Commencer

NetApp Data Classification

NetApp
February 06, 2026

Sommaire

Commencer

En savoir plus sur la NetApp Data Classification

NetApp Data Classification est un service de gouvernance des données pour la NetApp Console qui analyse vos sources de données d'entreprise sur site et dans le cloud pour mapper et classer les données et identifier les informations privées. Cela peut vous aider à réduire vos risques de sécurité et de conformité, à diminuer vos coûts de stockage et à vous aider dans vos projets de migration de données.

À partir de la version 1.31, la classification des données est disponible en tant que fonctionnalité principale dans la NetApp Console. Il n'y a pas de frais supplémentaires. Aucune licence de classification ni abonnement n'est requis. + Si vous avez utilisé la version héritée 1.30 ou une version antérieure, cette version est disponible jusqu'à l'expiration de votre abonnement.

NetApp Console

La classification des données est accessible via la NetApp Console.

La NetApp Console fournit une gestion centralisée des services de stockage et de données NetApp dans les environnements sur site et cloud à l'échelle de l'entreprise. La console est requise pour accéder aux services de données NetApp et les utiliser. En tant qu'interface de gestion, il vous permet de gérer de nombreuses ressources de stockage à partir d'une seule interface. Les administrateurs de console peuvent contrôler l'accès au stockage et aux services pour tous les systèmes de l'entreprise.

Vous n'avez pas besoin de licence ni d'abonnement pour commencer à utiliser NetApp Console et vous n'encourez des frais que lorsque vous devez déployer des agents de console dans votre cloud pour garantir la connectivité à vos systèmes de stockage ou à vos services de données NetApp . Cependant, certains services de données NetApp accessibles depuis la console sont sous licence ou basés sur un abonnement.

En savoir plus sur le"NetApp Console" .

Caractéristiques

La classification des données utilise l'intelligence artificielle (IA), le traitement du langage naturel (NLP) et l'apprentissage automatique (ML) pour comprendre le contenu qu'elle analyse afin d'extraire des entités et de catégoriser le contenu en conséquence. Cela permet à la classification des données de fournir les domaines de fonctionnalités suivants.

"En savoir plus sur les cas d'utilisation de la classification des données" .

Maintenir la conformité

La classification des données fournit plusieurs outils qui peuvent vous aider dans vos efforts de conformité. Vous pouvez utiliser la classification des données pour :

- Identifier les informations personnelles identifiables (PII).
- Identifiez un large éventail d'informations personnelles sensibles comme l'exigent les réglementations de confidentialité GDPR, CCPA, PCI et HIPAA.
- Répondre aux demandes d'accès aux données des personnes concernées (DSAR) en fonction du nom ou de l'adresse e-mail.

Renforcer la sécurité

La classification des données permet d'identifier les données potentiellement susceptibles d'être consultées à des fins criminelles. Vous pouvez utiliser la classification des données pour :

- Identifiez tous les fichiers et répertoires (partages et dossiers) avec des autorisations ouvertes qui sont exposés à l'ensemble de votre organisation ou au public.
- Identifiez les données sensibles qui résident en dehors de l'emplacement initial dédié.
- Respecter les politiques de conservation des données.
- Utilisez *Policies* pour détecter automatiquement les nouveaux problèmes de sécurité afin que le personnel de sécurité puisse agir immédiatement.

Optimiser l'utilisation du stockage

La classification des données fournit des outils qui peuvent vous aider à déterminer le coût total de possession (TCO) de votre stockage. Vous pouvez utiliser la classification des données pour :

- Augmentez l'efficacité du stockage en identifiant les données en double ou non liées à l'entreprise.
- Réduisez les coûts de stockage en identifiant les données inactives que vous pouvez hiérarchiser vers un stockage d'objets moins coûteux. "En savoir plus sur la hiérarchisation des systèmes Cloud Volumes ONTAP" . "En savoir plus sur la hiérarchisation des systèmes ONTAP sur site" .

Systèmes et sources de données pris en charge

La classification des données peut scanner et analyser des données structurées et non structurées provenant des types de systèmes et de sources de données suivants :

Systèmes

- Amazon FSx for NetApp ONTAP
- Azure NetApp Files
- Cloud Volumes ONTAP (déployé dans AWS, Azure ou GCP)
- Clusters ONTAP sur site
- StorageGRID
- Google Cloud NetApp Volumes

Sources de données

- Partages de fichiers NetApp
- Bases de données:
  - Service de base de données relationnelle Amazon (Amazon RDS)
  - MongoDB
  - MySQL
  - Oracle
  - PostgreSQL
  - SAP HANA
  - Serveur SQL (MSSQL)

La classification des données prend en charge les versions NFS 3.x, 4.0 et 4.1, ainsi que les versions CIFS 1.x, 2.0, 2.1 et 3.0.

Coût

La classification des données est gratuite. Aucune licence de classification ni abonnement payant n'est requis.

Coûts d'infrastructure

- L'installation de Data Classification dans le cloud nécessite le déploiement d'une instance cloud, ce qui entraîne des frais de la part du fournisseur cloud où elle est déployée. Voir le type d'instance déployé pour chaque fournisseur de cloud . L'installation de Data Classification sur un système local est gratuite.
- La classification des données nécessite que vous ayez déployé un agent de console. Dans de nombreux cas, vous disposez déjà d'un agent de console en raison d'autres stockages et services que vous utilisez dans la console. L'instance de l'agent de console entraîne des frais auprès du fournisseur de cloud où elle est déployée. Voir le "type d'instance déployée pour chaque fournisseur de cloud" . L'installation de l'agent de console sur un système local est gratuite.

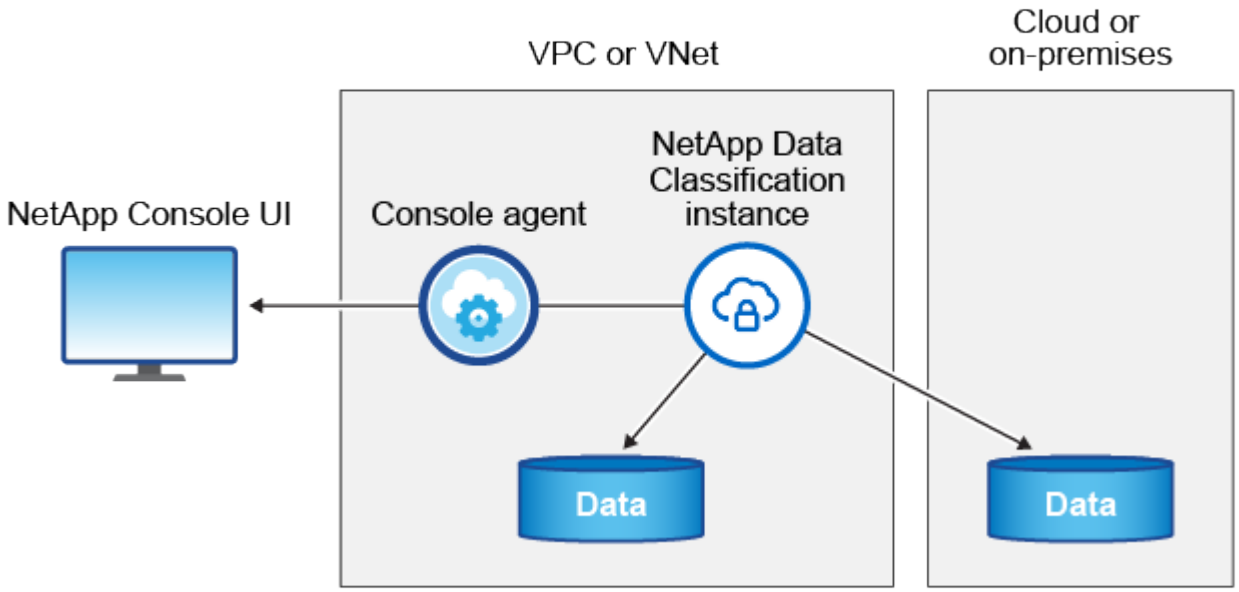
Coûts de transfert de données

Les coûts de transfert de données dépendent de votre configuration. Si l'instance de classification des données et la source de données se trouvent dans la même zone de disponibilité et la même région, il n'y a aucun coût de transfert de données. Mais si la source de données, comme un système Cloud Volumes ONTAP , se trouve dans une zone de disponibilité ou une région *différente*, les frais de transfert de données vous seront facturés par votre fournisseur de cloud. Consultez ces liens pour plus de détails :

- "AWS : Tarifs d'Amazon Elastic Compute Cloud (Amazon EC2)"
- "Microsoft Azure : Détails des tarifs de la bande passante"
- "Google Cloud : tarifs du service de transfert de stockage"

L'instance de classification des données

Lorsque vous déployez la classification des données dans le cloud, la console déploie l'instance dans le même sous-réseau que l'agent de la console. "En savoir plus sur l'agent de console."



Notez ce qui suit à propos de l'instance par défaut :

- Dans AWS, la classification des données s'exécute sur un "instance m6i.4xlarge" avec un disque GP2 de 500 Gio. L'image du système d'exploitation est Amazon Linux 2. Lorsqu'il est déployé dans AWS, vous pouvez choisir une taille d'instance plus petite si vous analysez une petite quantité de données.
- Dans Azure, la classification des données s'exécute sur un"VM Standard_D16s_v3" avec un disque de 500 Gio. L'image du système d'exploitation est Ubuntu 22.04.
- Dans GCP, la classification des données s'exécute sur un"n2-standard-16 VM" avec un disque persistant standard de 500 Gio. L'image du système d'exploitation est Ubuntu 22.04.
- Dans les régions où l'instance par défaut n'est pas disponible, la classification des données s'exécute sur une instance alternative. "Voir les types d'instances alternatifs" .
- L'instance est nommée *CloudCompliance* avec un hachage généré (UUID) concaténé. Par exemple : *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*
- Une seule instance de classification des données est déployée par agent de console.

Vous pouvez également déployer la classification des données sur un hôte Linux dans vos locaux ou sur un hôte chez votre fournisseur de cloud préféré. Le logiciel fonctionne exactement de la même manière, quelle que soit la méthode d'installation choisie. Les mises à niveau du logiciel de classification des données sont automatisées tant que l'instance dispose d'un accès Internet.

L'instance doit rester en cours d'exécution en permanence, car la classification des données analyse en permanence les données.

Déployer sur différents types d'instances

Consultez les spécifications suivantes pour les types d'instances :

Taille du système	Spécifications	Limites
Très grand	32 processeurs, 128 Go de RAM, 1 To de SSD	Peut numériser jusqu'à 500 millions de fichiers.
Grand (par défaut)	16 processeurs, 64 Go de RAM, 500 Go de SSD	Peut numériser jusqu'à 250 millions de fichiers.

Lors du déploiement de la classification des données dans Azure ou GCP, envoyez un e-mail à ng-contact-data-sense@netapp.com pour obtenir de l'aide si vous souhaitez utiliser un type d'instance plus petit.

Comment fonctionne l'analyse de classification des données

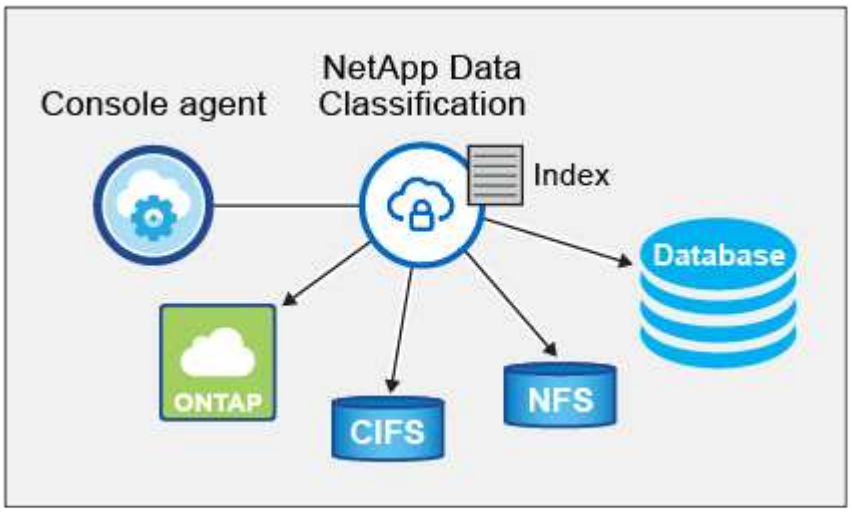
À un niveau élevé, l'analyse de classification des données fonctionne comme ceci :

1. Vous déployez une instance de classification des données dans la console.
2. Vous activez le mappage de haut niveau (appelé analyses *Mapping uniquement*) ou l'analyse de niveau profond (appelée analyses *Map & Classify*) sur une ou plusieurs sources de données.
3. La classification des données analyse les données à l'aide d'un processus d'apprentissage de l'IA.
4. Vous utilisez les tableaux de bord et les outils de reporting fournis pour vous aider dans vos efforts de conformité et de gouvernance.

Une fois que vous avez activé la classification des données et sélectionné les référentiels que vous souhaitez analyser (il s'agit des volumes, des schémas de base de données ou d'autres données utilisateur), l'analyse

des données commence immédiatement pour identifier les données personnelles et sensibles. Dans la plupart des cas, vous devez vous concentrer sur l'analyse des données de production en direct plutôt que sur les sauvegardes, les miroirs ou les sites de reprise après sinistre. Ensuite, la classification des données cartographie vos données organisationnelles, catégorise chaque fichier et identifie et extrait les entités et les modèles prédéfinis dans les données. Le résultat de l'analyse est un index des informations personnelles, des informations personnelles sensibles, des catégories de données et des types de fichiers.

La classification des données se connecte aux données comme n'importe quel autre client en montant des volumes NFS et CIFS. Les volumes NFS sont automatiquement accessibles en lecture seule, tandis que vous devez fournir les informations d'identification Active Directory pour analyser les volumes CIFS.



Après l'analyse initiale, Data Classification analyse en continu vos données de manière circulaire pour détecter les modifications incrémentielles. C'est pourquoi il est important de maintenir l'instance en cours d'exécution.

Vous pouvez activer et désactiver les analyses au niveau du volume ou au niveau du schéma de base de données.

La classification des données n'impose pas de limite à la quantité de données qu'elle peut analyser. Chaque agent de console prend en charge l'analyse et l'affichage de 500 Tio de données. Pour scanner plus de 500 Tio de données,"installer un autre agent de console" alors"déployer une autre instance de classification des données" . + L'interface utilisateur de la console affiche les données d'un seul connecteur. Pour obtenir des conseils sur l'affichage des données de plusieurs agents de console, consultez"Travailler avec plusieurs agents de console" .

Quelle est la différence entre les analyses de cartographie et de classification

Vous pouvez effectuer deux types d'analyses dans la classification des données :

- **Les analyses de cartographie uniquement** fournissent uniquement un aperçu de haut niveau de vos données et sont effectuées sur des sources de données sélectionnées. Les analyses de cartographie uniquement prennent moins de temps que les analyses de cartographie et de classification, car elles n'accèdent pas aux fichiers pour voir les données qu'ils contiennent. Vous souhaiterez peut-être procéder ainsi dans un premier temps pour identifier les domaines de recherche, puis effectuer une analyse de cartographie et de classification sur ces domaines.
- **Les analyses de cartographie et de classification** fournissent une analyse approfondie de vos données.

Pour plus de détails sur les différences entre les analyses de cartographie et de classification, voir"Quelle est

la différence entre les analyses de cartographie et de classification ?" .

Informations catégorisées par la classification des données

La classification des données collecte, indexe et attribue des catégories aux données suivantes :

- **Métadonnées standard** sur les fichiers : le type de fichier, sa taille, ses dates de création et de modification, etc.
- **Données personnelles** : Informations personnelles identifiables (PII) telles que les adresses e-mail, les numéros d'identification ou les numéros de carte de crédit, que Data Classification identifie à l'aide de mots, de chaînes et de modèles spécifiques dans les fichiers. "En savoir plus sur les données personnelles" .
- **Données personnelles sensibles** : Types particuliers d'informations personnelles sensibles (IPS), telles que les données de santé, l'origine ethnique ou les opinions politiques, telles que définies par le Règlement général sur la protection des données (RGPD) et d'autres réglementations sur la confidentialité. "En savoir plus sur les données personnelles sensibles" .
- **Catégories** : La classification des données prend les données numérisées et les divise en différents types de catégories. Les catégories sont des sujets basés sur l'analyse par l'IA du contenu et des métadonnées de chaque fichier. "En savoir plus sur les catégories".
- **Reconnaissance d'entité de nom** : la classification des données utilise l'IA pour extraire les noms naturels des personnes à partir de documents. "En savoir plus sur la réponse aux demandes d'accès aux données des personnes concernées" .

Présentation du réseau

Data Classification déploie un serveur unique, ou cluster, où vous le souhaitez : dans le cloud ou sur site. Les serveurs se connectent via des protocoles standard aux sources de données et indexent les résultats dans un cluster Elasticsearch, qui est également déployé sur les mêmes serveurs. Cela permet la prise en charge des environnements multicloud, cross-cloud, cloud privé et sur site.

La console déploie l'instance de classification des données avec un groupe de sécurité qui active les connexions HTTP entrantes à partir de l'agent de la console.

Lorsque vous utilisez la console en mode SaaS, la connexion à la console est effectuée via HTTPS et les données privées envoyées entre votre navigateur et l'instance de classification des données sont sécurisées par un cryptage de bout en bout à l'aide de TLS 1.2, ce qui signifie que NetApp et des tiers ne peuvent pas les lire.

Les règles sortantes sont complètement ouvertes. Un accès Internet est nécessaire pour installer et mettre à niveau le logiciel de classification des données et pour envoyer des mesures d'utilisation.

Si vous avez des exigences réseau strictes,"en savoir plus sur les points de terminaison contactés par la classification des données" .

Accéder à la NetApp Data Classification

Vous pouvez accéder à la NetApp Data Classification via la NetApp Console.

Pour vous connecter à la console, vous pouvez utiliser vos informations d'identification du site de support NetApp ou vous inscrire pour une connexion à la NetApp Console à l'aide de votre e-mail et d'un mot de passe. "En savoir plus sur la connexion à la console" .

Des tâches spécifiques nécessitent des rôles d'utilisateur de console spécifiques. "En savoir plus sur les rôles d'accès à la console pour tous les services" .

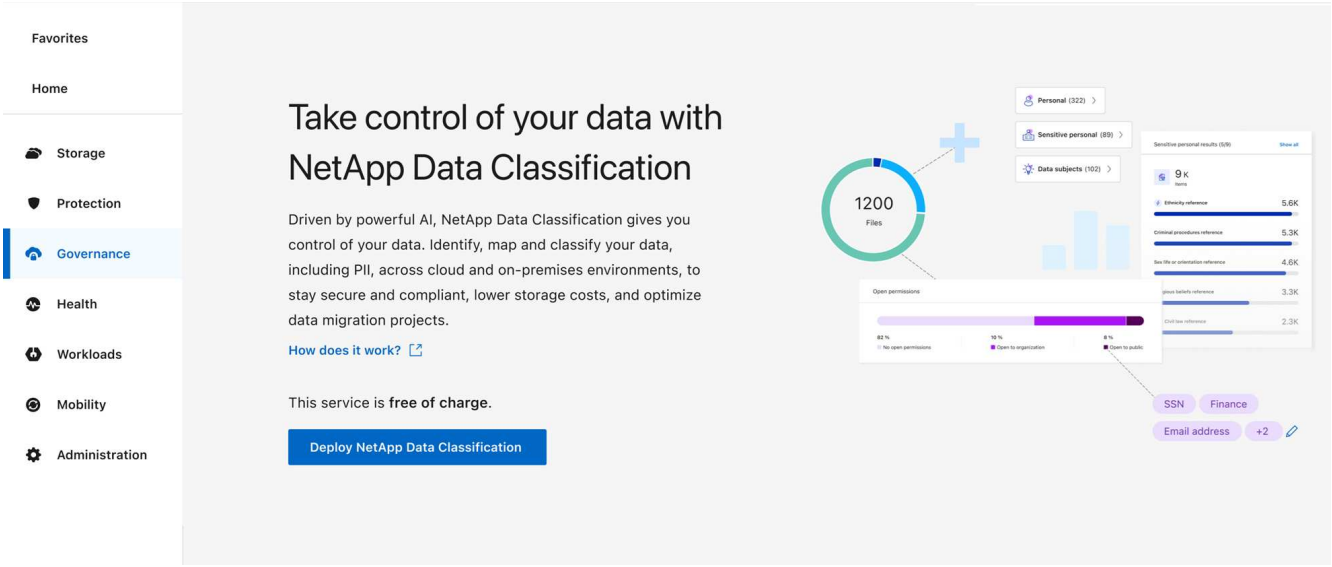
Avant de commencer

- "Vous devez ajouter un agent de console."
- "Comprendre quel style de déploiement de classification des données convient à votre charge de travail."

