réseau**

  Composant logiciel dans le cluster impliqué dans le traitement des E/S entre la couche de protocole et le cluster. Le traitement du réseau de traitement des nœuds a peut-être changé depuis la détection de l'événement. Si le composant de traitement de réseau est en conflit, son utilisation élevée au niveau du nœud de traitement réseau a un impact sur la latence d'une ou de plusieurs charges de travail.

  Lors de l'utilisation d'un cluster All SAN Array dans une configuration active/active, la valeur de latence de traitement réseau s'affiche pour les deux nœuds afin que vous puissiez vérifier que les nœuds partagent la charge de manière égale.

- **Limite de qualité de service max**

  Représente le paramètre de débit maximal (crête) du groupe de règles de qualité de service (QoS) de stockage affecté à la charge de travail. Si le composant de groupe de règles conflits, cela signifie que toutes les charges de travail du groupe de règles sont restreintes par la limite de débit définie, qui a un impact sur la latence d'une ou plusieurs de ces charges de travail.

- **Limite de qualité de service min**

  Représente la latence pour une charge de travail générée par le paramètre de débit de QoS minimal (attendu) attribué à d'autres workloads. Si, pour certaines charges de travail, la qualité de service minimale est définie sur la majorité de la bande passante pour garantir le débit promis, d'autres charges de travail sont restreintes et affichent une latence plus élevée.

- * Interconnexion de cluster*

  La représente les câbles et adaptateurs avec lesquels les nœuds en cluster sont physiquement connectés. Si le composant d'interconnexion de cluster est en conflit, cela signifie un temps d'attente élevé pour les demandes d'E/S au niveau de l'interconnexion de cluster se répercute sur la latence d'une ou de plusieurs charges de travail.

- **Traitement de données**

  Composant logiciel dans le cluster impliqué dans le traitement des E/S entre le cluster et l'agrégat de stockage qui contient la charge de travail. Le traitement des données de traitement du nœud peut avoir changé depuis la détection de l'événement. Si le composant de traitement des données conflit, une utilisation élevée au niveau du nœud de traitement des données affecte la latence d'un ou de plusieurs workloads.

- **Activation du volume**

  Processus permettant de suivre l'utilisation de tous les volumes actifs. Dans les environnements de grande taille où plus de 1000 volumes sont actifs, ce processus surveille en même temps le nombre de volumes stratégiques devant accéder aux ressources par le biais du nœud. Lorsque le nombre de volumes actifs simultanés dépasse le seuil maximal recommandé, certains volumes non critiques subissent une latence telle qu'elle est identifiée ici.

- **Ressources MetroCluster**

  La représente les ressources MetroCluster, y compris la NVRAM et les liens ISL, utilisés pour mettre en miroir les données entre les clusters dans une configuration MetroCluster. Si le composant MetroCluster rencontre des conflits, il s'agit d'un débit d'écriture élevé avec les charges de travail sur le cluster local ou d'un problème d'état de santé de la liaison ayant un impact sur la latence d'une ou de plusieurs charges de travail sur le cluster local. Si le cluster ne se trouve pas dans une configuration MetroCluster, cette icône n'est pas affichée.

- **Agrégat ou agrégat SSD**

  Agrégat de stockage sur lequel les charges de travail s'exécutent. Si le composant de l'agrégat est en conflit, une utilisation élevée de l'agrégat a un impact sur la latence d'une ou de plusieurs charges de travail. Un agrégat se compose de tous les disques durs, ou d'un mélange de disques durs et de disques SSD (un agrégat Flash Pool), ou d'une combinaison de disques durs et d'un niveau de cloud (un agrégat FabricPool). Un « agrégat SD » se compose de tous les SSD (un agrégat 100 % Flash), ou d'une combinaison de SSD et d'un niveau cloud (un agrégat FabricPool).

- * Latence cloud*

  Représente le composant logiciel du cluster impliqué dans le traitement des E/S entre le cluster et le niveau cloud sur lequel les données utilisateur sont stockées. Si le composant de latence dans le cloud conflits, une grande quantité de lectures sur les volumes hébergés sur le Tier cloud ont une incidence sur la latence d'un ou de plusieurs workloads.

- **SnapMirror de synchronisation**

  Représente le composant logiciel du cluster impliqué dans la réplication des données utilisateur depuis le volume primaire vers le volume secondaire dans une relation SnapMirror synchrone. Si le composant SnapMirror synchrone entre en conflit, l'activité des opérations SnapMirror synchrone a un impact sur la latence d'un ou de plusieurs workloads.

Rôles des charges de travail impliquées dans un événement de performance

Unified Manager utilise des rôles pour identifier la participation d'une charge de travail en cas de performance. Les rôles sont les victimes, les taureaux et les requins. Une charge de travail définie par l'utilisateur peut être une victime, un tyran et un requin en même temps.

Rôle	Description
Victime	Charge de travail définie par l'utilisateur dont les performances ont diminué en raison des autres charges de travail, appelées « bullies », qui sont sur-utilisées lors de l'utilisation d'un composant du cluster. Seules les charges de travail définies par l'utilisateur sont identifiées comme victimes. Unified Manager identifie les charges de travail victimes en fonction de leur écart de latence, où la latence réelle, pendant un événement, a été considérablement améliorée par rapport à sa prévision de latence (plage prévue).
Intimider	Une charge de travail définie par l'utilisateur ou définie par le système dont l'utilisation excessive d'un composant de cluster a entraîné une diminution des performances d'autres charges de travail, appelées « victimes ». Unified Manager identifie les workloads dominants en fonction de leur déviation par l'utilisation d'un composant de cluster, où l'utilisation réelle, au cours d'un événement, a considérablement augmenté à partir de sa plage d'utilisation prévue.

Rôle	Description
Requin	Charge de travail définie par l'utilisateur, avec l'utilisation la plus élevée d'un composant de cluster, et non pas toutes les charges de travail impliquées dans un événement. Unified Manager identifie les charges de travail Shark en fonction de leur utilisation d'un composant de cluster pendant un événement.

Les charges de travail d'un cluster peuvent partager la plupart des composants du cluster, tels que les agrégats et la CPU pour le traitement du réseau et des données. Lorsqu'une charge de travail, par exemple un volume, augmente l'utilisation d'un composant de cluster au point que le composant ne peut pas répondre efficacement aux exigences de la charge de travail, le composant engendre des conflits. La charge de travail sur-utilisation d'un composant de cluster est un phénomène tyran. Les autres charges de travail qui partagent ces composants, et dont la performance est impactée par le tyran, sont les victimes. L'activité provenant des charges de travail définies par le système, telles que la déduplication ou les copies Snapshot, peut également créer des « brimades ».

Lorsqu'Unified Manager détecte un événement, il identifie tous les workloads et composants de cluster impliqués, notamment les workloads dominants qui ont causé l'événement, le composant de cluster en conflit et les workloads victimes dont les performances ont diminué en raison de l'augmentation de l'activité des workloads dominants.

Si Unified Manager ne peut pas identifier les charges de travail dominantes, cette alerte s'applique uniquement aux charges de travail victimes et au composant de cluster concerné.

Unified Manager est capable d'identifier les charges de travail victimes de charges de travail dominantes. Il peut également y avoir une identification lorsque ces mêmes charges de travail deviennent des charges de travail dominantes. Un workload peut être un tyran à lui-même. Par exemple, une charge de travail élevée au ralenti par une limite de groupe de règles entraîne le restriction de toutes les charges de travail du groupe de règles, y compris de celles-ci. Une charge de travail dominante ou victime dans un événement de performance continu peut changer son rôle ou ne plus y participer.