



Surveiller la latence du volume

EDA workloads

NetApp
February 02, 2026

Sommaire

Surveiller la latence du volume	1
Surveiller la latence du volume	1
Aperçu	1
Avant de commencer	1
Configurer les seuils de latence	1
Comprendre les alertes	2
Afficher les événements de latence	3
Gérer la configuration de la latence	4
Meilleures pratiques	4

Surveiller la latence du volume

Surveiller la latence du volume

L'analyse de latence vous permet de surveiller proactivement les performances des volumes en suivant les métriques de latence de lecture et d'écriture sur vos systèmes de fichiers FSx for ONTAP. Configurez des seuils personnalisables pour les événements d'avertissement et critiques afin d'identifier les goulets d'étranglement potentiels avant qu'ils n'impactent vos charges de travail EDA.

Aperçu

L'analyse de latence collecte et surveille CloudWatch métriques pour les opérations de lecture et d'écriture de volumes. Lorsque les seuils de latence et d'IOPS sont dépassés pour tous les points de données dans une plage de temps spécifiée, le système génère des alertes qui apparaissent dans le tableau des événements de latence. Cela vous permet de :

- Identifier les volumes présentant une dégradation des performances.
- Distinguer entre les problèmes de performance de niveau avertissement et ceux de niveau critique.
- Suivez l'évolution de la latence au fil du temps pour optimiser les configurations de stockage.
- Prenez des mesures proactives avant que la latence n'affecte les performances de la charge de travail.

Avant de commencer

Pour utiliser l'analyse de latence, vous devez avoir configuré des identifiants AWS dans Workload Factory. Cette fonctionnalité nécessite l'accès aux métriques CloudWatch pour tous les volumes FSx for ONTAP associés à vos identifiants AWS.

Si vous n'avez pas configuré les informations d'identification AWS, consultez "[Ajouter les identifiants AWS](#)".

Configurer les seuils de latence

Vous pouvez configurer des seuils pour les événements d'avertissement et critiques. Chaque type d'événement comprend des seuils distincts pour les opérations de lecture et d'écriture. Le système évalue ces seuils en continu et génère des alertes lorsque les conditions sont remplies.

 Vous devez définir des seuils d'événements critiques supérieurs aux seuils d'événements d'avertissement pour garantir une remontée d'alertes correcte. Sinon, vous ne pouvez pas enregistrer votre configuration.

À propos de cette tâche

Pour qu'une alerte soit déclenchée, les seuils de latence et d'IOPS doivent être dépassés pour tous les points de données dans l'intervalle de temps spécifié. Cette logique à double condition permet de réduire les faux positifs en garantissant que la latence élevée est maintenue sous une charge importante.

Mesures

1. Connectez-vous en utilisant l'un des "[expériences sur console](#)".
2. Sélectionnez le menu  et sélectionnez ensuite **EDA**.

3. Dans le menu EDA, sélectionnez **Latency**.

4. Sur la page de configuration de la latence EDA, configurez les seuils suivants :

◦ **Événements d'avertissement**

- **Seuil de latence de lecture** : Saisissez le seuil de latence en millisecondes. Par défaut : 6 ms.
- **Seuil IOPS de lecture** : Saisissez le seuil IOPS en opérations par seconde. Par défaut : 100 ops/sec.
- **Plage de temps de lecture** : Saisissez la plage de temps en minutes (5-20). Par défaut : 10 minutes.
- **Seuil de latence d'écriture** : Saisissez le seuil de latence en millisecondes. Par défaut : 8 ms.
- **Seuil d'IOPS d'écriture** : Saisissez le seuil d'IOPS en opérations par seconde. Par défaut : 100 ops/sec.
- **Plage de temps d'écriture** : Entrez la plage de temps en minutes (5-20). Par défaut : 10 minutes.

◦ **Événements critiques**

- **Seuil de latence de lecture** : Saisissez le seuil de latence en millisecondes. Par défaut : 12 ms.
- **Seuil IOPS de lecture** : Saisissez le seuil IOPS en opérations par seconde. Par défaut : 100 ops/sec.
- **Plage de temps de lecture** : Saisissez la plage de temps en minutes (5-20). Par défaut : 10 minutes.
- **Seuil de latence d'écriture** : Saisissez le seuil de latence en millisecondes. Par défaut : 15 ms.
- **Seuil d'IOPS d'écriture** : Saisissez le seuil d'IOPS en opérations par seconde. Par défaut : 100 ops/sec.
- **Plage de temps d'écriture** : Entrez la plage de temps en minutes (5-20). Par défaut : 10 minutes.

5. Sélectionnez **Appliquer**.

Résultat

Workload Factory commence à collecter les métriques de latence pour tous les volumes FSx for ONTAP associés à vos identifiants AWS. Les métriques sont collectées au moins toutes les 20 minutes. Le tableau des événements de latence affiche tous les volumes qui dépassent les seuils que vous avez configurés.

Comprendre les alertes

La fonction d'analyse de latence utilise des alarmes CloudWatch pour surveiller les performances du volume. Comprendre comment les alertes sont déclenchées vous aide à configurer les seuils appropriés et à interpréter les résultats.