Étapes

1. Dans un navigateur Web, accédez à l'"Console" .
2. Connectez-vous à la console.
3. Depuis la page principale de la NetApp Console, sélectionnez **Gouvernance** > **Classification des données**.
4. Si c'est la première fois que vous accédez à la classification des données, la page de destination apparaît.

   Sélectionnez **Déployer la classification sur site ou dans le cloud** pour commencer à déployer votre instance de classification. Pour plus d'informations, voir"Quel déploiement de classification des données devez-vous utiliser ?"



   Sinon, le tableau de bord de classification des données s'affiche.

Déployer la classification des données

Quel déploiement de NetApp Data Classification devez-vous utiliser ?

Vous pouvez déployer NetApp Data Classification de différentes manières. Découvrez quelle méthode répond à vos besoins.

La classification des données peut être déployée des manières suivantes :

- "Déployer dans le cloud à l'aide de la console" . La console déploie l'instance de classification des données dans le même réseau de fournisseur de cloud que l'agent de la console.
- "Installer sur un hôte Linux avec accès Internet" . Installez Data Classification sur un hôte Linux de votre réseau ou sur un hôte Linux dans le cloud, disposant d'un accès Internet. Ce type d'installation peut être

une bonne option si vous préférez analyser les systèmes ONTAP sur site à l'aide d'une instance de classification des données également située sur site, bien que cela ne soit pas une exigence.

- "Installer sur un hôte Linux dans un site sur site sans accès Internet", également connu sous le nom de *mode privé*. Ce type d'installation, qui utilise un script d'installation, n'a aucune connectivité à la couche SaaS de la console.

> Le mode privé BlueXP (interface BlueXP héritée) est généralement utilisé avec des environnements locaux qui n'ont pas de connexion Internet et avec des régions cloud sécurisées, notamment AWS Secret Cloud, AWS Top Secret Cloud et Azure IL6. NetApp continue de prendre en charge ces environnements avec l'interface BlueXP héritée. Pour la documentation du mode privé dans l'ancienne interface BlueXP , voir"Documentation PDF pour le mode privé BlueXP" .

L'installation sur un hôte Linux avec accès Internet et l'installation sur site sur un hôte Linux sans accès Internet utilisent un script d'installation. Le script commence par vérifier si le système et l'environnement répondent aux prérequis. Si les conditions préalables sont remplies, l'installation démarre. Si vous souhaitez vérifier les conditions préalables indépendamment de l'exécution de l'installation de la classification des données, vous pouvez télécharger un progiciel distinct qui teste uniquement les conditions préalables.

"Vérifiez que votre hôte Linux est prêt à installer la classification des données" .

Déployer la NetApp Data Classification dans le cloud à l'aide de la NetApp Console

Vous pouvez déployer NetApp Data Classification dans le cloud avec la NetApp Console. La console déploie l'instance de classification des données dans le même réseau de fournisseur de cloud que l'agent de la console.

Notez que vous pouvez également"installer Data Classification sur un hôte Linux disposant d'un accès Internet" . Ce type d'installation peut être une bonne option si vous préférez analyser les systèmes ONTAP sur site à l'aide d'une instance de classification des données également située sur site, mais ce n'est pas une exigence. Le logiciel fonctionne exactement de la même manière, quelle que soit la méthode d'installation choisie.

Démarrage rapide

Commencez rapidement en suivant ces étapes ou faites défiler les sections restantes pour obtenir tous les détails.

Créer un agent de console

Si vous n'avez pas encore d'agent de console, créez-en un. Voir"création d'un agent de console dans AWS" ,"création d'un agent de console dans Azure" , ou"création d'un agent de console dans GCP" .

Vous pouvez également"installer l'agent de console sur site" sur un serveur Linux de votre réseau ou sur un serveur Linux dans le cloud.

Prérequis

Assurez-vous que votre environnement répond aux prérequis. Cela inclut un accès Internet sortant pour l'instance, une connectivité entre l'agent Console et Data Classification via le port 443, et plus encore. Voir la liste complète.

3 Déployer la classification des données

Lancez l'assistant d'installation pour déployer l'instance de classification des données dans le cloud.

Créer un agent de console

Si vous ne disposez pas encore d'un agent de console, créez un agent de console chez votre fournisseur de cloud. Voir "création d'un agent de console dans AWS" ou "création d'un agent de console dans Azure" , ou "création d'un agent de console dans GCP" . Dans la plupart des cas, vous devrez probablement configurer un agent de console avant de tenter d'activer la classification des données, car la plupart "Les fonctionnalités de la console nécessitent un agent de console" mais il existe des cas où vous devrez en configurer un dès maintenant.

Il existe certains scénarios dans lesquels vous devez utiliser un agent de console déployé chez un fournisseur de cloud spécifique :

- Lors de l'analyse des données dans Cloud Volumes ONTAP dans AWS ou Amazon FSx pour les compartiments ONTAP , vous utilisez un agent de console dans AWS.
- Lors de l'analyse des données dans Cloud Volumes ONTAP dans Azure ou dans Azure NetApp Files, vous utilisez un agent de console dans Azure.
  - Pour Azure NetApp Files, il doit être déployé dans la même région que les volumes que vous souhaitez analyser.
- Lors de l'analyse des données dans Cloud Volumes ONTAP dans GCP, vous utilisez un agent de console dans GCP.

Les systèmes ONTAP sur site, les partages de fichiers NetApp et les bases de données peuvent être analysés lors de l'utilisation de l'un de ces agents de console cloud.

Notez que vous pouvez également "installer l'agent de console sur site" sur un serveur Linux de votre réseau ou dans le cloud. Certains utilisateurs prévoyant d'installer la classification des données sur site peuvent également choisir d'installer l'agent de console sur site.

Il peut y avoir des situations où vous devez utiliser "plusieurs agents de console" .

> La classification des données n'impose pas de limite à la quantité de données qu'elle peut analyser. Chaque agent de console prend en charge l'analyse et l'affichage de 500 Tio de données. Pour scanner plus de 500 Tio de données,"installer un autre agent de console" alors"déployer une autre instance de classification des données" . + L'interface utilisateur de la console affiche les données d'un seul connecteur. Pour obtenir des conseils sur l'affichage des données de plusieurs agents de console, consultez"Travailler avec plusieurs agents de console" .

Soutien gouvernemental régional

La classification des données est prise en charge lorsque l'agent de console est déployé dans une région gouvernementale (AWS GovCloud, Azure Gov ou Azure DoD). Lorsqu'elle est déployée de cette manière, la classification des données présente les restrictions suivantes :

"Découvrez comment déployer l'agent Console dans une région gouvernementale".

Prérequis

Passez en revue les conditions préalables suivantes pour vous assurer que vous disposez d'une configuration prise en charge avant de déployer la classification des données dans le cloud. Lorsque vous déployez la classification des données dans le cloud, elle est située dans le même sous-réseau que l'agent de la console.

Activer l'accès Internet sortant à partir de la classification des données

La classification des données nécessite un accès Internet sortant. Si votre réseau virtuel ou physique utilise un serveur proxy pour l'accès à Internet, assurez-vous que l'instance de classification des données dispose d'un accès Internet sortant pour contacter les points de terminaison suivants. Le proxy doit être non transparent. Les proxys transparents ne sont actuellement pas pris en charge.

Consultez le tableau approprié ci-dessous selon que vous déployez la classification des données dans AWS, Azure ou GCP.

Points de terminaison requis pour AWS

Points de terminaison	But
\ https://api.console.netapp.com	Communication avec le service Console, qui inclut les comptes NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Communication avec le site Web de la console pour l'authentification centralisée des utilisateurs.
\ https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	Fournit un accès aux images logicielles, aux manifestes et aux modèles.
\ https://kinesis.us-east-1.amazonaws.com	Permet à NetApp de diffuser des données à partir des enregistrements d'audit.
\ https://cognito-idp.us-east-1.amazonaws.com \ https://cognito-identity.us-east-1.amazonaws.com \ https://user-feedback-store-prod.s3.us-west-2.amazonaws.com \ https://customer-data-production.s3.us-west-2.amazonaws.com	Permet à la classification des données d'accéder et de télécharger des manifestes et des modèles, et d'envoyer des journaux et des métriques.

Points de terminaison requis pour Azure

Points de terminaison	But
\ https://api.console.netapp.com	Communication avec le service Console, qui inclut les comptes NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Communication avec le site Web de la console pour l'authentification centralisée des utilisateurs.
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	Fournit un accès aux images logicielles, aux manifestes, aux modèles et permet d'envoyer des journaux et des métriques.
\ https://support.compliance.api.console.netapp.com/	Permet à NetApp de diffuser des données à partir des enregistrements d'audit.

Points de terminaison requis pour GCP

Points de terminaison	But
\ https://api.console.netapp.com	Communication avec le service Console, qui inclut les comptes NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Communication avec le site Web de la console pour l'authentification centralisée des utilisateurs.

Points de terminaison	But
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	Fournit un accès aux images logicielles, aux manifestes, aux modèles et permet d'envoyer des journaux et des métriques.
\ https://support.compliance.api.console.netapp.com/	Permet à NetApp de diffuser des données à partir des enregistrements d'audit.

Assurez-vous que la classification des données dispose des autorisations requises

Assurez-vous que Data Classification dispose des autorisations nécessaires pour déployer des ressources et créer des groupes de sécurité pour l'instance Data Classification.

- "Autorisations Google Cloud"
- "Autorisations AWS"
- "Autorisations Azure"

Assurez-vous que l'agent de la console peut accéder à la classification des données

Assurez la connectivité entre l'agent de console et l'instance de classification des données. Le groupe de sécurité de l'agent de console doit autoriser le trafic entrant et sortant sur le port 443 vers et depuis l'instance de classification des données. Cette connexion permet le déploiement de l'instance de classification des données et vous permet d'afficher les informations dans les onglets Conformité et Gouvernance. La classification des données est prise en charge dans les régions gouvernementales dans AWS et Azure.

Des règles de groupe de sécurité entrantes et sortantes supplémentaires sont requises pour les déploiements AWS et AWS GovCloud. Voir "Règles pour l'agent de console dans AWS" pour plus de détails.

Des règles de groupe de sécurité entrantes et sortantes supplémentaires sont requises pour les déploiements Azure et Azure Government. Voir "Règles pour l'agent de console dans Azure" pour plus de détails.

Assurez-vous de pouvoir maintenir la classification des données en cours d'exécution

L'instance de classification des données doit rester active pour analyser en continu vos données.

Assurer la connectivité du navigateur Web à la classification des données

Une fois la classification des données activée, assurez-vous que les utilisateurs accèdent à l'interface de la console à partir d'un hôte disposant d'une connexion à l'instance de classification des données.

L'instance de classification des données utilise une adresse IP privée pour garantir que les données indexées ne sont pas accessibles sur Internet. Par conséquent, le navigateur Web que vous utilisez pour accéder à la console doit disposer d'une connexion à cette adresse IP privée. Cette connexion peut provenir d'une connexion directe à votre fournisseur de cloud (par exemple, un VPN) ou d'un hôte situé sur le même réseau que l'instance de classification des données.

Vérifiez vos limites de vCPU

Assurez-vous que la limite vCPU de votre fournisseur de cloud permet le déploiement d'une instance avec le nombre de cœurs nécessaire. Vous devrez vérifier la limite de vCPU pour la famille d'instances concernée dans la région où la console s'exécute. "Voir les types d'instances requis" .

Consultez les liens suivants pour plus de détails sur les limites du vCPU :

- "Documentation AWS : quotas de service Amazon EC2"
- "Documentation Azure : Quotas de processeurs virtuels pour machines virtuelles"
- "Documentation Google Cloud : quotas de ressources"

Déployer la classification des données dans le cloud

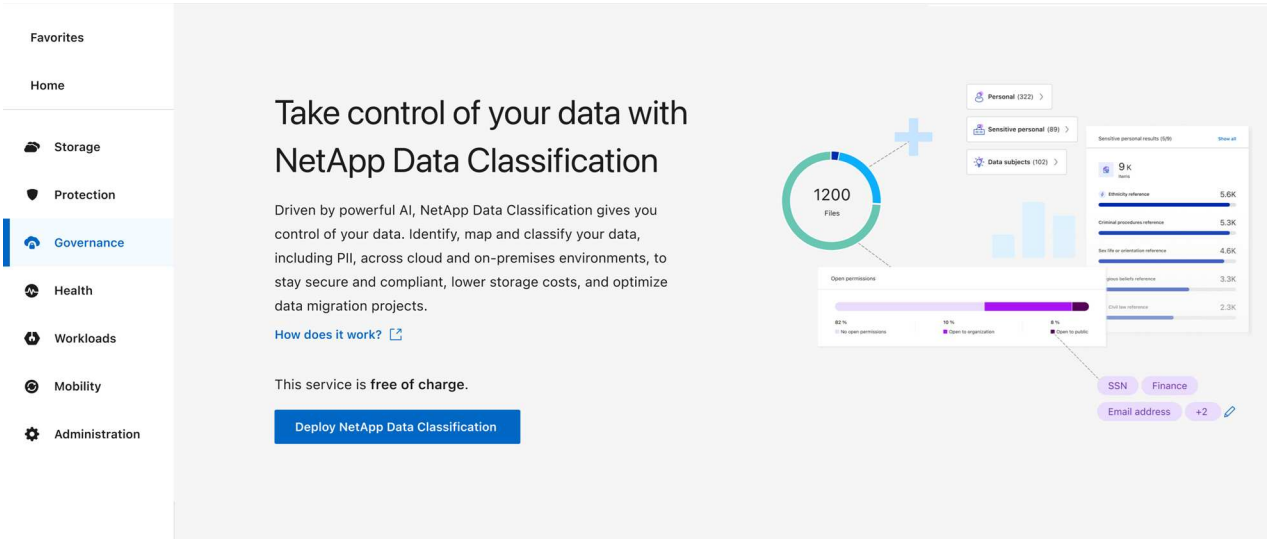
Suivez ces étapes pour déployer une instance de classification des données dans le cloud. L'agent de console déploiera l'instance dans le cloud, puis installera le logiciel de classification des données sur cette instance.

Dans les régions où le type d'instance par défaut n'est pas disponible, la classification des données s'exécute sur un"type d'instance alternatif" .

Déployer dans AWS

Étapes

1. Depuis la page principale de la classification des données, sélectionnez **Déployer la classification sur site ou dans le cloud**.

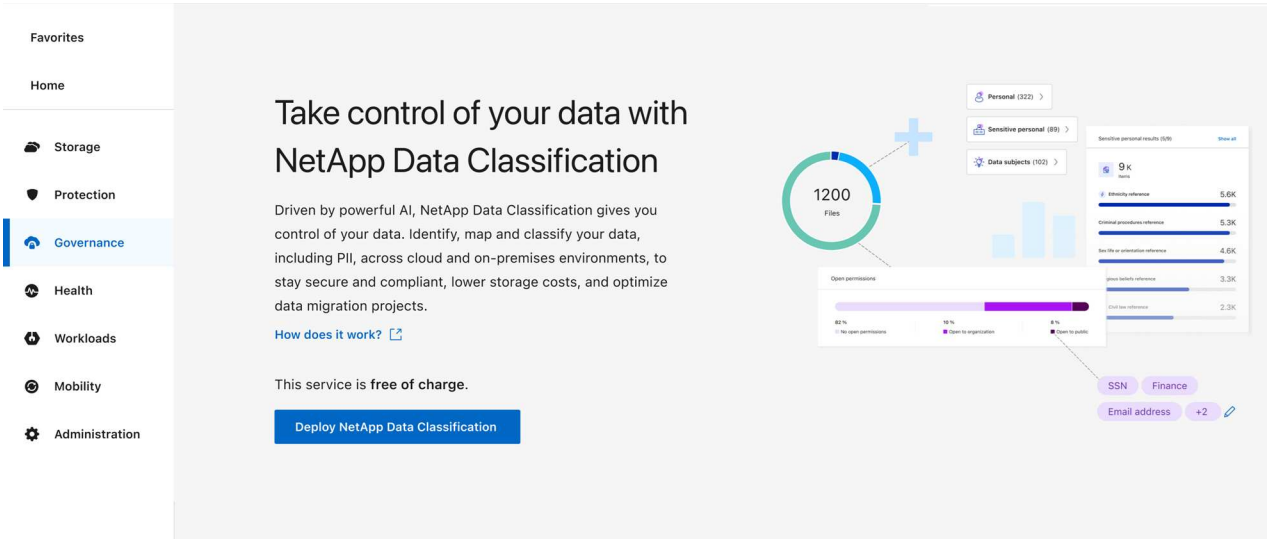


2. Depuis la page *Installation*, sélectionnez **Déployer > Déployer** pour utiliser la taille d'instance « Grande » et démarrer l'assistant de déploiement cloud.
3. L'assistant affiche la progression au fur et à mesure des étapes de déploiement. Lorsque des entrées sont requises ou si vous rencontrez des problèmes, vous êtes invité à le faire.
4. Une fois l'instance déployée et la classification des données installée, sélectionnez **Continuer vers la configuration** pour accéder à la page *Configuration*.

Déployer dans Azure

Étapes

1. Depuis la page principale de la classification des données, sélectionnez **Déployer la classification sur site ou dans le cloud**.



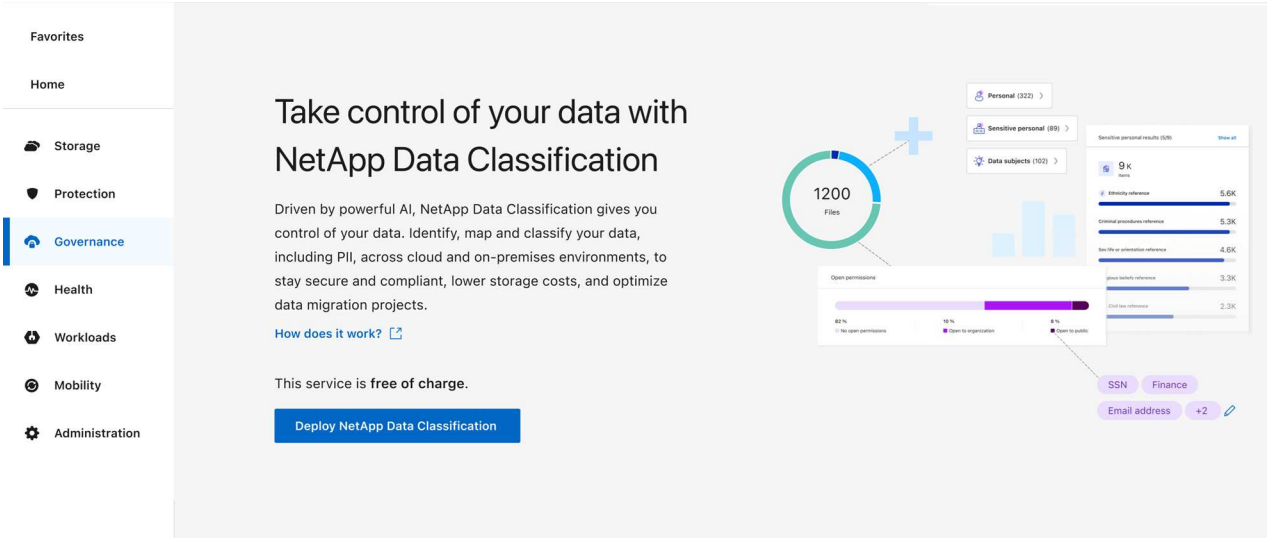
2. Sélectionnez **Déployer** pour démarrer l'assistant de déploiement cloud.

3. L'assistant affiche la progression au fur et à mesure des étapes de déploiement. Il s'arrêtera et demandera une saisie s'il rencontre des problèmes.
4. Une fois l'instance déployée et la classification des données installée, sélectionnez **Continuer vers la configuration** pour accéder à la page *Configuration*.

Déployer dans Google Cloud

Étapes

1. Depuis la page principale de la classification des données, sélectionnez **Gouvernance > Classification**.
2. Sélectionnez **Déployer la classification sur site ou dans le cloud**.



3. Sélectionnez **Déployer** pour démarrer l'assistant de déploiement cloud.
4. L'assistant affiche la progression au fur et à mesure des étapes de déploiement. Il s'arrêtera et demandera une saisie s'il rencontre des problèmes.
5. Une fois l'instance déployée et la classification des données installée, sélectionnez **Continuer vers la configuration** pour accéder à la page *Configuration*.

Résultat

La console déploie l'instance de classification des données dans votre fournisseur de cloud.

Les mises à niveau de l'agent de console et du logiciel de classification des données sont automatisées tant que les instances disposent d'une connectivité Internet.

Et ensuite?

Depuis la page Configuration, vous pouvez sélectionner les sources de données que vous souhaitez analyser.

Installer NetApp Data Classification sur un hôte disposant d'un accès Internet

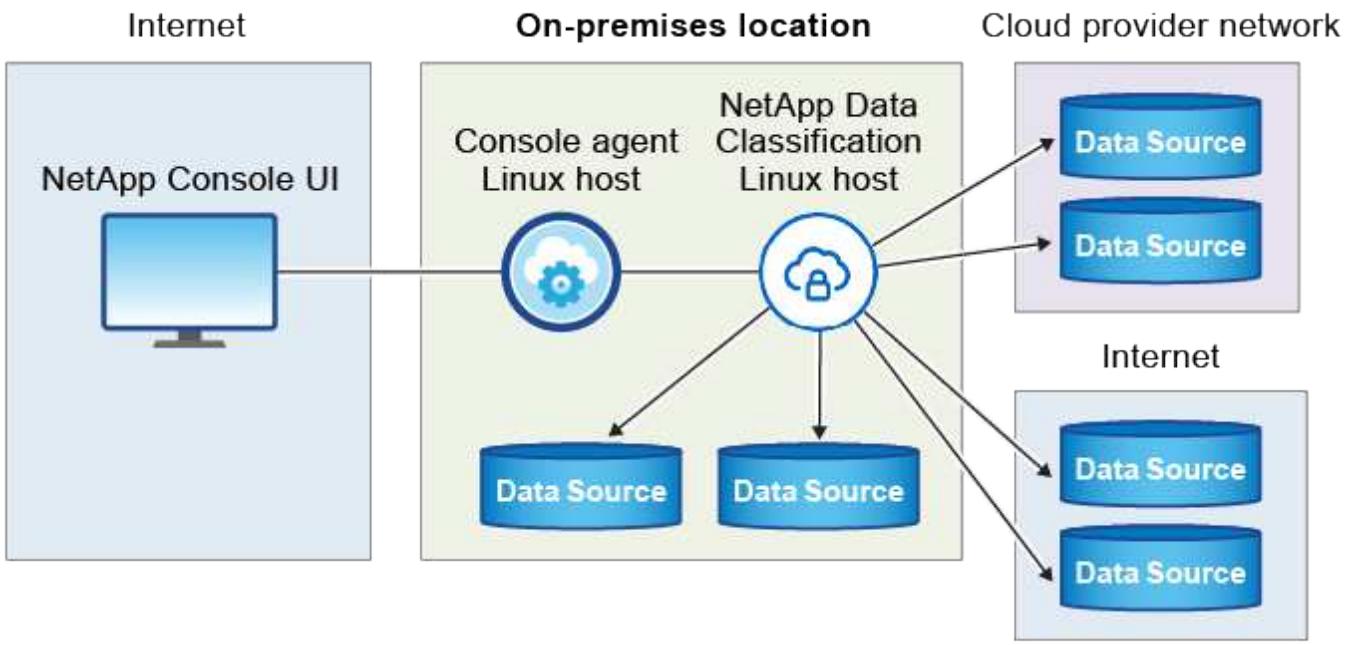
Pour déployer NetApp Data Classification sur un hôte Linux de votre réseau ou sur un hôte Linux dans le cloud disposant d'un accès Internet, vous devez déployer l'hôte Linux manuellement sur votre réseau ou dans le cloud.

L'installation sur site est une bonne option si vous préférez analyser les systèmes ONTAP sur site à l'aide

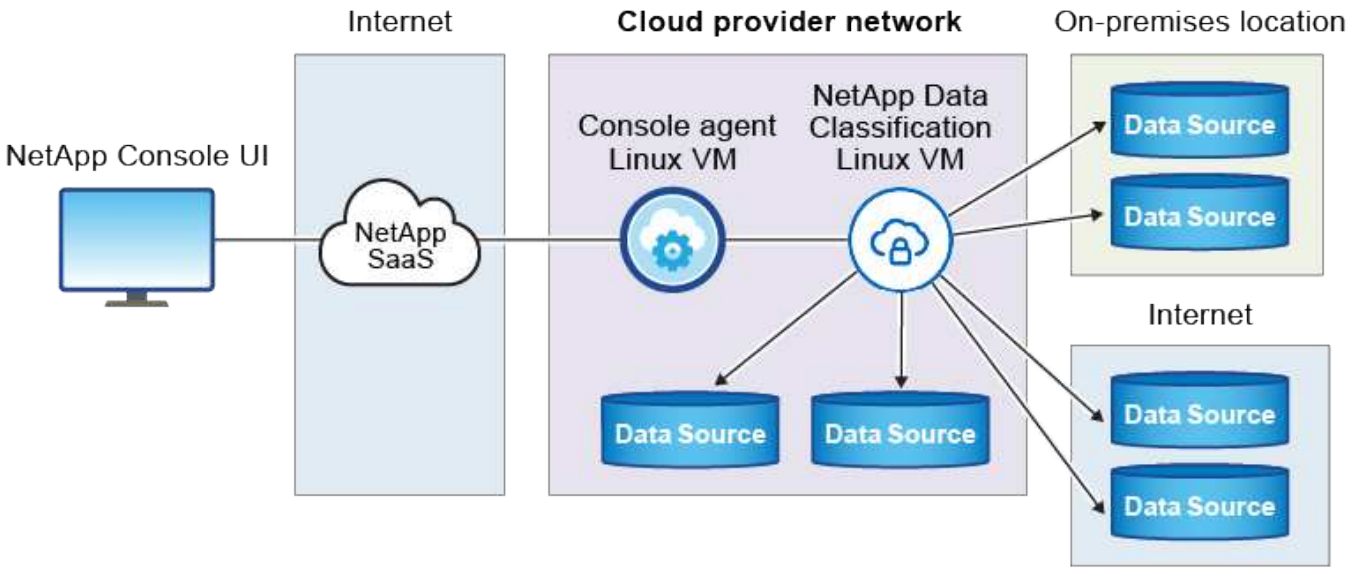
d'une instance de classification des données également située sur site. Ce n'est pas une exigence. Le logiciel fonctionne de la même manière quelle que soit la méthode d'installation choisie.

Le script d'installation de la classification des données commence par vérifier si le système et l'environnement répondent aux prérequis requis. Si toutes les conditions préalables sont remplies, l'installation démarre. Si vous souhaitez vérifier les conditions préalables indépendamment de l'exécution de l'installation de la classification des données, vous pouvez télécharger un progiciel distinct qui teste uniquement les conditions préalables. "Découvrez comment vérifier si votre hôte Linux est prêt à installer la classification des données" .

L'installation typique sur un hôte Linux *dans vos locaux* comporte les composants et connexions suivants.



L'installation typique sur un hôte Linux *dans le cloud* comporte les composants et connexions suivants.



Démarrage rapide

Commencez rapidement en suivant ces étapes ou faites défiler les sections restantes pour obtenir tous les

détails.

1 Créer un agent de console

Si vous n'avez pas encore d'agent de console, "déployer l'agent de console sur site" sur un hôte Linux de votre réseau ou sur un hôte Linux dans le cloud.

Vous pouvez également créer un agent de console avec votre fournisseur de cloud. Voir "création d'un agent de console dans AWS" , "création d'un agent de console dans Azure" , ou "création d'un agent de console dans GCP" .

2 Réviser les prérequis

Assurez-vous que votre environnement peut répondre aux prérequis. Cela inclut l'accès Internet sortant pour l'instance, la connectivité entre l'agent de console et la classification des données via le port 443, et bien plus encore. Voir la liste complète .

Vous avez également besoin d'un système Linux qui répond auxexigences suivantes .

3 Téléchargez et déployez la classification des données

Téléchargez le logiciel Cloud Data Classification à partir du site de support NetApp et copiez le fichier d'installation sur l'hôte Linux que vous prévoyez d'utiliser. Lancez ensuite l'assistant d'installation et suivez les invites pour déployer l'instance de classification des données.

Créer un agent de console

Un agent de console est requis avant de pouvoir installer et utiliser la classification des données. Dans la plupart des cas, vous aurez probablement configuré un agent de console avant de tenter d'activer la classification des données, car la plupart "Les fonctionnalités de la console nécessitent un agent de console" , mais il y a des cas où vous devrez en créer un maintenant.

Pour en créer un dans votre environnement de fournisseur de cloud, consultez "création d'un agent de console dans AWS" , "création d'un agent de console dans Azure" , ou "création d'un agent de console dans GCP" .

Il existe certains scénarios dans lesquels vous devez utiliser un agent de console déployé chez un fournisseur de cloud spécifique :

- Lors de l'analyse des données dans Cloud Volumes ONTAP dans AWS ou Amazon FSx for ONTAP, vous utilisez un agent de console dans AWS.
- Lors de l'analyse des données dans Cloud Volumes ONTAP dans Azure ou dans Azure NetApp Files, vous utilisez un agent de console dans Azure.

  Pour Azure NetApp Files, il doit être déployé dans la même région que les volumes que vous souhaitez analyser.

- Lors de l'analyse des données dans Cloud Volumes ONTAP dans GCP, vous utilisez un agent de console dans GCP.

Les systèmes ONTAP sur site, les partages de fichiers NetApp et les comptes de base de données peuvent être analysés à l'aide de l'un de ces agents de console cloud.

Notez que vous pouvez également "déployer l'agent de console sur site" sur un hôte Linux de votre réseau ou sur un hôte Linux dans le cloud. Certains utilisateurs prévoyant d'installer la classification des données sur site peuvent également choisir d'installer l'agent de console sur site.

Vous aurez besoin de l'adresse IP ou du nom d'hôte du système d'agent de la console lors de l'installation de la classification des données. Vous disposerez de ces informations si vous avez installé l'agent Console dans vos locaux. Si l'agent de la console est déployé dans le cloud, vous pouvez trouver ces informations depuis la console : sélectionnez l'icône Aide puis **Support** puis **Agent de la console**.

Préparer le système hôte Linux

Le logiciel de classification des données doit s'exécuter sur un hôte qui répond aux exigences spécifiques du système d'exploitation, aux exigences de RAM, aux exigences logicielles, etc. L'hôte Linux peut être dans votre réseau ou dans le cloud.

Assurez-vous de pouvoir maintenir la classification des données en cours d'exécution. La machine de classification des données doit rester allumée pour analyser en continu vos données.

- La classification des données doit être hébergée sur un serveur dédié. L'hôte ne peut pas être partagé avec d'autres applications ou logiciels tiers tels que les antivirus.
- Choisissez la taille qui correspond à l'ensemble de données que vous prévoyez d'analyser avec la classification des données.

Taille du système	processeur	RAM (la mémoire d'échange doit être désactivée)	Disque
Extra Large	32 processeurs	128 Go de RAM	• 1 Tio SSD sur /, ou 100 Gio disponibles sur /opt • 895 Gio disponibles sur /var/lib/docker • 5 Gio sur /tmp • **Pour Podman, 30 Go sur /var/tmp**
Grand	16 processeurs	64 Go de RAM	• 500 Gio SSD sur /, ou 100 Gio disponibles sur /opt • 400 Gio disponibles sur /var/lib/docker ou pour Podman /var/lib/containers • 5 Gio sur /tmp • **Pour Podman, 30 Go sur /var/tmp**

- Lors du déploiement d'une instance de calcul dans le cloud pour votre installation de classification des données, il est recommandé d'utiliser un système qui répond aux exigences système « Large » ci-dessus :
  - **Type d'instance Amazon Elastic Compute Cloud (Amazon EC2)** : « m6i.4xlarge ». "Voir d'autres types d'instances AWS" .
  - **Taille de la machine virtuelle Azure** : « Standard_D16s_v3 ». "Voir d'autres types d'instances Azure" .

- **Type de machine GCP** : « n2-standard-16 ». "Voir les types d'instances GCP supplémentaires" .

- **Autorisations de dossier UNIX** : Les autorisations UNIX minimales suivantes sont requises :

Dossier	autorisations minimales
/tmp	`rwxrwxrwt`
/opter	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/système	`rwxr-xr-x`

- **Système opérateur**:
  - Les systèmes d'exploitation suivants nécessitent l'utilisation du moteur de conteneur Docker :
    - Red Hat Enterprise Linux versions 7.8 et 7.9
    - Ubuntu 22.04 (nécessite la version 1.23 ou supérieure de Data Classification)
    - Ubuntu 24.04 (nécessite la version 1.23 ou supérieure de Data Classification)
  - Les systèmes d'exploitation suivants nécessitent l'utilisation du moteur de conteneur Podman et nécessitent la version 1.30 ou supérieure de Data Classification :
    - Red Hat Enterprise Linux versions 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 et 9.6.
  - Les extensions vectorielles avancées (AVX2) doivent être activées sur le système hôte.
- **Red Hat Subscription Management** : L'hôte doit être enregistré auprès de Red Hat Subscription Management. S'il n'est pas enregistré, le système ne peut pas accéder aux référentiels pour mettre à jour les logiciels tiers requis lors de l'installation.
- **Logiciel supplémentaire** : Vous devez installer le logiciel suivant sur l'hôte avant d'installer Data Classification :
  - Selon le système d'exploitation que vous utilisez, vous devez installer l'un des moteurs de conteneurs :
    - Docker Engine version 19.3.1 ou supérieure. "Voir les instructions d'installation" .
    - Podman version 4 ou supérieure. Pour installer Podman, entrez(`sudo yum install podman netavark -y`).
- Version Python 3.6 ou supérieure. "Voir les instructions d'installation" .
  - **Considérations NTP** : NetApp recommande de configurer le système de classification des données pour utiliser un service NTP (Network Time Protocol). L'heure doit être synchronisée entre le système de classification des données et le système d'agent de la console.
- **Considérations relatives au pare-feu** : Si vous envisagez d'utiliser `firewalld` , nous vous recommandons de l'activer avant d'installer Data Classification. Exécutez les commandes suivantes pour configurer `firewalld` afin qu'il soit compatible avec la classification des données :

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

Si vous prévoyez d'utiliser des hôtes de classification de données supplémentaires comme nœuds de scanner, ajoutez ces règles à votre système principal à ce stade :

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Notez que vous devez redémarrer Docker ou Podman chaque fois que vous activez ou mettez à jour `firewalld` paramètres.

L'adresse IP du système hôte de classification des données ne peut pas être modifiée après l'installation.

Activer l'accès Internet sortant à partir de la classification des données

La classification des données nécessite un accès Internet sortant. Si votre réseau virtuel ou physique utilise un serveur proxy pour l'accès à Internet, assurez-vous que l'instance de classification des données dispose d'un accès Internet sortant pour contacter les points de terminaison suivants.

Points de terminaison	But
\ https://api.console.netapp.com	Communication avec la console, qui inclut les comptes NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Communication avec le site Web de la console pour l'authentification centralisée des utilisateurs.
\ https://support.compliance.api.bluexp.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	Fournit un accès aux images logicielles, aux manifestes, aux modèles et permet d'envoyer des journaux et des métriques.
https://support.compliance.api.bluexp.netapp.com/	Permet à NetApp de diffuser des données à partir des enregistrements d'audit.
\ https://github.com/docker \ https://download.docker.com	Fournit des packages prérequis pour l'installation de Docker.
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Fournit des packages prérequis pour l'installation d'Ubuntu.

Vérifiez que tous les ports requis sont activés

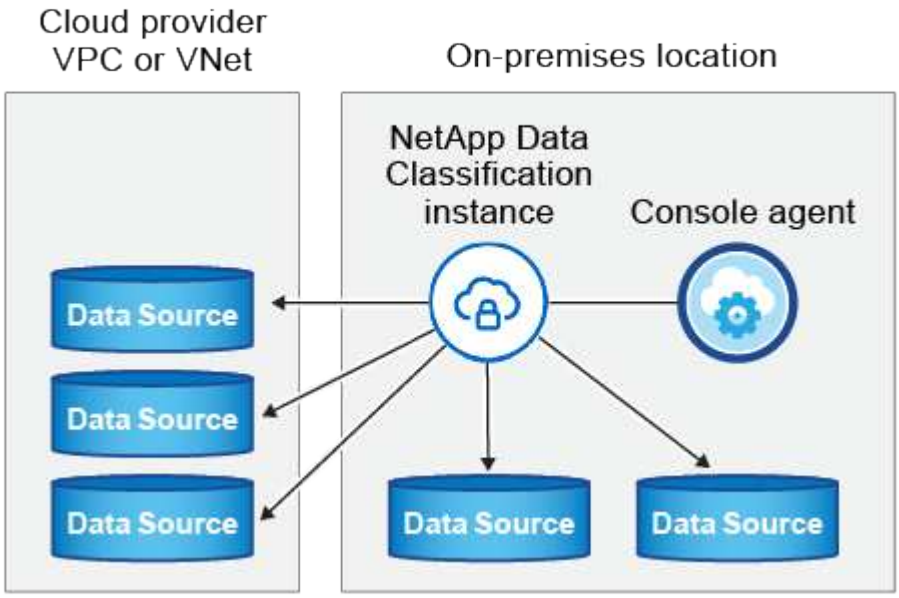
Vous devez vous assurer que tous les ports requis sont ouverts pour la communication entre l'agent de console, la classification des données, Active Directory et vos sources de données.

Type de connexion	Ports	Description
Agent de console <> Classification des données	8080 (TCP), 443 (TCP) et 80. 9000	Les règles de pare-feu ou de routage de l'agent de console doivent autoriser le trafic entrant et sortant sur le port 443 vers et depuis l'instance de classification des données. Assurez-vous que le port 8080 est ouvert afin de pouvoir voir la progression de l'installation dans la console. Si un pare-feu est utilisé sur l'hôte Linux, le port 9000 est requis pour les processus internes au sein d'un serveur Ubuntu.
Agent de console <> cluster ONTAP (NAS)	443 (TCP)	La console découvre les clusters ONTAP à l'aide de HTTPS. Si vous utilisez des stratégies de pare-feu personnalisées, elles doivent répondre aux exigences suivantes : • L'hôte de l'agent de console doit autoriser l'accès HTTPS sortant via le port 443. Si l'agent de la console est dans le cloud, toutes les communications sortantes sont autorisées par le pare-feu prédéfini ou les règles de routage. • Le cluster ONTAP doit autoriser l'accès HTTPS entrant via le port 443. La stratégie de pare-feu « mgmt » par défaut autorise l'accès HTTPS entrant à partir de toutes les adresses IP. Si vous avez modifié cette politique par défaut ou si vous avez créé votre propre politique de pare-feu, vous devez associer le protocole HTTPS à cette politique et activer l'accès à partir de l'hôte de l'agent de la console.
Classification des données <> cluster ONTAP	• Pour NFS - 111 (TCP\UDP) et 2049 (TCP\UDP) • Pour CIFS - 139 (TCP\UDP) et 445 (TCP\UDP)	La classification des données nécessite une connexion réseau à chaque sous-réseau Cloud Volumes ONTAP ou à un système ONTAP sur site. Les pare-feu ou les règles de routage pour Cloud Volumes ONTAP doivent autoriser les connexions entrantes à partir de l'instance de classification des données. Assurez-vous que ces ports sont ouverts à l'instance de classification des données : • Pour NFS - 111 et 2049 • Pour CIFS - 139 et 445 Les stratégies d'exportation de volume NFS doivent autoriser l'accès à partir de l'instance de classification des données.

Type de connexion	Ports	Description
Classification des données <> Active Directory	389 (TCP et UDP), 636 (TCP), 3268 (TCP) et 3269 (TCP)	Vous devez déjà avoir un Active Directory configuré pour les utilisateurs de votre entreprise. De plus, la classification des données nécessite des informations d'identification Active Directory pour analyser les volumes CIFS. Vous devez disposer des informations pour Active Directory : • Adresse IP du serveur DNS ou plusieurs adresses IP • Nom d'utilisateur et mot de passe pour le serveur • Nom de domaine (nom Active Directory) • Que vous utilisiez ou non un LDAP sécurisé (LDAPS) • Port du serveur LDAP (généralement 389 pour LDAP et 636 pour LDAP sécurisé)

Installer la classification des données sur l'hôte Linux

Pour les configurations typiques, vous installerez le logiciel sur un seul système hôte. Voir ces étapes ici .



VoirPréparation du système hôte Linux etRévision des prérequis pour obtenir la liste complète des exigences avant de déployer la classification des données.

Les mises à niveau du logiciel de classification des données sont automatisées tant que l'instance dispose d'une connexion Internet.

La classification des données ne peut actuellement pas analyser les compartiments S3, Azure NetApp Files ou FSx pour ONTAP lorsque le logiciel est installé sur site. Dans ces cas, vous devrez déployer un agent de console distinct et une instance de classification des données dans le cloud et "basculer entre les connecteurs" pour vos différentes sources de données.

Installation sur un seul hôte pour les configurations typiques

Passez en revue les exigences et suivez ces étapes lors de l'installation du logiciel de classification des données sur un seul hôte local.

"Regardez cette vidéo"pour voir comment installer Data Classification.

Notez que toutes les activités d'installation sont enregistrées lors de l'installation de Data Classification. Si vous rencontrez des problèmes lors de l'installation, vous pouvez afficher le contenu du journal d'audit d'installation. Il est écrit à `/opt/netapp/install_logs/` .