Métriques collectées

Le système collecte les indicateurs CloudWatch suivants pour chaque volume :

- **Seuil de latence de lecture** : Calculé comme $1000 * m2/(m1+0,000001)$ où m1 = DataReadOperations et m2 = DataReadOperationTime
- **Seuil de latence d'écriture** : Calculé comme $1000 * m2/(m1+0,000001)$ où m1 = DataWriteOperations et m2 = DataWriteOperationTime

Conditions de déclenchement d'alerte

Une alerte est déclenchée lorsque toutes les conditions suivantes sont réunies :

- Le seuil de latence est dépassé pour le type d'opération (lecture ou écriture).
- Le seuil d'IOPS est dépassé pour le type d'opération.
- Les deux conditions persistent pour tous les points de données dans la plage horaire configurée.

Par exemple, avec les seuils d'avertissement par défaut, une alerte de lecture ne se déclenche que si la latence de lecture dépasse 6 ms ET si les IOPS de lecture dépassent 100 ops/sec pour tous les points de données sur une période de 10 minutes.

Gravité de l'événement

- **Événements d'alerte** : Indiquent une latence élevée qui pourrait nécessiter une attention.
- **Événements critiques** : Indiquent une latence sévère qui nécessite une enquête immédiate.

Afficher les événements de latence

Le tableau des événements de latence affiche tous les événements d'avertissement et critiques détectés au cours des dernières 72 heures. Utilisez ce tableau pour surveiller les performances des volumes et identifier les volumes qui nécessitent une optimisation.

Informations complémentaires

- Seule la dernière violation pour chaque volume apparaît dans le tableau. Si un volume subit plusieurs violations, seul l'événement le plus récent est affiché.
- Les événements sont automatiquement supprimés après 72 heures.
- Le tableau affiche un maximum de 200 événements. Les événements les plus anciens sont supprimés au fur et à mesure que de nouveaux événements sont ajoutés.

Mesures

1. Dans l'onglet **Latence**, consultez le tableau des événements de latence.
2. Consultez les informations pour chaque événement, notamment :
 - **Gravité** : Indique si l'événement est Critical ou Warning.
 - **Nom du volume** : Le nom du volume concerné.
 - **Volume ID** : L'ID du volume concerné.
 - **Système de fichiers** : Le système de fichiers FSx for ONTAP contenant le volume.
 - **Heure de détection** : Lorsque la brèche a été détectée
 - **Latence médiane** : La valeur médiane de la latence pendant la période de breach.
3. Pour trier le tableau, sélectionnez l'en-tête d'une colonne. Par défaut, les événements critiques apparaissent en premier triés par heure, suivis des événements d'avertissement triés par heure.
4. Pour rejeter un ou plusieurs événements, sélectionnez **Rejeter** à côté de chaque événement.
5. Pour ajouter des colonnes au tableau, sélectionnez l'icône de colonne, choisissez les colonnes et sélectionnez **Appliquer**.

Gérer la configuration de la latence

Après la configuration initiale, vous pouvez modifier vos seuils.

Mesures

1. Sur la page **Latency**, sélectionnez **Edit**.
2. Modifiez les valeurs de seuil selon vos besoins.



Veillez à ce que les seuils critiques restent supérieurs aux seuils d'alerte. Le système affiche une erreur si vous configurez des seuils critiques inférieurs aux seuils d'alerte.

3. Sélectionnez **Apply** pour enregistrer vos modifications.

Meilleures pratiques

Tenez compte de ces recommandations lors de la configuration et de l'utilisation de l'analyse de latence :

- **Définissez des seuils réalistes** : configurez les seuils en fonction des exigences de votre charge de travail. Les valeurs par défaut constituent un point de départ mais peuvent nécessiter un ajustement pour votre environnement spécifique.
- **Commencez par des seuils d'avertissement** : Utilisez les événements d'avertissement pour établir des attentes de performance de base avant d'affiner les seuils critiques.
- **Considérez soigneusement les intervalles de temps** : des intervalles plus courts (5-10 minutes) détectent les problèmes plus rapidement mais peuvent générer plus d'alertes. Des intervalles plus longs (15-20 minutes) réduisent les faux positifs mais peuvent retarder la détection.
- **Surveillez les tendances** : Examinez régulièrement le tableau des événements de latence pour identifier des schémas ou des problèmes récurrents qui pourraient indiquer des problèmes de configuration sous-jacents.
- **Coordonnez les seuils d'IOPS et de latence** : la logique à double condition signifie que les deux doivent être dépassés. Définir des seuils d'IOPS très élevés peut empêcher les alertes même lorsque la latence est problématique.
- **Examinez les événements rejetés** : Passez périodiquement en revue les raisons pour lesquelles les événements ont été rejetés afin d'identifier des opportunités d'ajustement des seuils ou d'amélioration de l'infrastructure.

Informations sur le copyright

Copyright © 2026 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUSSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTUELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

Informations sur les marques commerciales

NETAPP, le logo NETAPP et les marques citées sur le site <http://www.netapp.com/TM> sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.