Avant de commencer

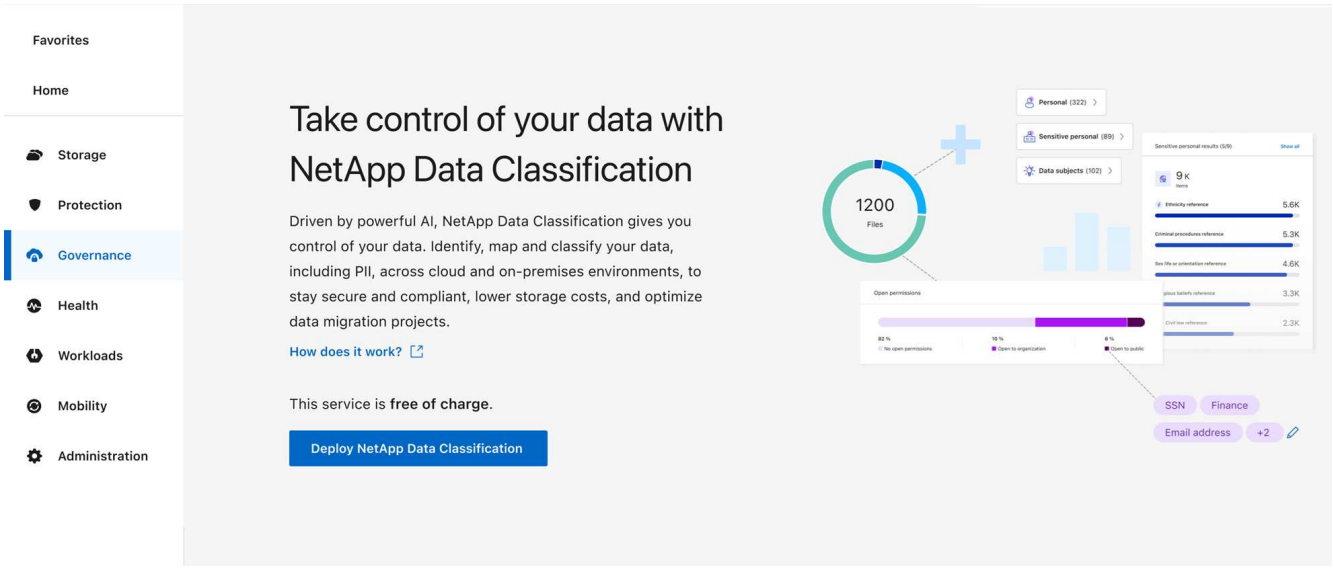
- Vérifiez que votre système Linux répond auxexigences de l'hôte .
- Vérifiez que le système dispose des deux packages logiciels prérequis installés (Docker Engine ou Podman et Python 3).
- Assurez-vous que vous disposez des privilèges root sur le système Linux.
- Si vous utilisez un proxy pour accéder à Internet :
  - Vous aurez besoin des informations du serveur proxy (adresse IP ou nom d'hôte, port de connexion, schéma de connexion : https ou http, nom d'utilisateur et mot de passe).
  - Si le proxy effectue une interception TLS, vous devez connaître le chemin sur le système Linux de classification des données où les certificats CA TLS sont stockés.
  - Le proxy doit être non transparent. La classification des données ne prend actuellement pas en charge les proxys transparents.
  - L'utilisateur doit être un utilisateur local. Les utilisateurs de domaine ne sont pas pris en charge.
- Vérifiez que votre environnement hors ligne répond aux exigences requisesautorisations et connectivité .

Étapes

1. Téléchargez le logiciel de classification des données à partir du "Site de support NetApp" . Le fichier que vous devez sélectionner est nommé **DATASENSE-INSTALLER-<version>.tar.gz**.
2. Copiez le fichier d'installation sur l'hôte Linux que vous prévoyez d'utiliser (en utilisant `scp` ou une autre méthode).
3. Décompressez le fichier d'installation sur la machine hôte, par exemple :

```
tar -xzf DATASENSE-INSTALLER-V1.25.0.tar.gz
```

4. Dans la console, sélectionnez **Gouvernance > Classification**.
5. Sélectionnez **Déployer la classification sur site ou dans le cloud**.



6. Selon que vous installez Data Classification sur une instance que vous avez préparée dans le cloud ou sur une instance que vous avez préparée dans vos locaux, sélectionnez l'option **Déployer** appropriée pour démarrer l'installation de Data Classification.
7. La boîte de dialogue *Déployer la classification des données sur site* s'affiche. Copiez la commande fournie (par exemple : `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq`) et collez-le dans un fichier texte pour pouvoir l'utiliser plus tard. Sélectionnez ensuite **Fermer** pour fermer la boîte de dialogue.
8. Sur la machine hôte, entrez la commande que vous avez copiée, puis suivez une série d'invites, ou vous pouvez fournir la commande complète, y compris tous les paramètres requis, comme arguments de ligne de commande.

   Notez que le programme d'installation effectue une pré-vérification pour s'assurer que votre système et vos exigences réseau sont en place pour une installation réussie. "Regardez cette vidéo" pour comprendre les messages et les implications du pré-contrôle.

Entrez les paramètres comme demandé :	Entrez la commande complète :
a. Collez la commande que vous avez copiée à l'étape 7 : `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` Si vous effectuez l'installation sur une instance cloud (pas dans vos locaux), ajoutez `--manual-cloud-install <cloud_provider>`. b. Saisissez l'adresse IP ou le nom d'hôte de la machine hôte de classification des données afin que le système d'agent de la console puisse y accéder. c. Saisissez l'adresse IP ou le nom d'hôte de la machine hôte de l'agent de console afin que le système de classification des données puisse y accéder. d. Saisissez les détails du proxy lorsque vous y êtes invité. Si votre agent de console utilise déjà un proxy, il n'est pas nécessaire de saisir à nouveau ces informations ici, car la classification des données utilisera automatiquement le proxy utilisé par l'agent de console.	Alternativement, vous pouvez créer la commande entière à l'avance, en fournissant les paramètres d'hôte et de proxy nécessaires : `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy-user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

Valeurs des variables :

- *account_id* = ID de compte NetApp
- *client_id* = ID client de l'agent de console (ajoutez le suffixe « clients » à l'ID client s'il n'est pas déjà présent)
- *user_token* = jeton d'accès utilisateur JWT
- *ds_host* = Adresse IP ou nom d'hôte du système Linux de classification des données.
- *cm_host* = Adresse IP ou nom d'hôte du système agent de la console.
- *cloud_provider* = Lors de l'installation sur une instance cloud, saisissez « AWS », « Azure » ou « Gcp » selon le fournisseur de cloud.
- *proxy_host* = IP ou nom d'hôte du serveur proxy si l'hôte est derrière un serveur proxy.
- *proxy_port* = Port de connexion au serveur proxy (par défaut 80).
- *proxy_scheme* = Schéma de connexion : https ou http (par défaut http).
- *proxy_user* = Utilisateur authentifié pour se connecter au serveur proxy, si une authentification de base est requise. L'utilisateur doit être un utilisateur local - les utilisateurs de domaine ne sont pas pris en charge.
- *proxy_password* = Mot de passe pour le nom d'utilisateur que vous avez spécifié.
- *ca_cert_dir* = Chemin sur le système Linux de classification des données contenant des ensembles de certificats CA TLS supplémentaires. Requis uniquement si le proxy effectue une interception TLS.

Résultat

Le programme d'installation de la classification des données installe les packages, enregistre l'installation et installe la classification des données. L'installation peut prendre 10 à 20 minutes.

S'il existe une connectivité via le port 8080 entre la machine hôte et l'instance de l'agent de la console, vous verrez la progression de l'installation dans l'onglet Classification des données de la console.

Et ensuite?

Depuis la page Configuration, vous pouvez sélectionner les sources de données que vous souhaitez analyser.

Installer NetApp Data Classification sur un hôte Linux sans accès Internet

L'installation de NetApp Data Classification sur un hôte Linux dans un site local qui n'a pas accès à Internet est appelée *mode privé*. Ce type d'installation, qui utilise un script d'installation, n'a aucune connectivité à la couche SaaS de la NetApp Console .

> Le mode privé BlueXP (interface BlueXP héritée) est généralement utilisé avec des environnements locaux qui n'ont pas de connexion Internet et avec des régions cloud sécurisées, notamment AWS Secret Cloud, AWS Top Secret Cloud et Azure IL6. NetApp continue de prendre en charge ces environnements avec l'interface BlueXP héritée. Pour la documentation du mode privé dans l'ancienne interface BlueXP , voir"Documentation PDF pour le mode privé BlueXP" .

Vérifiez que votre hôte Linux est prêt à installer NetApp Data Classification

Avant d'installer manuellement NetApp Data Classification sur un hôte Linux, exécutez éventuellement un script sur l'hôte pour vérifier que toutes les conditions préalables sont réunies pour l'installation de Data Classification. Vous pouvez exécuter ce script sur un hôte Linux de votre réseau ou sur un hôte Linux dans le cloud. L'hôte peut être connecté à Internet ou résider sur un site qui n'a pas accès à Internet (un *site sombre*).

Le script d'installation de la classification des données comprend un script de test pour garantir que votre environnement répond aux exigences. Vous pouvez exécuter ce script séparément pour vérifier que le système hôte Linux est prêt avant d'exécuter le script d'installation.

Commencer

Vous effectuerez les tâches suivantes.

- Vous pouvez également installer un agent de console si vous n'en avez pas déjà un installé. Vous pouvez exécuter le script de test sans avoir installé d'agent de console, mais le script vérifie la connectivité entre l'agent de console et la machine hôte de classification des données. Il est donc recommandé de disposer d'un agent de console.
- Préparez la machine hôte et vérifiez qu'elle répond à toutes les exigences.
- Activez l'accès Internet sortant à partir de la machine hôte de classification des données.
- Vérifiez que tous les ports requis sont activés sur tous les systèmes.
- Téléchargez et exécutez le script de test prérequis.

Créer un agent de console

Un agent de console est requis avant de pouvoir installer et utiliser la classification des données. Vous pouvez toutefois exécuter le script Prérequis sans agent de console.

Tu peux "installer l'agent de console sur site" sur un serveur Linux de votre réseau ou sur un serveur Linux dans le cloud. Vous pouvez également installer la classification des données sur site si l'agent de la console est installé sur site.

Pour créer un agent Console dans votre environnement de fournisseur de cloud, consultez :

- "création d'un agent de console dans AWS"
- "création d'un agent de console dans Azure"
- "création d'un agent de console dans GCP"

Vous aurez besoin de l'adresse IP ou du nom d'hôte du système de l'agent Console lors de l'exécution du script Prérequis. Vous disposez de ces informations si vous avez installé l'agent Console dans vos locaux. Si l'agent Console est déployé dans le cloud, vous pouvez trouver ces informations depuis la Console : sélectionnez l'icône Aide puis **Support** ; dans la section Agent et Audit, sélectionnez **Accéder à l'agent**.

Vérifier les exigences de l'hôte

Le logiciel de classification des données doit s'exécuter sur un hôte répondant à des exigences spécifiques en matière de système d'exploitation, de mémoire vive et de logiciel.

- La classification des données doit être hébergée sur un serveur dédié. L'hôte ne peut pas être partagé avec d'autres applications ou logiciels tiers tels que les antivirus.
- Choisissez la taille qui correspond à l'ensemble de données que vous prévoyez d'analyser avec la classification des données.

Taille du système	processeur	RAM (la mémoire d'échange doit être désactivée)	Disque
Extra Large	32 processeurs	128 Go de RAM	• 1 Tio SSD sur /, ou 100 Gio disponibles sur /opt • 895 Gio disponibles sur /var/lib/docker • 5 Gio sur /tmp • **Pour Podman, 30 Go sur /var/tmp**
Grand	16 processeurs	64 Go de RAM	• 500 Gio SSD sur /, ou 100 Gio disponibles sur /opt • 400 Gio disponibles sur /var/lib/docker ou pour Podman /var/lib/containers • 5 Gio sur /tmp • **Pour Podman, 30 Go sur /var/tmp**

- Lors du déploiement d'une instance de calcul dans le cloud pour votre installation de classification des données, il est recommandé d'utiliser un système qui répond aux exigences système « Large » ci-dessus :
  - **Type d'instance Amazon Elastic Compute Cloud (Amazon EC2)** : « m6i.4xlarge ». "Voir d'autres types d'instances AWS" .
  - **Taille de la machine virtuelle Azure** : « Standard_D16s_v3 ». "Voir d'autres types d'instances Azure" .
  - **Type de machine GCP** : « n2-standard-16 ». "Voir les types d'instances GCP supplémentaires" .
- **Autorisations de dossier UNIX** : Les autorisations UNIX minimales suivantes sont requises :

Dossier	autorisations minimales
/tmp	`rwxrwxrwt`
/opter	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/système	`rwxr-xr-x`

- **Système opérateur**:
  - Les systèmes d'exploitation suivants nécessitent l'utilisation du moteur de conteneur Docker :
    - Red Hat Enterprise Linux versions 7.8 et 7.9
    - Ubuntu 22.04 (nécessite la version 1.23 ou supérieure de Data Classification)
    - Ubuntu 24.04 (nécessite la version 1.23 ou supérieure de Data Classification)
  - Les systèmes d'exploitation suivants nécessitent l'utilisation du moteur de conteneur Podman et nécessitent la version 1.30 ou supérieure de Data Classification :
    - Red Hat Enterprise Linux versions 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 et 9.6.
  - Les extensions vectorielles avancées (AVX2) doivent être activées sur le système hôte.
- **Red Hat Subscription Management** : L'hôte doit être enregistré auprès de Red Hat Subscription Management. S'il n'est pas enregistré, le système ne peut pas accéder aux référentiels pour mettre à jour les logiciels tiers requis lors de l'installation.
- **Logiciel supplémentaire** : Vous devez installer le logiciel suivant sur l'hôte avant d'installer Data Classification :
  - Selon le système d'exploitation que vous utilisez, vous devez installer l'un des moteurs de conteneurs :
    - Docker Engine version 19.3.1 ou supérieure. "Voir les instructions d'installation" .
    - Podman version 4 ou supérieure. Pour installer Podman, entrez(`sudo yum install podman netavark -y`).
- Version Python 3.6 ou supérieure. "Voir les instructions d'installation" .
  - **Considérations NTP** : NetApp recommande de configurer le système de classification des données pour utiliser un service NTP (Network Time Protocol). L'heure doit être synchronisée entre le système de classification des données et le système d'agent de la console.
- **Considérations relatives au pare-feu** : Si vous envisagez d'utiliser `firewalld` , nous vous recommandons de l'activer avant d'installer Data Classification. Exécutez les commandes suivantes pour configurer `firewalld` afin qu'il soit compatible avec la classification des données :

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

Si vous prévoyez d'utiliser des hôtes de classification de données supplémentaires comme nœuds de scanner (dans un modèle distribué), ajoutez ces règles à votre système principal à ce stade :

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Notez que vous devez redémarrer Docker ou Podman chaque fois que vous activez ou mettez à jour `firewalld` paramètres.

Activer l'accès Internet sortant à partir de la classification des données

La classification des données nécessite un accès Internet sortant. Si votre réseau virtuel ou physique utilise un serveur proxy pour l'accès à Internet, assurez-vous que l'instance de classification des données dispose d'un accès Internet sortant pour contacter les points de terminaison suivants.

Cette section n'est pas requise pour les systèmes hôtes installés sur des sites sans connectivité Internet.

Points de terminaison	But
\ https://api.console.netapp.com	Communication avec le service Console, qui inclut les comptes NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Communication avec le site Web de la console pour l'authentification centralisée des utilisateurs.
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ \ https://production.cloudflare.docker.com/	Fournit un accès aux images logicielles, aux manifestes, aux modèles et permet d'envoyer des journaux et des métriques.
\ https://support.compliance.api.console.netapp.com/	Permet à NetApp de diffuser des données à partir des enregistrements d'audit.
\ https://github.com/docker \ https://download.docker.com	Fournit des packages prérequis pour l'installation de Docker.

Points de terminaison	But
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Fournit des packages prérequis pour l'installation d'Ubuntu.

Vérifiez que tous les ports requis sont activés

Vous devez vous assurer que tous les ports requis sont ouverts pour la communication entre l'agent de console, la classification des données, Active Directory et vos sources de données.

Type de connexion	Ports	Description
Agent de console <> Classification des données	8080 (TCP), 443 (TCP) et 80. 9000	Les règles de pare-feu ou de routage de l'agent de console doivent autoriser le trafic entrant et sortant sur le port 443 vers et depuis l'instance de classification des données. Assurez-vous que le port 8080 est ouvert afin de pouvoir voir la progression de l'installation dans la console. Si un pare-feu est utilisé sur l'hôte Linux, le port 9000 est requis pour les processus internes au sein d'un serveur Ubuntu.
Agent de console <> cluster ONTAP (NAS)	443 (TCP)	La console découvre les clusters ONTAP à l'aide de HTTPS. Si vous utilisez des stratégies de pare-feu personnalisées, l'hôte de l'agent de console doit autoriser l'accès HTTPS sortant via le port 443. Si l'agent de la console est dans le cloud, toutes les communications sortantes sont autorisées par le pare-feu prédéfini ou les règles de routage.

Exécutez le script des prérequis de classification des données

Suivez ces étapes pour exécuter le script des prérequis de classification des données.

"Regardez cette vidéo"pour voir comment exécuter le script Prérequis et interpréter les résultats.

Avant de commencer

- Vérifiez que votre système Linux répond auxexigences de l'hôte .
- Vérifiez que le système dispose des deux packages logiciels prérequis installés (Docker Engine ou Podman et Python 3).
- Assurez-vous que vous disposez des privilèges root sur le système Linux.

Étapes

1. Téléchargez le script des prérequis de classification des données à partir du "Site de support NetApp" . Le fichier que vous devez sélectionner est nommé **standalone-pre-requisite-tester-<version>**.
2. Copiez le fichier sur l'hôte Linux que vous prévoyez d'utiliser (en utilisant `scp` ou une autre méthode).
3. Attribuer des autorisations pour exécuter le script.

```
chmod +x standalone-pre-requisite-tester-v1.25.0
```

4. Exécutez le script à l'aide de la commande suivante.

```
./standalone-pre-requisite-tester-v1.25.0 <--darksite>
```

   Ajoutez l'option « --darksite » uniquement si vous exécutez le script sur un hôte qui n'a pas accès à Internet. Certains tests prérequis sont ignorés lorsque l'hôte n'est pas connecté à Internet.

5. Le script vous demande l'adresse IP de la machine hôte de classification des données.
   - Entrez l'adresse IP ou le nom d'hôte.
6. Le script vous demande si vous disposez d'un agent de console installé.
   - Entrez **N** si vous n'avez pas d'agent de console installé.
   - Entrez **Y** si vous avez un agent de console installé. Ensuite, entrez l'adresse IP ou le nom d'hôte de l'agent de la console afin que le script de test puisse tester cette connectivité.
7. Le script exécute une variété de tests sur le système et affiche les résultats au fur et à mesure de sa progression. Une fois terminé, il écrit un journal de la session dans un fichier nommé `prerequisites-test-<timestamp>.log` dans le répertoire `/opt/netapp/install_logs`.

Résultat

Si tous les tests prérequis se sont déroulés avec succès, vous pouvez installer Data Classification sur l'hôte lorsque vous êtes prêt.

Si des problèmes sont détectés, ils sont classés comme « Recommandé » ou « Obligatoire » pour être résolus. Les problèmes recommandés sont généralement des éléments qui ralentiraient l'exécution des tâches d'analyse et de catégorisation de la classification des données. Ces éléments n'ont pas besoin d'être corrigés, mais vous souhaiterez peut-être les corriger.

Si vous rencontrez des problèmes « obligatoires », vous devez les résoudre et exécuter à nouveau le script de test des prérequis.

Activer l'analyse sur vos sources de données

Analyser les sources de données avec la NetApp Data Classification

NetApp Data Classification analyse les données dans les référentiels (volumes, schémas de base de données ou autres données utilisateur) que vous sélectionnez pour identifier les données personnelles et sensibles. La classification des données cartographie ensuite vos données organisationnelles, catégorise chaque fichier et identifie des modèles prédéfinis dans les données. Le résultat de l'analyse est un index des informations personnelles, des informations personnelles sensibles, des catégories de données et des types de fichiers.

Après l'analyse initiale, Data Classification analyse en continu vos données de manière circulaire pour détecter les modifications incrémentielles. C'est pourquoi il est important de maintenir l'instance en cours d'exécution.

Vous pouvez activer et désactiver les analyses au niveau du volume ou au niveau du schéma de base de données.

Quelle est la différence entre les analyses de cartographie et de classification

Vous pouvez effectuer deux types d'analyses dans la classification des données :

- **Les analyses de cartographie uniquement** fournissent uniquement un aperçu de haut niveau de vos données et sont effectuées sur des sources de données sélectionnées. Les analyses de cartographie uniquement prennent moins de temps que les analyses de cartographie et de classification, car elles n'accèdent pas aux fichiers pour voir les données qu'ils contiennent. Vous souhaiterez peut-être procéder ainsi dans un premier temps pour identifier les domaines de recherche, puis effectuer une analyse de cartographie et de classification sur ces domaines.
- **Les analyses de cartographie et de classification** fournissent une analyse approfondie de vos données.

Le tableau ci-dessous montre certaines des différences :

Fonctionnalité	Cartographier et classer les scans	Analyses de cartographie uniquement
Vitesse de numérisation	Lent	Rapide
Tarifs	Gratuit	Gratuit
Capacité	Limité à 500 Tio*	Limité à 500 Tio*
Liste des types de fichiers et de la capacité utilisée	Oui	Oui
Nombre de fichiers et capacité utilisée	Oui	Oui
Âge et taille des fichiers	Oui	Oui
Capacité à exécuter un"Rapport de mappage des données"	Oui	Oui
Page d'enquête sur les données pour afficher les détails du fichier	Oui	Non
Rechercher des noms dans les fichiers	Oui	Non
Créer"requêtes enregistrées" qui fournissent des résultats de recherche personnalisés	Oui	Non
Possibilité d'exécuter d'autres rapports	Oui	Non
Possibilité de voir les métadonnées des fichiers**	Non	Oui

* La classification des données n'impose pas de limite à la quantité de données qu'elle peut analyser. Chaque agent de console prend en charge l'analyse et l'affichage de 500 Tio de données. Pour scanner plus de 500 Tio de données,"installer un autre agent de console" alors"déployer une autre instance de classification des données" . + L'interface utilisateur de la console affiche les données d'un seul connecteur. Pour obtenir des conseils sur l'affichage des données de plusieurs agents de console, consultez"Travailler avec plusieurs agents de console" .

** Les métadonnées suivantes sont extraites des fichiers lors des analyses de mappage :

- Système
- Type de système
- Référentiel de stockage
- Type de fichier

- Capacité utilisée
- Nombre de fichiers
- Taille du fichier
- Création de fichier
- Dernier accès au fichier
- Fichier modifié pour la dernière fois
- Heure de découverte du fichier
- Extraction des autorisations

Différences entre les tableaux de bord de gouvernance :

Fonctionnalité	Cartographier et classer	Carte
Données obsolètes	Oui	Oui
Données non commerciales	Oui	Oui
Fichiers dupliqués	Oui	Oui
Requêtes enregistrées prédéfinies	Oui	Non
Requêtes enregistrées par défaut	Oui	Oui
Rapport DDA	Oui	Oui
Rapport de cartographie	Oui	Oui
Détection du niveau de sensibilité	Oui	Non
Données sensibles avec des autorisations étendues	Oui	Non
Autorisations ouvertes	Oui	Oui
L'âge des données	Oui	Oui
Taille des données	Oui	Oui
Catégories	Oui	Non
Types de fichiers	Oui	Oui

Différences entre les tableaux de bord de conformité :

Fonctionnalité	Cartographier et classer	Carte
Informations personnelles	Oui	Non
Informations personnelles sensibles	Oui	Non
Rapport d'évaluation des risques liés à la vie privée	Oui	Non
Rapport HIPAA	Oui	Non
Rapport PCI DSS	Oui	Non

Différences entre les filtres d'investigation :

Fonctionnalité	Cartographier et classer	Carte
Requêtes enregistrées	Oui	Oui
Type de système	Oui	Oui
Système	Oui	Oui
Référentiel de stockage	Oui	Oui
Type de fichier	Oui	Oui
Taille du fichier	Oui	Oui
Temps de création	Oui	Oui
Temps découvert	Oui	Oui
Dernière modification	Oui	Oui
Dernier accès	Oui	Oui
Autorisations ouvertes	Oui	Oui
Chemin du répertoire de fichiers	Oui	Oui
Catégorie	Oui	Non
Niveau de sensibilité	Oui	Non
Nombre d'identifiants	Oui	Non
Données personnelles	Oui	Non
Données personnelles sensibles	Oui	Non
Personne concernée	Oui	Non
Doublons	Oui	Oui
Statut de classification	Oui	Le statut est toujours « Informations limitées »
Événement d'analyse d'analyse	Oui	Oui
Hachage de fichier	Oui	Oui
Nombre d'utilisateurs avec accès	Oui	Oui
Autorisations utilisateur/groupe	Oui	Oui
Propriétaire du fichier	Oui	Oui
Type de répertoire	Oui	Oui

Analyser Amazon FSx pour les volumes ONTAP avec la NetApp Data Classification

Suivez quelques étapes pour analyser Amazon FSx pour les volumes ONTAP avec NetApp Data Classification.

Avant de commencer

- Vous avez besoin d'un agent de console actif dans AWS pour déployer et gérer la classification des données.
- Le groupe de sécurité que vous avez sélectionné lors de la création du système doit autoriser le trafic provenant de l'instance de classification des données. Vous pouvez trouver le groupe de sécurité associé à l'aide de l'ENI connecté au système de fichiers FSx for ONTAP et le modifier à l'aide de la console de gestion AWS.

  "Groupes de sécurité AWS pour les instances Linux"

  "Groupes de sécurité AWS pour les instances Windows"

  "Interfaces réseau élastiques AWS (ENI)"

- Assurez-vous que les ports suivants sont ouverts sur l'instance de classification des données :
  - Pour NFS – ports 111 et 2049.
  - Pour CIFS – ports 139 et 445.

Déployer l'instance de classification des données

"Déployer la classification des données"s'il n'y a pas déjà une instance déployée.

Vous devez déployer la classification des données sur le même réseau AWS que l'agent de console pour AWS et les volumes FSx que vous souhaitez analyser.

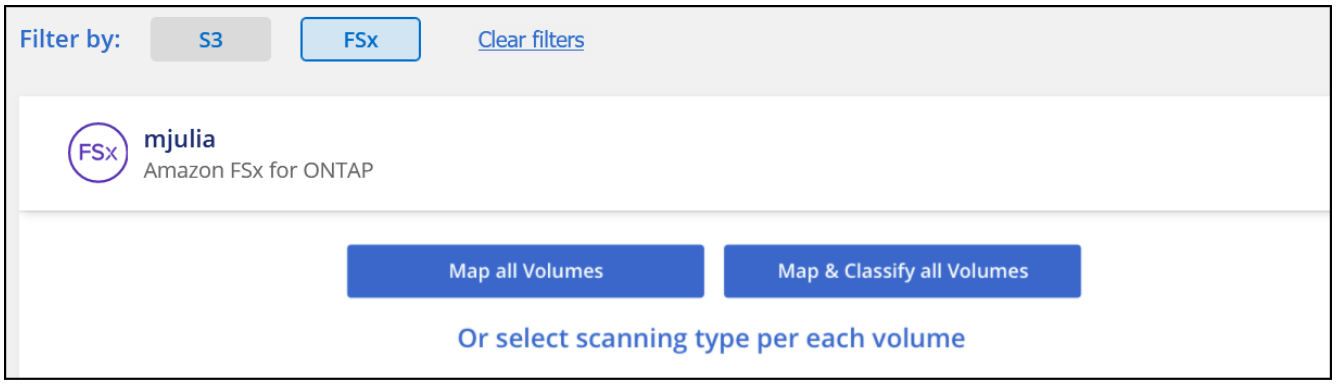
Remarque : le déploiement de la classification des données dans un emplacement local n'est actuellement pas pris en charge lors de l'analyse des volumes FSx.

Les mises à niveau du logiciel de classification des données sont automatisées tant que l'instance dispose d'une connexion Internet.

Activez la classification des données dans vos systèmes

Vous pouvez activer la classification des données pour les volumes FSx for ONTAP .

1. Depuis la NetApp Console, **Gouvernance > Classification**.
2. Dans le menu Classification des données, sélectionnez **Configuration**.



3. Sélectionnez la manière dont vous souhaitez analyser les volumes de chaque système. "En savoir plus sur les analyses de cartographie et de classification":

- Pour mapper tous les volumes, sélectionnez **Mapper tous les volumes**.
- Pour cartographier et classer tous les volumes, sélectionnez **Cartographier et classer tous les volumes**.
- Pour personnaliser l'analyse de chaque volume, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**, puis choisissez les volumes que vous souhaitez mapper et/ou classer.

4. Dans la boîte de dialogue de confirmation, sélectionnez **Approuver** pour que la classification des données commence à analyser vos volumes.

Résultat

La classification des données commence à analyser les volumes que vous avez sélectionnés dans le système. Les résultats seront disponibles dans le tableau de bord de conformité dès que la classification des données aura terminé les analyses initiales. Le temps nécessaire dépend de la quantité de données : cela peut prendre quelques minutes ou quelques heures. Vous pouvez suivre la progression de l'analyse initiale en accédant au menu **Configuration** puis en sélectionnant la **Configuration système**. Suivez la progression de chaque analyse dans la barre de progression ; vous pouvez survoler la barre de progression pour voir le nombre de fichiers analysés par rapport au nombre total de fichiers dans le volume.

- Par défaut, si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers de vos volumes car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**. La page résultante contient un paramètre que vous pouvez activer pour que la classification des données analyse les volumes quelles que soient les autorisations.
- La classification des données analyse un seul partage de fichiers sous un volume. Si vous avez plusieurs partages dans vos volumes, vous devrez analyser ces autres partages séparément en tant que groupe de partages. "Voir plus de détails sur cette limitation de classification des données" .

Vérifiez que la classification des données a accès aux volumes

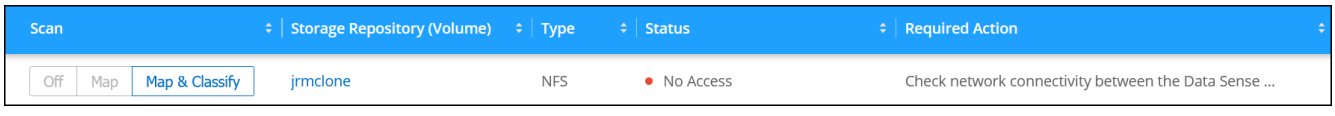
Assurez-vous que la classification des données peut accéder aux volumes en vérifiant votre réseau, vos groupes de sécurité et vos politiques d'exportation.

Vous devrez fournir à Data Classification les informations d'identification CIFS afin qu'il puisse accéder aux volumes CIFS.

Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Sur la page Configuration, sélectionnez **Afficher les détails** pour vérifier l'état et corriger les erreurs.

   Par exemple, l'image suivante montre un volume que Data Classification ne peut pas analyser en raison de problèmes de connectivité réseau entre l'instance Data Classification et le volume.



3. Assurez-vous qu'il existe une connexion réseau entre l'instance de classification des données et chaque réseau qui inclut des volumes pour FSx pour ONTAP.

Pour FSx for ONTAP, la classification des données peut analyser les volumes uniquement dans la même région que la console.

4. Assurez-vous que les stratégies d'exportation de volume NFS incluent l'adresse IP de l'instance de classification des données afin qu'elle puisse accéder aux données sur chaque volume.
5. Si vous utilisez CIFS, fournissez à Data Classification les informations d'identification Active Directory afin qu'il puisse analyser les volumes CIFS.

   a. Dans le menu Classification des données, sélectionnez **Configuration**.
   b. Pour chaque système, sélectionnez **Modifier les informations d'identification CIFS** et saisissez le nom d'utilisateur et le mot de passe dont Data Classification a besoin pour accéder aux volumes CIFS sur le système.

      Les informations d'identification peuvent être en lecture seule, mais la fourniture d'informations d'identification d'administrateur garantit que la classification des données peut lire toutes les données nécessitant des autorisations élevées. Les informations d'identification sont stockées sur l'instance de classification des données.

      Si vous souhaitez vous assurer que les « dernières heures d'accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, il est recommandé que l'utilisateur dispose des autorisations d'écriture d'attributs dans CIFS ou des autorisations d'écriture dans NFS. Si possible, configurez l'utilisateur Active Directory en tant que membre d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

   Après avoir saisi les informations d'identification, vous devriez voir un message indiquant que tous les volumes CIFS ont été authentifiés avec succès.

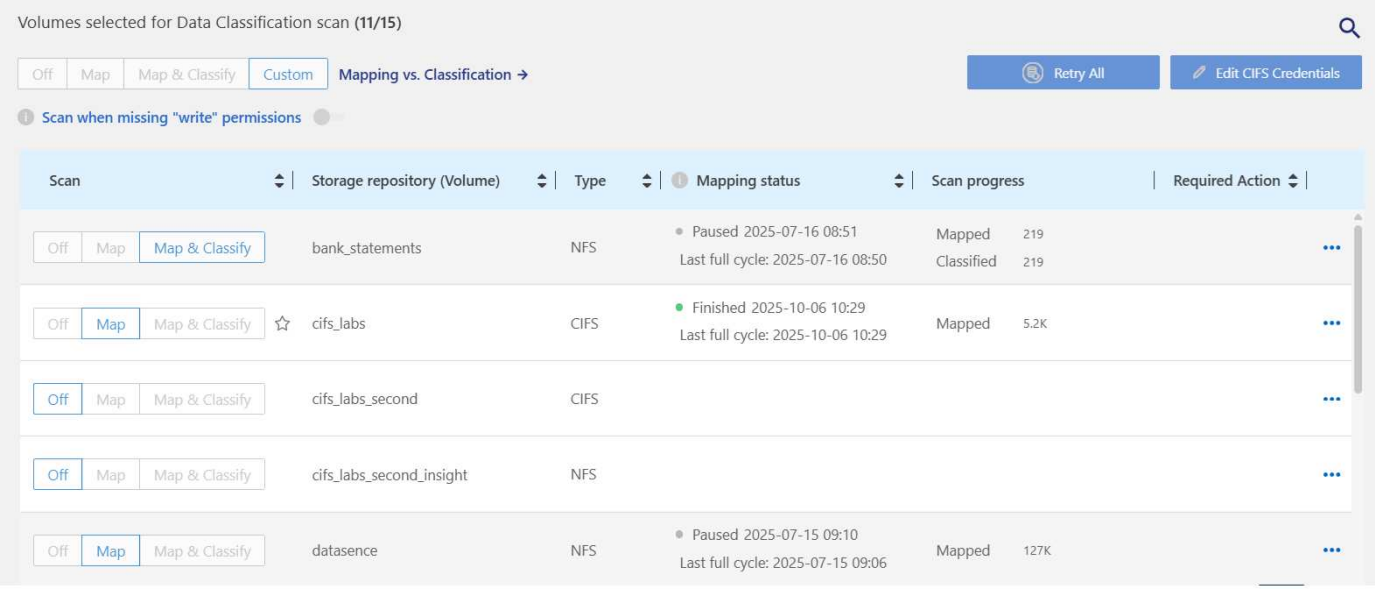
Activer et désactiver les analyses sur les volumes

Vous pouvez démarrer ou arrêter les analyses sur n'importe quel système à tout moment à partir de la page de configuration. Vous pouvez également passer d'analyses de cartographie uniquement à des analyses de cartographie et de classification, et vice-versa. Il est recommandé d'analyser tous les volumes d'un système.

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez sélectionné le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque cette option est définie sur **Personnalisé** ou **Désactivé** dans la zone d'en-tête, vous devrez activer le mappage et/ou l'analyse complète sur chaque nouveau volume que vous ajoutez au système.

Le commutateur en haut de la page pour **Analyser en cas d'absence d'autorisations « d'écriture »** est désactivé par défaut. Cela signifie que si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, activez l'interrupteur et tous les fichiers sont analysés quelles que soient les autorisations. "Apprendre encore plus".

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez défini le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque le paramètre pour tous les volumes est **Personnalisé** ou **Désactivé**, vous devez activer l'analyse manuellement pour chaque nouveau volume que vous ajoutez.



Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Choisissez un système, puis sélectionnez **Configuration**.
3. Pour activer ou désactiver les analyses pour tous les volumes, sélectionnez **Map**, **Map & Classify** ou **Off** dans l'en-tête au-dessus de tous les volumes.

   Pour activer ou désactiver les analyses de volumes individuels, recherchez les volumes dans la liste, puis sélectionnez **Map**, **Map & Classify** ou **Off** à côté du nom du volume.

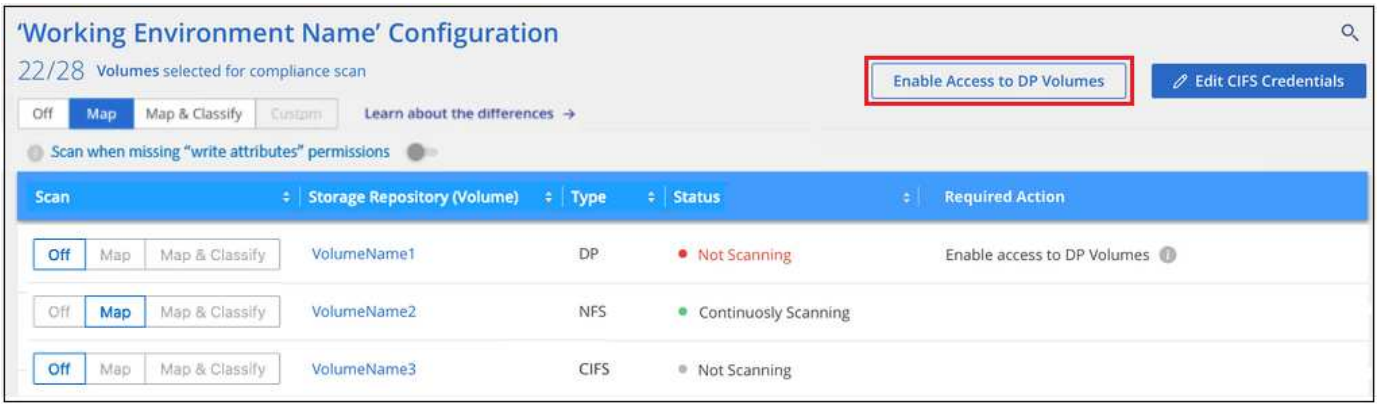
Résultat

Lorsque vous activez l'analyse, la classification des données démarre l'analyse des volumes que vous avez sélectionnés dans le système. Les résultats commencent à apparaître dans le tableau de bord Conformité dès que la classification des données démarre l'analyse. Le temps d'exécution de l'analyse dépend de la quantité de données, allant de quelques minutes à quelques heures.

Analyser les volumes de protection des données

Par défaut, les volumes de protection des données (DP) ne sont pas analysés car ils ne sont pas exposés en externe et Data Classification ne peut pas y accéder. Il s'agit des volumes de destination pour les opérations SnapMirror à partir d'un système de fichiers FSx pour ONTAP .

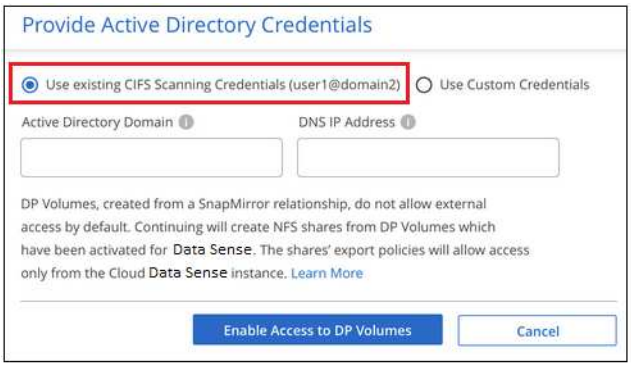
Initialement, la liste des volumes identifie ces volumes comme *Type* **DP** avec le *Statut* **Pas d'analyse** et l'*Action requise* **Activer l'accès aux volumes DP**.

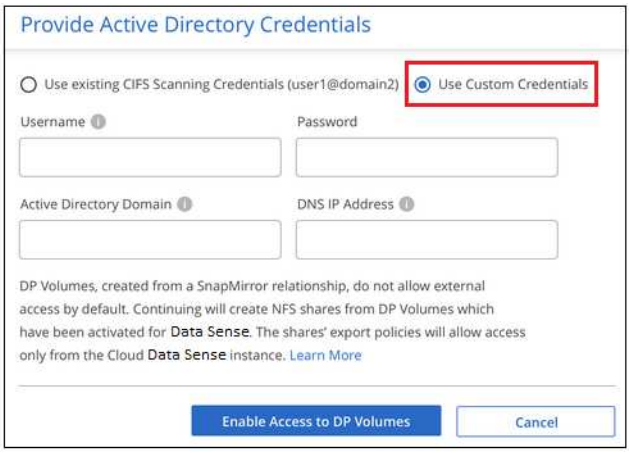


Étapes

Si vous souhaitez analyser ces volumes de protection des données :

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Sélectionnez **Activer l'accès aux volumes DP** en haut de la page.
3. Vérifiez le message de confirmation et sélectionnez à nouveau **Activer l'accès aux volumes DP**.
   - Les volumes initialement créés en tant que volumes NFS dans le système de fichiers source FSx pour ONTAP sont activés.
   - Les volumes initialement créés en tant que volumes CIFS dans le système de fichiers source FSx for ONTAP nécessitent que vous saisissiez les informations d'identification CIFS pour analyser ces volumes DP. Si vous avez déjà saisi les informations d'identification Active Directory pour que la classification des données puisse analyser les volumes CIFS, vous pouvez utiliser ces informations d'identification ou spécifier un autre ensemble d'informations d'identification d'administrateur.





4. Activez chaque volume DP que vous souhaitez analyser.

Résultat

Une fois activée, la classification des données crée un partage NFS à partir de chaque volume DP activé pour l'analyse. Les politiques d'exportation de partage autorisent uniquement l'accès à partir de l'instance de classification des données.

Si vous n'aviez aucun volume de protection des données CIFS lorsque vous avez initialement activé l'accès aux volumes DP, et que vous en avez ajouté ultérieurement, le bouton **Activer l'accès à CIFS DP** apparaît en haut de la page de configuration. Sélectionnez ce bouton et ajoutez les informations d'identification CIFS pour activer l'accès à ces volumes CIFS DP.

Les informations d'identification Active Directory sont enregistrées uniquement dans la machine virtuelle de stockage du premier volume DP CIFS. Par conséquent, tous les volumes DP sur cette SVM seront analysés. Tous les volumes résidant sur d'autres SVM n'auront pas les informations d'identification Active Directory enregistrées, de sorte que ces volumes DP ne seront pas analysés.

Analyser les volumes Azure NetApp Files avec la NetApp Data Classification

Suivez quelques étapes pour démarrer avec NetApp Data Classification pour Azure NetApp Files.

Découvrez le système Azure NetApp Files que vous souhaitez analyser

Si le système Azure NetApp Files que vous souhaitez analyser n'est pas déjà présent dans la NetApp Console en tant que système,"ajoutez-le dans la page Systèmes" .

Déployer l'instance de classification des données

"Déployer la classification des données"s'il n'y a pas déjà une instance déployée.

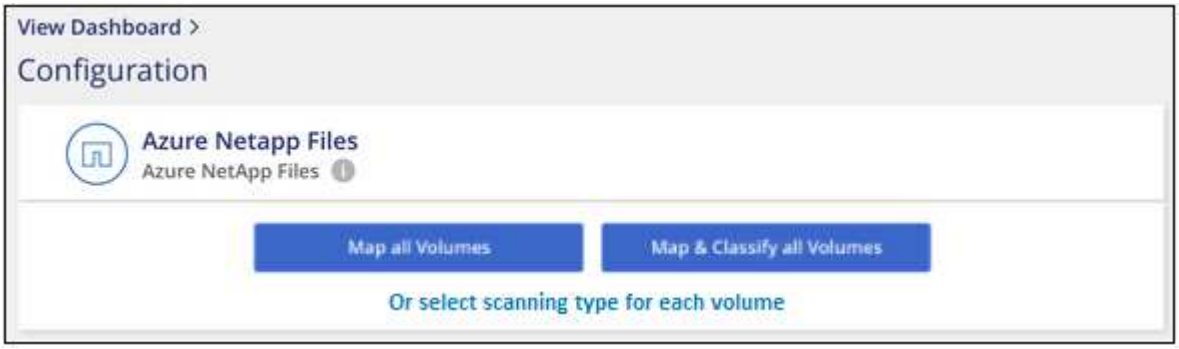
La classification des données doit être déployée dans le cloud lors de l'analyse des volumes Azure NetApp Files et doit être déployée dans la même région que les volumes que vous souhaitez analyser.

Remarque : le déploiement de la classification des données dans un emplacement local n'est actuellement pas pris en charge lors de l'analyse des volumes Azure NetApp Files .

Activez la classification des données dans vos systèmes

Vous pouvez activer la classification des données sur vos volumes Azure NetApp Files .

1. Dans le menu Classification des données, sélectionnez **Configuration**.



2. Sélectionnez la manière dont vous souhaitez analyser les volumes de chaque système. "En savoir plus sur les analyses de cartographie et de classification":
   - Pour mapper tous les volumes, sélectionnez **Mapper tous les volumes**.
   - Pour cartographier et classer tous les volumes, sélectionnez **Cartographier et classer tous les volumes**.
   - Pour personnaliser l'analyse de chaque volume, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**, puis choisissez les volumes que vous souhaitez mapper ou mapper et classer.

     VoirActiver ou désactiver les analyses sur les volumes pour plus de détails.

3. Dans la boîte de dialogue de confirmation, sélectionnez **Approuver**.

Résultat

La classification des données commence à analyser les volumes que vous avez sélectionnés dans le système. Les résultats sont disponibles dans le tableau de bord Conformité dès que la classification des données termine les analyses initiales. Le temps nécessaire dépend de la quantité de données : cela peut prendre quelques minutes ou quelques heures. Vous pouvez suivre la progression de l'analyse initiale en accédant au menu **Configuration** puis en sélectionnant la **Configuration système**. La classification des données affiche une barre de progression pour chaque analyse. Vous pouvez survoler la barre de progression pour voir le nombre de fichiers analysés par rapport au nombre total de fichiers dans le volume.

- Par défaut, si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers de vos volumes car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**. La page résultante contient un paramètre que vous pouvez activer pour que la classification des données analyse les volumes quelles que soient les autorisations.
- La classification des données analyse un seul partage de fichiers sous un volume. Si vous avez plusieurs partages dans vos volumes, vous devrez analyser ces autres partages séparément en tant que groupe de partages. "En savoir plus sur cette limitation de classification des données" .

Vérifiez que la classification des données a accès aux volumes

Assurez-vous que la classification des données peut accéder aux volumes en vérifiant votre réseau, vos groupes de sécurité et vos politiques d'exportation. Vous devez fournir à Data Classification les informations d'identification CIFS afin qu'elle puisse accéder aux volumes CIFS.

Pour Azure NetApp Files, la classification des données ne peut analyser que les volumes dans la même région que la console.

Liste de contrôle

- Assurez-vous qu'il existe une connexion réseau entre l'instance de classification des données et chaque réseau qui inclut des volumes pour Azure NetApp Files.
- Assurez-vous que les ports suivants sont ouverts sur l'instance de classification des données :
  - Pour NFS – ports 111 et 2049.
  - Pour CIFS – ports 139 et 445.
- Assurez-vous que les stratégies d'exportation de volume NFS incluent l'adresse IP de l'instance de classification des données afin qu'elle puisse accéder aux données sur chaque volume.

Étapes

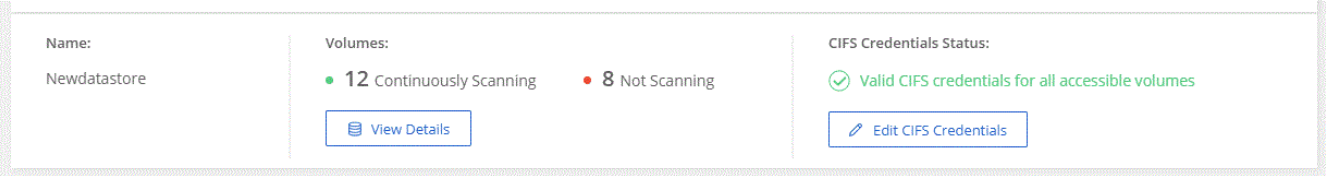
1. Dans le menu Classification des données, sélectionnez **Configuration**.

   a. Si vous utilisez CIFS (SMB), assurez-vous que les informations d'identification Active Directory sont correctes. Pour chaque système, sélectionnez **Modifier les informations d'identification CIFS**, puis saisissez le nom d'utilisateur et le mot de passe dont Data Classification a besoin pour accéder aux volumes CIFS sur le système.

      Les informations d'identification peuvent être en lecture seule ; la fourniture d'informations d'identification d'administrateur garantit que Data Classification peut lire toutes les données nécessitant des autorisations élevées. Les informations d'identification sont stockées sur l'instance de classification des données.

Si vous souhaitez vous assurer que les « dernières heures d'accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, il est recommandé que l'utilisateur dispose des autorisations d'écriture d'attributs dans CIFS ou des autorisations d'écriture dans NFS. Si possible, configurez l'utilisateur Active Directory en tant que membre d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

Après avoir saisi les informations d'identification, vous devriez voir un message indiquant que tous les volumes CIFS ont été authentifiés avec succès.



2. Sur la page Configuration, sélectionnez **Afficher les détails** pour vérifier l'état de chaque volume CIFS et NFS. Si nécessaire, corrigez les erreurs telles que les problèmes de connectivité réseau.

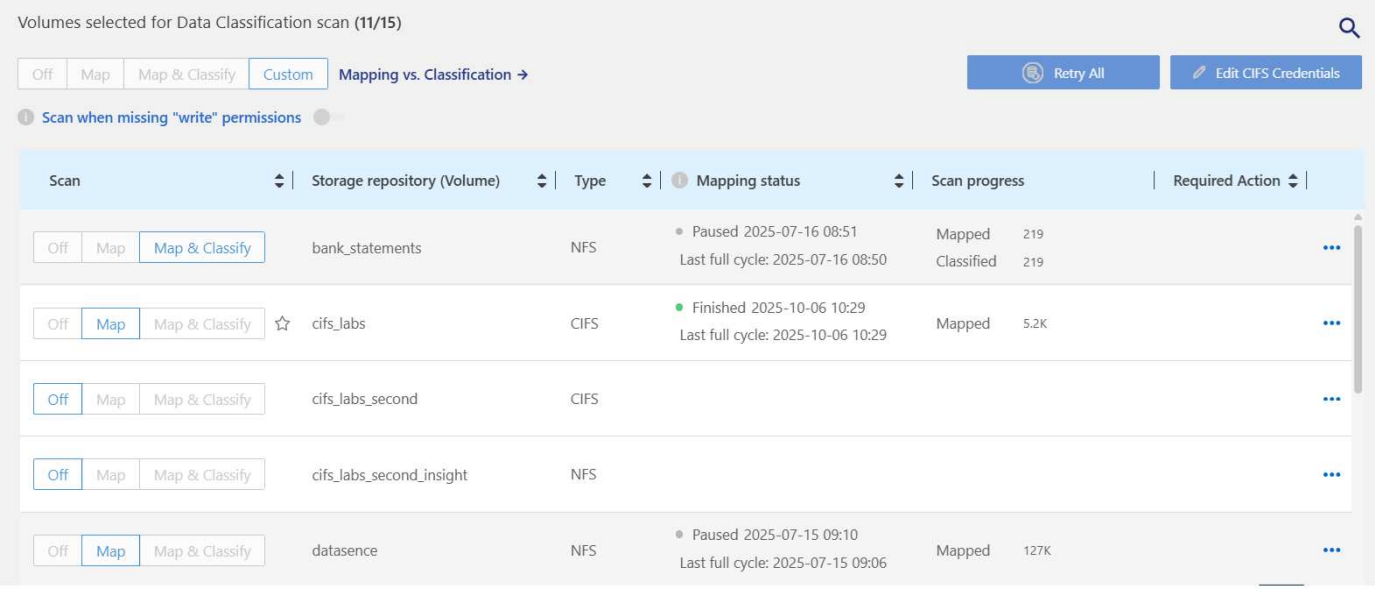
Activer ou désactiver les analyses sur les volumes

Vous pouvez démarrer ou arrêter les analyses sur n'importe quel système à tout moment à partir de la page de configuration. Vous pouvez également passer d'analyses de cartographie uniquement à des analyses de cartographie et de classification, et vice-versa. Il est recommandé d'analyser tous les volumes d'un système.

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez sélectionné le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque cette option est définie sur **Personnalisé** ou **Désactivé** dans la zone d'en-tête, vous devrez activer le mappage et/ou l'analyse complète sur chaque nouveau volume que vous ajoutez au système.

Le commutateur en haut de la page pour **Analyser en cas d'absence d'autorisations « d'écriture »** est désactivé par défaut. Cela signifie que si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, activez l'interrupteur et tous les fichiers sont analysés quelles que soient les autorisations. "Apprendre encore plus".

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez défini le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque le paramètre pour tous les volumes est **Personnalisé** ou **Désactivé**, vous devez activer l'analyse manuellement pour chaque nouveau volume que vous ajoutez.



Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Choisissez un système, puis sélectionnez **Configuration**.
3. Pour activer ou désactiver les analyses pour tous les volumes, sélectionnez **Map**, **Map & Classify** ou **Off** dans l'en-tête au-dessus de tous les volumes.

   Pour activer ou désactiver les analyses de volumes individuels, recherchez les volumes dans la liste, puis sélectionnez **Map**, **Map & Classify** ou **Off** à côté du nom du volume.

Résultat

Lorsque vous activez l'analyse, la classification des données démarre l'analyse des volumes que vous avez sélectionnés dans le système. Les résultats commencent à apparaître dans le tableau de bord Conformité dès que la classification des données démarre l'analyse. Le temps d'exécution de l'analyse dépend de la quantité de données, allant de quelques minutes à quelques heures.

Analysez les Cloud Volumes ONTAP et les volumes ONTAP sur site avec la NetApp Data Classification

Suivez quelques étapes pour commencer à analyser vos Cloud Volumes ONTAP et vos volumes ONTAP sur site à l'aide de NetApp Data Classification.

Prérequis

Avant d'activer la classification des données, assurez-vous que vous disposez d'une configuration prise en charge.

- Si vous numérisez des Cloud Volumes ONTAP et des systèmes ONTAP sur site accessibles via Internet, vous pouvez"déployer la classification des données dans le cloud" ou"dans un local équipé d'un accès Internet" .
- Si vous analysez des systèmes ONTAP sur site qui ont été installés sur un site sombre sans accès Internet, vous devez"déployer la classification des données dans le même emplacement sur site qui n'a pas d'accès Internet" . Cela nécessite que l'agent de console soit déployé dans le même emplacement sur site.

Vérifiez que la classification des données a accès aux volumes

Assurez-vous que la classification des données peut accéder aux volumes en vérifiant votre réseau, vos groupes de sécurité et vos politiques d'exportation. Vous devrez fournir à Data Classification les informations d'identification CIFS afin qu'il puisse accéder aux volumes CIFS.

Liste de contrôle

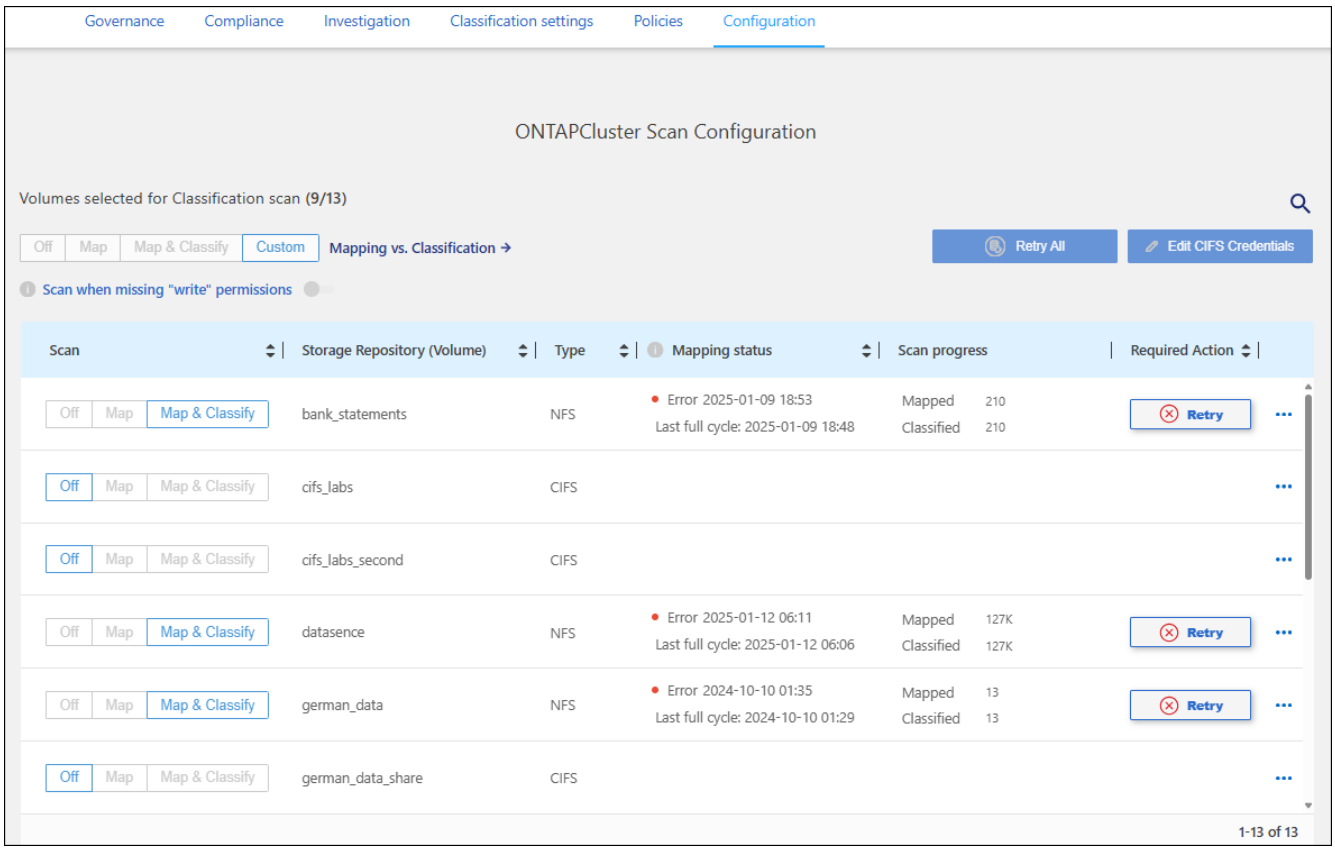
- Assurez-vous qu'il existe une connexion réseau entre l'instance de classification des données et chaque réseau qui inclut des volumes pour les clusters Cloud Volumes ONTAP ou ONTAP sur site.
- Assurez-vous que le groupe de sécurité pour Cloud Volumes ONTAP autorise le trafic entrant depuis l'instance de classification des données.

  Vous pouvez soit ouvrir le groupe de sécurité pour le trafic provenant de l'adresse IP de l'instance de classification des données, soit ouvrir le groupe de sécurité pour tout le trafic provenant de l'intérieur du réseau virtuel.

- Assurez-vous que les stratégies d'exportation de volume NFS incluent l'adresse IP de l'instance de classification des données afin qu'elle puisse accéder aux données sur chaque volume.

Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.



2. Si vous utilisez CIFS, fournissez à Data Classification les informations d'identification Active Directory afin qu'il puisse analyser les volumes CIFS. Pour chaque système, sélectionnez **Modifier les informations d'identification CIFS** et saisissez le nom d'utilisateur et le mot de passe dont Data Classification a besoin pour accéder aux volumes CIFS sur le système.

   Les informations d'identification peuvent être en lecture seule, mais la fourniture d'informations d'identification d'administrateur garantit que la classification des données peut lire toutes les données

nécessitant des autorisations élevées. Les informations d'identification sont stockées sur l'instance de classification des données.

Si vous souhaitez vous assurer que les « dernières heures d'accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, il est recommandé que l'utilisateur dispose des autorisations d'écriture d'attributs dans CIFS ou des autorisations d'écriture dans NFS. Si possible, configurez l'utilisateur Active Directory en tant que membre d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

Si vous avez correctement saisi les informations d'identification, un message confirme que tous les volumes CIFS ont été authentifiés avec succès.

3. Sur la page Configuration, sélectionnez **Configuration** pour vérifier l'état de chaque volume CIFS et NFS et corriger les erreurs éventuelles.

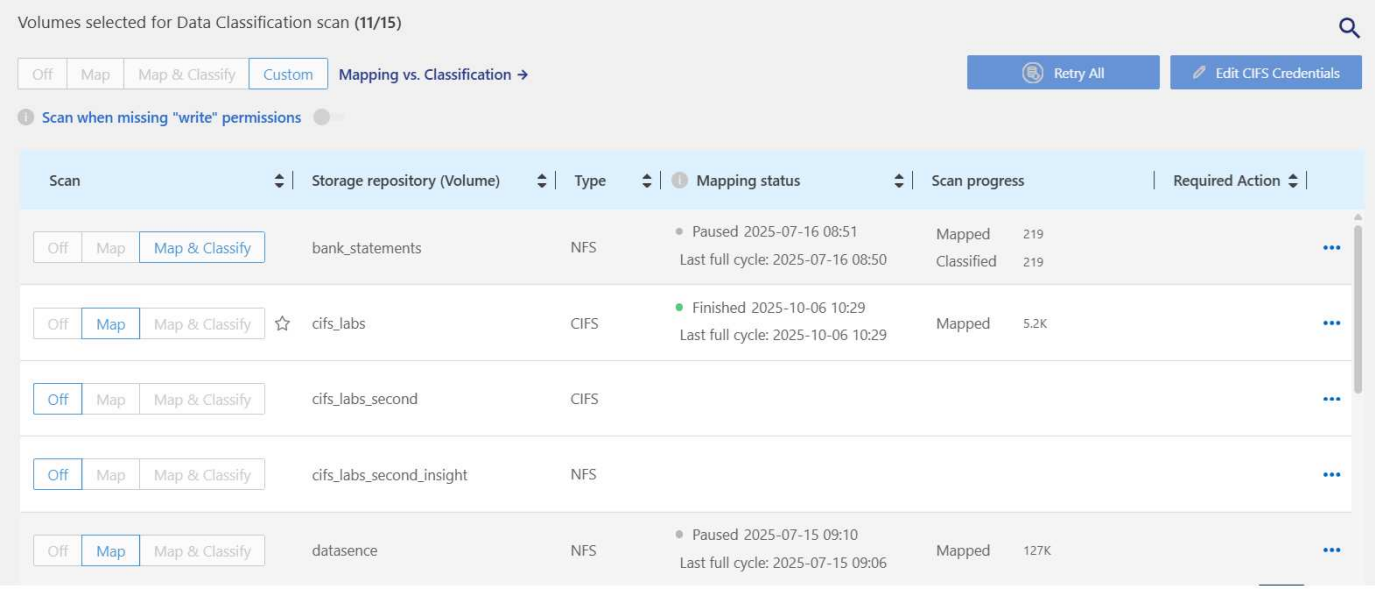
Activer ou désactiver les analyses sur les volumes

Vous pouvez démarrer ou arrêter les analyses sur n'importe quel système à tout moment à partir de la page de configuration. Vous pouvez également passer d'analyses de cartographie uniquement à des analyses de cartographie et de classification, et vice-versa. Il est recommandé d'analyser tous les volumes d'un système.

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez sélectionné le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque cette option est définie sur **Personnalisé** ou **Désactivé** dans la zone d'en-tête, vous devrez activer le mappage et/ou l'analyse complète sur chaque nouveau volume que vous ajoutez au système.

Le commutateur en haut de la page pour **Analyser en cas d'absence d'autorisations « d'écriture »** est désactivé par défaut. Cela signifie que si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, activez l'interrupteur et tous les fichiers sont analysés quelles que soient les autorisations. "Apprendre encore plus".

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez défini le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque le paramètre pour tous les volumes est **Personnalisé** ou **Désactivé**, vous devez activer l'analyse manuellement pour chaque nouveau volume que vous ajoutez.



Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Choisissez un système, puis sélectionnez **Configuration**.
3. Pour activer ou désactiver les analyses pour tous les volumes, sélectionnez **Map**, **Map & Classify** ou **Off** dans l'en-tête au-dessus de tous les volumes.

   Pour activer ou désactiver les analyses de volumes individuels, recherchez les volumes dans la liste, puis sélectionnez **Map**, **Map & Classify** ou **Off** à côté du nom du volume.

Résultat

Lorsque vous activez l'analyse, la classification des données démarre l'analyse des volumes que vous avez sélectionnés dans le système. Les résultats commencent à apparaître dans le tableau de bord Conformité dès que la classification des données démarre l'analyse. Le temps d'exécution de l'analyse dépend de la quantité de données, allant de quelques minutes à quelques heures.

La classification des données analyse un seul partage de fichiers sous un volume. Si vous avez plusieurs partages dans vos volumes, vous devrez analyser ces autres partages séparément en tant que groupe de partages. "Voir plus de détails sur cette limitation de classification des données" .

Analyser les schémas de base de données avec la NetApp Data Classification

Suivez quelques étapes pour commencer à analyser vos schémas de base de données avec NetApp Data Classification.

Réviser les prérequis

Passez en revue les conditions préalables suivantes pour vous assurer que vous disposez d'une configuration prise en charge avant d'activer la classification des données.

Bases de données prises en charge

La classification des données peut analyser les schémas des bases de données suivantes :

- Service de base de données relationnelle Amazon (Amazon RDS)
- MongoDB
- MySQL
- Oracle
- PostgreSQL
- SAP HANA
- Serveur SQL (MSSQL)

La fonctionnalité de collecte de statistiques **doit être activée** dans la base de données.

Exigences relatives à la base de données

Toute base de données connectée à l'instance de classification des données peut être analysée, quel que soit l'endroit où elle est hébergée. Vous avez juste besoin des informations suivantes pour vous connecter à la base de données :

- Adresse IP ou nom d'hôte
- Port
- Nom du service (uniquement pour l'accès aux bases de données Oracle)
- Informations d'identification permettant l'accès en lecture aux schémas

  Lors du choix d'un nom d'utilisateur et d'un mot de passe, il est important d'en choisir un qui dispose de toutes les autorisations de lecture sur tous les schémas et tables que vous souhaitez analyser. Nous vous recommandons de créer un utilisateur dédié au système de classification des données avec toutes les autorisations requises.

Pour MongoDB, un rôle d'administrateur en lecture seule est requis.

Déployer l'instance de classification des données

Déployez la classification des données s'il n'existe pas déjà d'instance déployée.

Si vous numérisez des schémas de bases de données accessibles sur Internet, vous pouvez"déployer la classification des données dans le cloud" ou"déployer la classification des données dans un emplacement sur site disposant d'un accès Internet" .

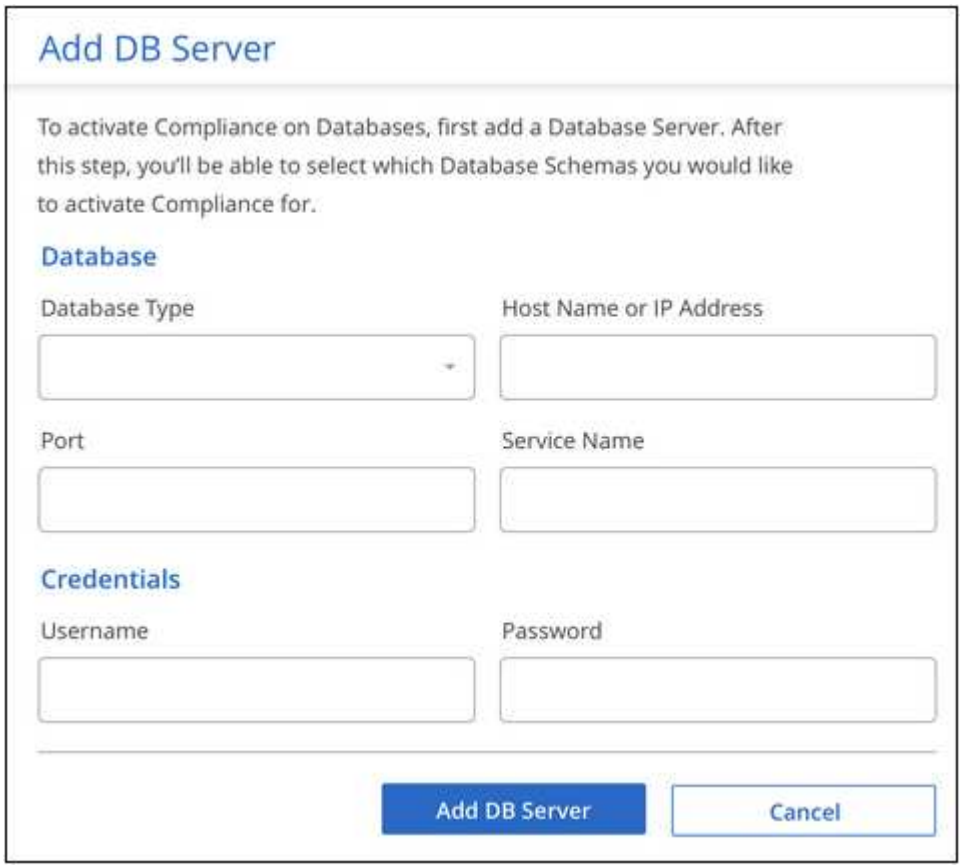
Si vous numérisez des schémas de base de données qui ont été installés sur un site sombre qui n'a pas d'accès Internet, vous devez"déployer la classification des données dans le même emplacement sur site qui n'a pas d'accès Internet" . Cela nécessite également que l'agent de console soit déployé dans le même emplacement sur site.

Ajouter le serveur de base de données

Ajoutez le serveur de base de données sur lequel résident les schémas.

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Depuis la page Configuration, sélectionnez **Ajouter un système** > **Ajouter un serveur de base de données**.

3. Saisissez les informations requises pour identifier le serveur de base de données.
   a. Sélectionnez le type de base de données.
   b. Entrez le port et le nom d'hôte ou l'adresse IP pour vous connecter à la base de données.
   c. Pour les bases de données Oracle, entrez le nom du service.
   d. Saisissez les informations d'identification pour que Data Classification puisse accéder au serveur.
   e. Sélectionnez **Ajouter un serveur de base de données**.



La base de données est ajoutée à la liste des systèmes.

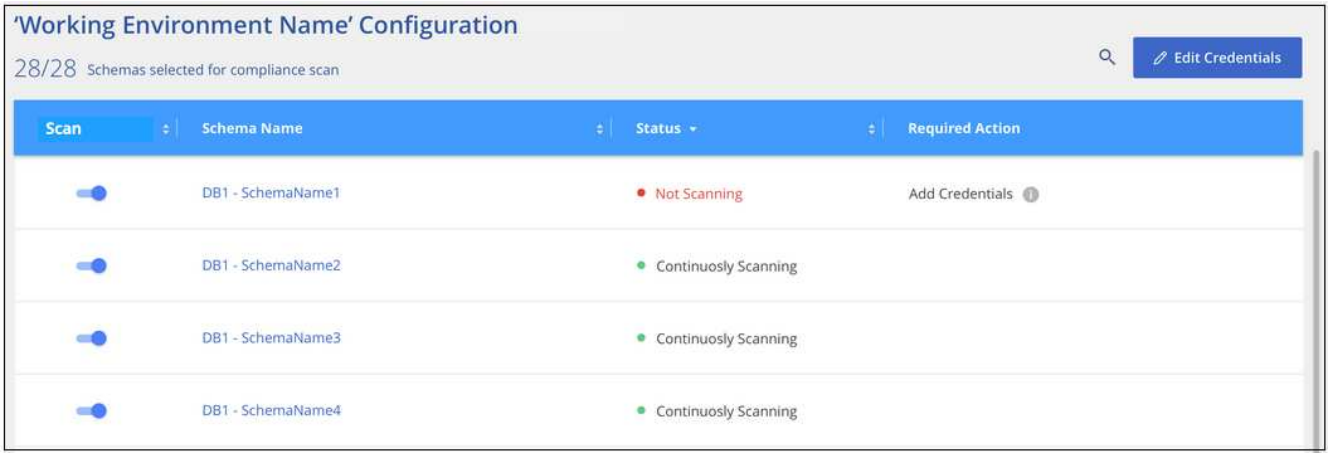
Activer et désactiver les analyses sur les schémas de base de données

Vous pouvez arrêter ou démarrer l'analyse complète de vos schémas à tout moment.

Il n'existe aucune option permettant de sélectionner des analyses de mappage uniquement pour les schémas de base de données.

1. Depuis la page Configuration, sélectionnez le bouton **Configuration** correspondant à la base de données que vous souhaitez configurer.

2. Sélectionnez les schémas que vous souhaitez analyser en déplaçant le curseur vers la droite.



Résultat

La classification des données commence à analyser les schémas de base de données que vous avez activés. Vous pouvez suivre la progression de l'analyse initiale en accédant au menu **Configuration** puis en sélectionnant la **Configuration système**. La progression de chaque analyse est affichée sous forme de barre de progression. Vous pouvez également survoler la barre de progression pour voir le nombre de fichiers analysés par rapport au nombre total de fichiers dans le volume. S'il y a des erreurs, elles apparaîtront dans la colonne Statut, à côté des actions requises pour corriger l'erreur.

Data Classification analyse vos bases de données une fois par jour ; les bases de données ne sont pas analysées en continu comme les autres sources de données.

Analyser les Google Cloud NetApp Volumes avec la NetApp Data Classification

NetApp Data Classification prend en charge Google Cloud NetApp Volumes en tant que système. Découvrez comment analyser votre système Google Cloud NetApp Volumes .

Découvrez le système Google Cloud NetApp Volumes que vous souhaitez analyser

Si le système Google Cloud NetApp Volumes que vous souhaitez analyser n'est pas déjà présent dans la NetApp Console en tant que système,"ajoutez-le à la page Systèmes" .

Déployer l'instance de classification des données

"Déployer la classification des données"s'il n'y a pas déjà une instance déployée.

La classification des données doit être déployée dans le cloud lors de l'analyse des Google Cloud NetApp

Volumes et doit être déployée dans la même région que les volumes que vous souhaitez analyser.

Remarque : le déploiement de la classification des données dans un emplacement sur site n'est actuellement pas pris en charge lors de l'analyse des Google Cloud NetApp Volumes.

Activez la classification des données dans vos systèmes

Vous pouvez activer la classification des données sur votre système Google Cloud NetApp Volumes .

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Sélectionnez la manière dont vous souhaitez analyser les volumes de chaque système. "En savoir plus sur les analyses de cartographie et de classification":
   - Pour mapper tous les volumes, sélectionnez **Mapper tous les volumes**.
   - Pour cartographier et classer tous les volumes, sélectionnez **Cartographier et classer tous les volumes**.
   - Pour personnaliser l'analyse de chaque volume, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**, puis choisissez les volumes que vous souhaitez mapper et/ou classer.

     VoirActiver et désactiver les analyses sur les volumes pour plus de détails.
3. Dans la boîte de dialogue de confirmation, sélectionnez **Approuver**.

Résultat

La classification des données commence à analyser les volumes que vous avez sélectionnés dans le système. Les résultats sont disponibles dans le tableau de bord Conformité dès que la classification des données termine les analyses initiales. Le temps nécessaire dépend de la quantité de données : de quelques minutes à quelques heures. Vous pouvez suivre la progression de l'analyse initiale dans la section **Configuration système** du menu **Configuration**. La classification des données affiche une barre de progression pour chaque analyse. Vous pouvez également survoler la barre de progression pour voir le nombre de fichiers analysés par rapport au nombre total de fichiers dans le volume.

- Par défaut, si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers de vos volumes car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, sélectionnez **Ou sélectionnez le type d'analyse pour chaque volume**. La page résultante contient un paramètre que vous pouvez activer pour que la classification des données analyse les volumes quelles que soient les autorisations.
- La classification des données analyse un seul partage de fichiers sous un volume. Si vous avez plusieurs partages dans vos volumes, vous devez analyser ces autres partages séparément en tant que groupe de partages. "En savoir plus sur cette limitation de classification des données".

Vérifiez que la classification des données a accès aux volumes

Assurez-vous que la classification des données peut accéder aux volumes en vérifiant votre réseau, vos groupes de sécurité et vos politiques d'exportation. Pour les volumes CIFS, vous devez fournir une classification des données avec les informations d'identification CIFS.

Pour les Google Cloud NetApp Volumes, la classification des données ne peut analyser que les volumes situés dans la même région que la console.

Liste de contrôle

- Assurez-vous qu'il existe une connexion réseau entre l'instance de classification des données et chaque réseau qui inclut des volumes pour Google Cloud NetApp Volumes.
- Assurez-vous que les ports suivants sont ouverts sur l'instance de classification des données :
  - Pour NFS – ports 111 et 2049.
  - Pour CIFS – ports 139 et 445.
- Assurez-vous que les stratégies d'exportation de volume NFS incluent l'adresse IP de l'instance de classification des données afin qu'elle puisse accéder aux données sur chaque volume.

Étapes

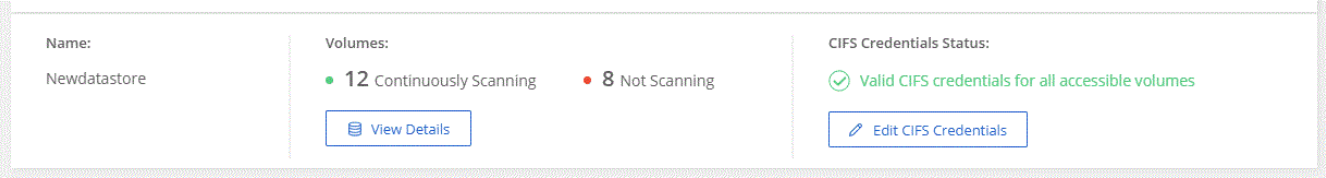
1. Dans le menu Classification des données, sélectionnez **Configuration**.

   a. Si vous utilisez CIFS (SMB), assurez-vous que les informations d'identification Active Directory sont correctes. Pour chaque système, sélectionnez **Modifier les informations d'identification CIFS**, puis saisissez le nom d'utilisateur et le mot de passe dont Data Classification a besoin pour accéder aux volumes CIFS sur le système.

      Les informations d'identification peuvent être en lecture seule, mais la fourniture d'informations d'identification d'administrateur garantit que la classification des données peut lire toutes les données nécessitant des autorisations élevées. Les informations d'identification sont stockées sur l'instance de classification des données.

      Si vous souhaitez vous assurer que les « dernières heures d'accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, il est recommandé que l'utilisateur dispose des autorisations d'écriture d'attributs dans CIFS ou des autorisations d'écriture dans NFS. Si possible, configurez l'utilisateur Active Directory en tant que membre d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

      Après avoir saisi les informations d'identification, vous devriez voir un message indiquant que tous les volumes CIFS ont été authentifiés avec succès.



2. Sur la page Configuration, sélectionnez **Afficher les détails** pour vérifier l'état de chaque volume CIFS et NFS et corriger les erreurs éventuelles.

Activer et désactiver les analyses sur les volumes

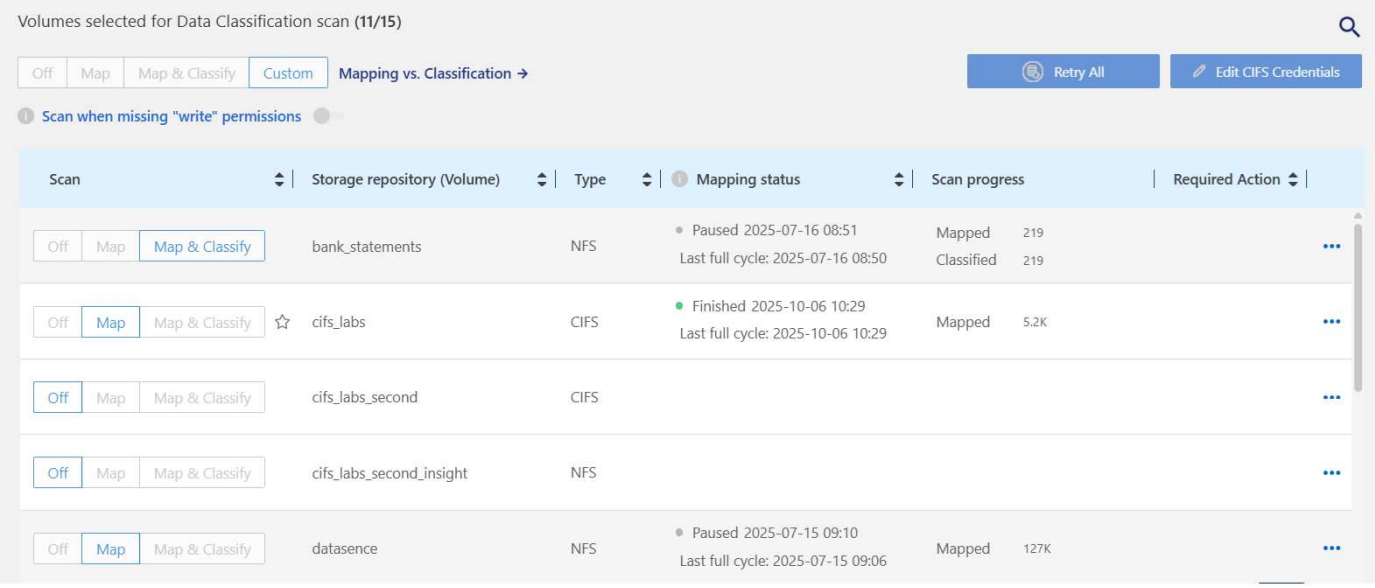
Vous pouvez démarrer ou arrêter les analyses sur n'importe quel système à tout moment à partir de la page de configuration. Vous pouvez également passer d'analyses de cartographie uniquement à des analyses de cartographie et de classification, et vice-versa. Il est recommandé d'analyser tous les volumes d'un système.

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez sélectionné le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque cette option est définie sur **Personnalisé** ou **Désactivé** dans la zone d'en-tête, vous devrez activer le mappage et/ou l'analyse complète sur chaque nouveau volume que vous ajoutez au système.

Le commutateur en haut de la page pour **Analyser en cas d'absence d'autorisations « d'écriture »** est

désactivé par défaut. Cela signifie que si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. Si vous ne vous souciez pas de savoir si la dernière heure d'accès est réinitialisée, activez l'interrupteur et tous les fichiers sont analysés quelles que soient les autorisations. "Apprendre encore plus".

Les nouveaux volumes ajoutés au système sont automatiquement analysés uniquement lorsque vous avez défini le paramètre **Carte** ou **Carte et classification** dans la zone d'en-tête. Lorsque le paramètre pour tous les volumes est **Personnalisé** ou **Désactivé**, vous devez activer l'analyse manuellement pour chaque nouveau volume que vous ajoutez.



Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Choisissez un système, puis sélectionnez **Configuration**.
3. Pour activer ou désactiver les analyses pour tous les volumes, sélectionnez **Map**, **Map & Classify** ou **Off** dans l'en-tête au-dessus de tous les volumes.

   Pour activer ou désactiver les analyses de volumes individuels, recherchez les volumes dans la liste, puis sélectionnez **Map**, **Map & Classify** ou **Off** à côté du nom du volume.

Résultat

Lorsque vous activez l'analyse, la classification des données démarre l'analyse des volumes que vous avez sélectionnés dans le système. Les résultats commencent à apparaître dans le tableau de bord Conformité dès que la classification des données démarre l'analyse. Le temps d'exécution de l'analyse dépend de la quantité de données, allant de quelques minutes à quelques heures.

Analyser les partages de fichiers avec la NetApp Data Classification

Pour analyser les partages de fichiers, vous devez d'abord créer un groupe de partages de fichiers dans NetApp Data Classification. Les groupes de partages de fichiers sont destinés aux partages NFS ou CIFS (SMB) hébergés sur site ou dans le cloud.

L'analyse des données provenant de partages de fichiers non NetApp n'est pas prise en charge dans la version principale de la classification des données.

Prérequis

Passez en revue les conditions préalables suivantes pour vous assurer que vous disposez d'une configuration prise en charge avant d'activer la classification des données.

- Les actions peuvent être hébergées n'importe où, y compris dans le cloud ou sur site. Les partages CIFS des anciens systèmes de stockage NetApp 7-Mode peuvent être analysés en tant que partages de fichiers.
  - La classification des données ne peut pas extraire les autorisations ou la « dernière heure d'accès » des systèmes 7-Mode.
  - En raison d'un problème connu entre certaines versions Linux et les partages CIFS sur les systèmes 7-Mode, vous devez configurer le partage pour utiliser uniquement SMBv1 avec l'authentification NTLM activée.
- Une connectivité réseau est nécessaire entre l'instance de classification des données et les partages.
- Vous pouvez ajouter un partage DFS (Distributed File System) en tant que partage CIFS standard. Étant donné que la classification des données ne sait pas que le partage est basé sur plusieurs serveurs/volumes combinés en un seul partage CIFS, vous risquez de recevoir des erreurs d'autorisation ou de connectivité concernant le partage lorsque le message s'applique réellement uniquement à l'un des dossiers/partages situés sur un serveur/volume différent.
- Pour les partages CIFS (SMB), assurez-vous que vous disposez des informations d'identification Active Directory qui fournissent un accès en lecture aux partages. Les informations d'identification d'administrateur sont préférables au cas où la classification des données doit analyser des données nécessitant des autorisations élevées.

  Si vous souhaitez vous assurer que les « dernières heures d'accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, il est recommandé que l'utilisateur dispose des autorisations d'écriture d'attributs dans CIFS ou des autorisations d'écriture dans NFS. Si possible, configurez l'utilisateur Active Directory en tant que membre d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

- Tous les partages de fichiers CIFS d'un groupe doivent utiliser les mêmes informations d'identification Active Directory.
- Vous pouvez mélanger les partages NFS et CIFS (en utilisant Kerberos ou NTLM). Vous devez ajouter les actions au groupe séparément. Autrement dit, vous devez effectuer le processus deux fois, une fois par protocole.
  - Vous ne pouvez pas créer un groupe de partage de fichiers qui mélange les types d'authentification CIFS (Kerberos et NTLM).
- Si vous utilisez CIFS avec l'authentification Kerberos, assurez-vous que l'adresse IP fournie est accessible à la classification des données. Les partages de fichiers ne peuvent pas être ajoutés si l'adresse IP est inaccessible.

Créer un groupe de partage de fichiers

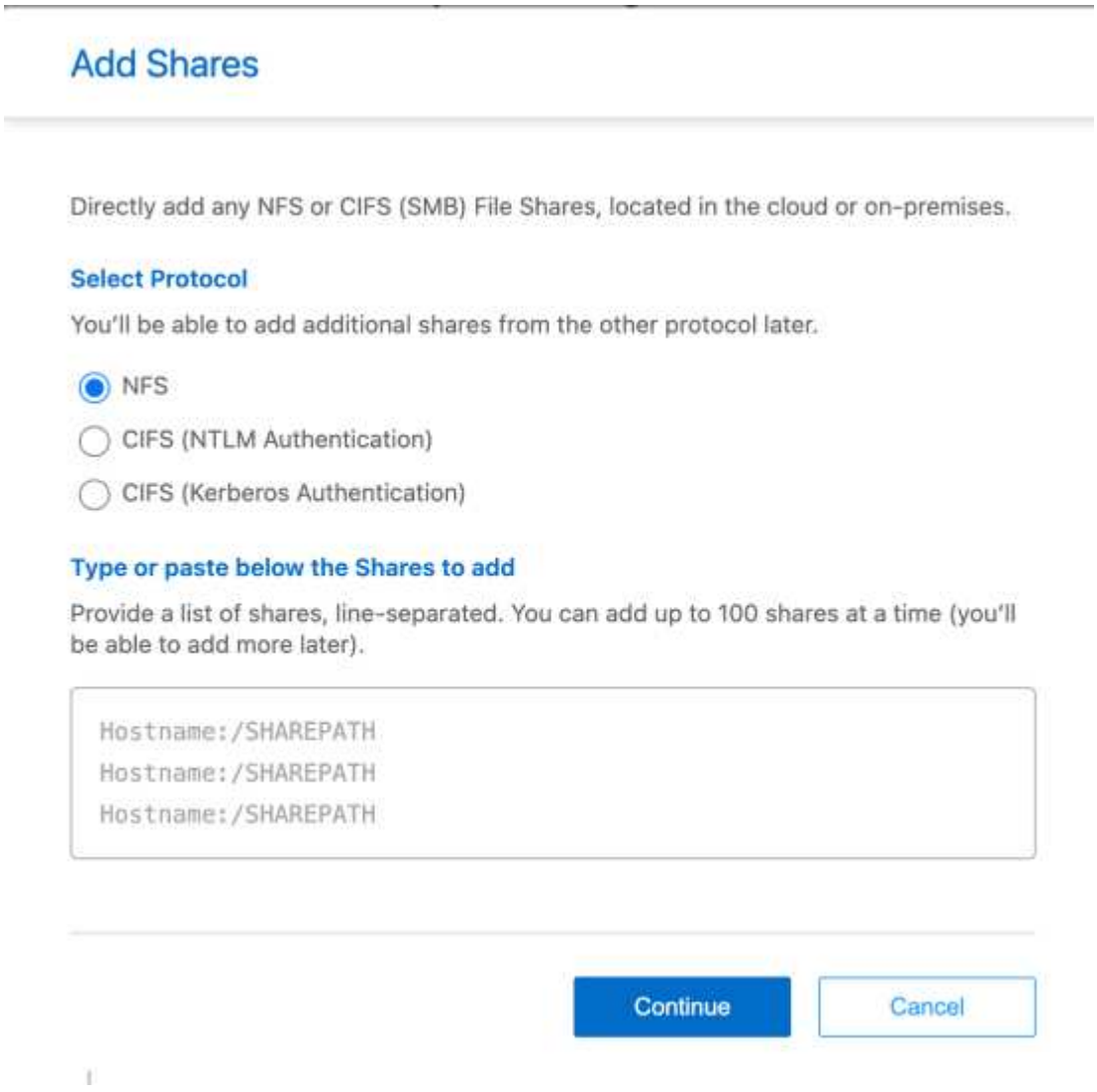
Lorsque vous ajoutez des partages de fichiers au groupe, vous devez utiliser le format `<host_name>:/<share_path>`.

Vous pouvez ajouter des partages de fichiers individuellement ou saisir une liste séparée par des lignes des

partages de fichiers que vous souhaitez analyser. Vous pouvez ajouter jusqu'à 100 actions à la fois.

Étapes

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Depuis la page Configuration, sélectionnez **Ajouter un système** > **Ajouter un groupe de partages de fichiers**.
3. Dans la boîte de dialogue Ajouter un groupe de partages de fichiers, saisissez le nom du groupe de partages, puis sélectionnez **Continuer**.
4. Sélectionnez le protocole pour les partages de fichiers que vous ajoutez.



   a. Si vous ajoutez des partages CIFS avec l'authentification NTLM, entrez les informations d'identification Active Directory pour accéder aux volumes CIFS. Bien que les informations d'identification en lecture seule soient prises en charge, il est recommandé de fournir un accès complet avec les informations d'identification d'administrateur. Sélectionnez **Enregistrer**.

5. Ajoutez les partages de fichiers que vous souhaitez analyser (un partage de fichiers par ligne). Sélectionnez ensuite **Continuer**.
6. Une boîte de dialogue de confirmation affiche le nombre de partages qui ont été ajoutés.

   Si la boîte de dialogue répertorie des partages qui n'ont pas pu être ajoutés, capturez ces informations afin

de pouvoir résoudre le problème. Si le problème concerne une convention de dénomination, vous pouvez rajouter le partage avec un nom corrigé.

7. Configurer l'analyse sur le volume :
   - Pour activer les analyses de mappage uniquement sur les partages de fichiers, sélectionnez **Map**.
   - Pour activer les analyses complètes sur les partages de fichiers, sélectionnez **Mappez et classez**.
   - Pour désactiver l'analyse sur les partages de fichiers, sélectionnez **Désactivé**.

> Le commutateur en haut de la page pour **Analyser lorsque les autorisations « attributs d'écriture » sont manquantes** est désactivé par défaut. Cela signifie que si Data Classification ne dispose pas d'autorisations d'attributs d'écriture dans CIFS ou d'autorisations d'écriture dans NFS, le système n'analysera pas les fichiers car Data Classification ne peut pas rétablir l'« heure du dernier accès » à l'horodatage d'origine. + Si vous activez **Analyser en cas d'absence d'autorisations « attributs d'écriture »**, l'analyse réinitialise l'heure du dernier accès et analyse tous les fichiers, quelles que soient les autorisations. + Pour en savoir plus sur l'horodatage du dernier accès, voir"Métadonnées collectées à partir de sources de données dans la classification des données" .

Résultat

La classification des données commence à analyser les fichiers dans les partages de fichiers que vous avez ajoutés. Tu peuxSuivre la progression de la numérisation et afficher les résultats de l'analyse dans le **Tableau de bord**.

> Si l'analyse ne se termine pas correctement pour une configuration CIFS avec authentification Kerberos, vérifiez l'onglet **Configuration** pour détecter d'éventuelles erreurs.

Modifier un groupe de partage de fichiers

Après avoir créé un groupe de partages de fichiers, vous pouvez modifier le protocole CIFS ou ajouter et supprimer des partages de fichiers.

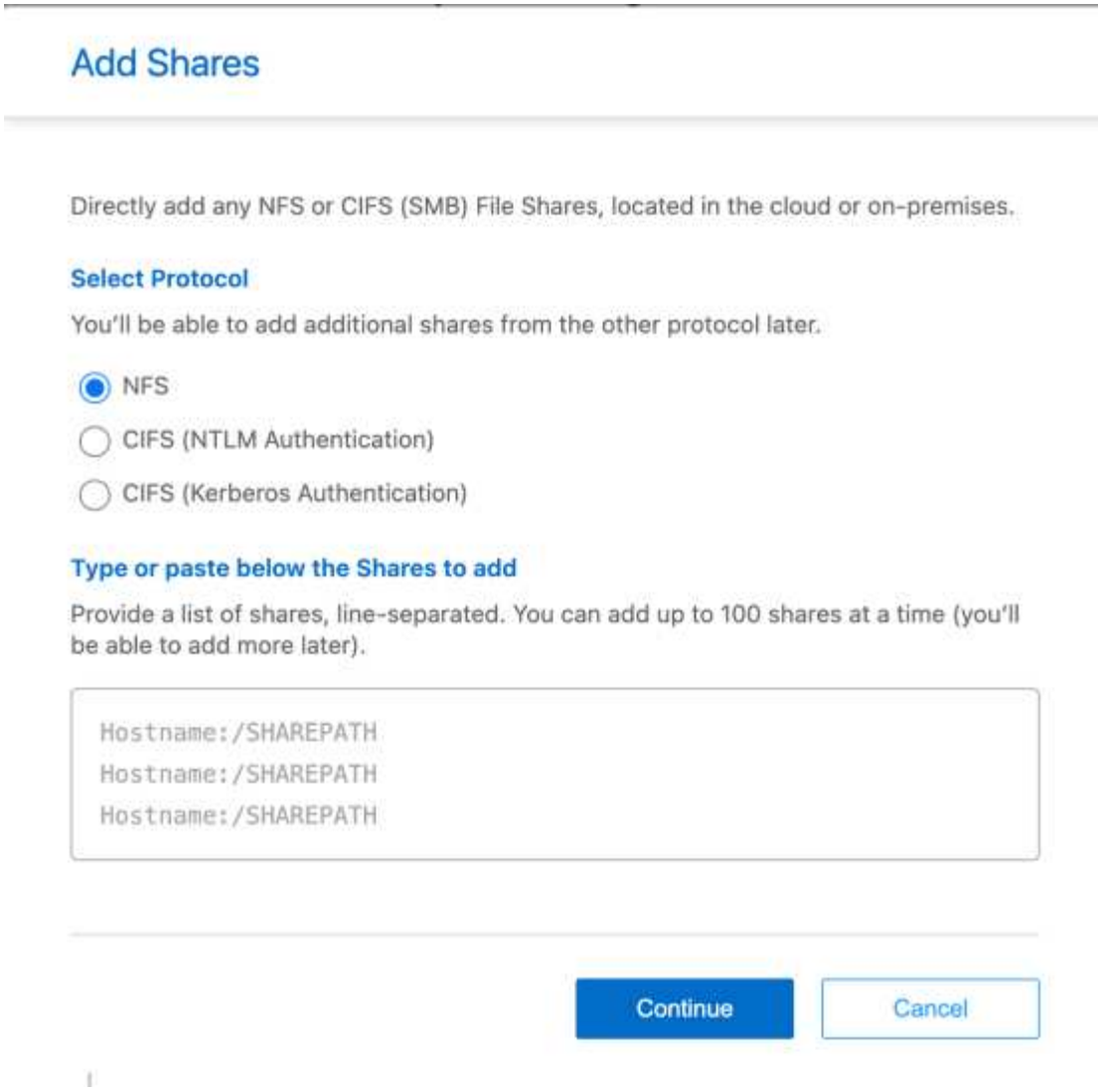
Modifier la configuration du protocole CIFS

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Depuis la page Configuration, sélectionnez le groupe de partages de fichiers que vous souhaitez modifier.
3. Sélectionnez **Modifier les informations d'identification CIFS**.

4. Choisissez la méthode d'authentification : **NTLM** ou **Kerberos**.
5. Saisissez le **Nom d'utilisateur** et le **Mot de passe** d'Active Directory.
6. Sélectionnez **Enregistrer** pour terminer le processus.

Ajouter des partages de fichiers aux analyses

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Depuis la page Configuration, sélectionnez le groupe de partages de fichiers que vous souhaitez modifier.
3. Sélectionnez **+ Ajouter des partages**.
4. Sélectionnez le protocole pour les partages de fichiers que vous ajoutez.



Si vous ajoutez des partages de fichiers à un protocole que vous avez déjà configuré, aucune modification n'est requise.

Si vous ajoutez des partages de fichiers avec un deuxième protocole, assurez-vous d'avoir correctement configuré l'authentification comme détaillé dans le"prérequis" .

5. Ajoutez les partages de fichiers que vous souhaitez analyser (un partage de fichiers par ligne) en utilisant le format `<host_name>:/<share_path>` .
6. Sélectionnez **Continuer** pour terminer l'ajout des partages de fichiers.

Supprimer un partage de fichiers des analyses

1. Dans le menu Classification des données, sélectionnez **Configuration**.
2. Sélectionnez le système dont vous souhaitez supprimer les partages de fichiers.
3. Sélectionnez **Configuration**.
4. Depuis la page Configuration, sélectionnez les Actions ••• pour le partage de fichiers que vous souhaitez supprimer.
5. Dans le menu Actions, sélectionnez **Supprimer le partage**.

Suivre la progression de la numérisation

Vous pouvez suivre la progression de l'analyse initiale.

1. Sélectionnez le menu **Configuration**.
2. Sélectionnez la **Configuration système**.
3. Pour le référentiel de stockage, vérifiez la colonne Progression de l'analyse pour afficher son état.

Analyser les données StorageGRID avec la NetApp Data Classification

Suivez quelques étapes pour commencer à analyser les données dans StorageGRID directement avec NetApp Data Classification.

Examiner les exigences de StorageGRID

Passez en revue les conditions préalables suivantes pour vous assurer que vous disposez d'une configuration prise en charge avant d'activer la classification des données.

- Vous devez disposer de l'URL du point de terminaison pour vous connecter au service de stockage d'objets.
- Vous devez disposer de la clé d'accès et de la clé secrète de StorageGRID afin que la classification des données puisse accéder aux buckets.

Déployer l'instance de classification des données

Déployez la classification des données s'il n'existe pas déjà d'instance déployée.

Si vous numérisez des données de StorageGRID accessibles sur Internet, vous pouvez"déployer la classification des données dans le cloud" ou"déployer la classification des données dans un emplacement sur site disposant d'un accès Internet" .

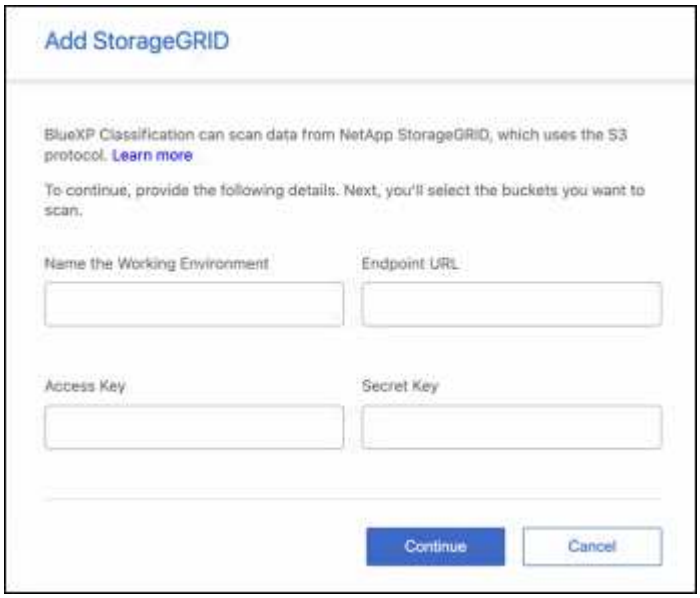
Si vous numérisez des données à partir de StorageGRID qui a été installé sur un site sombre qui n'a pas d'accès Internet, vous devez"déployer la classification des données dans le même emplacement sur site qui n'a pas d'accès Internet" . Cela nécessite également que l'agent de console soit déployé dans le même emplacement sur site.

Ajoutez le service StorageGRID à la classification des données

Ajoutez le service StorageGRID .

Étapes

1. Dans le menu Classification des données, sélectionnez l'option **Configuration**.
2. Depuis la page Configuration, sélectionnez **Ajouter un système** > **Ajouter StorageGRID**.
3. Dans la boîte de dialogue Ajouter un service StorageGRID , saisissez les détails du service StorageGRID et sélectionnez **Continuer**.
   a. Entrez le nom que vous souhaitez utiliser pour le système. Ce nom doit refléter le nom du service StorageGRID auquel vous vous connectez.
   b. Saisissez l'URL du point de terminaison pour accéder au service de stockage d'objets.
   c. Saisissez la clé d'accès et la clé secrète afin que la classification des données puisse accéder aux buckets dans StorageGRID.



Résultat

StorageGRID est ajouté à la liste des systèmes.

Activer et désactiver les analyses sur les buckets StorageGRID

Après avoir activé la classification des données sur StorageGRID, l'étape suivante consiste à configurer les buckets que vous souhaitez analyser. La classification des données découvre ces compartiments et les affiche dans le système que vous avez créé.

Étapes

1. Dans la page Configuration, recherchez le système StorageGRID .
2. Sur la mosaïque système StorageGRID , sélectionnez **Configuration**.
3. Effectuez l'une des étapes suivantes pour activer ou désactiver l'analyse :
   - Pour activer les analyses de mappage uniquement sur un bucket, sélectionnez **Carte**.
   - Pour activer les analyses complètes sur un bucket, sélectionnez **Cartographier et classer**.
   - Pour désactiver l'analyse sur un bucket, sélectionnez **Désactivé**.

Résultat

La classification des données commence à analyser les compartiments que vous avez activés. Vous pouvez suivre la progression de l'analyse initiale en accédant au menu **Configuration** puis en sélectionnant la **Configuration système**. La progression de chaque analyse est affichée sous forme de barre de progression. Vous pouvez également survoler la barre de progression pour voir le nombre de fichiers analysés par rapport au nombre total de fichiers dans le volume. S'il y a des erreurs, elles apparaîtront dans la colonne Statut, à côté de l'action requise pour corriger l'erreur.

Intégrez votre Active Directory à la NetApp Data Classification

Vous pouvez intégrer un Active Directory global à NetApp Data Classification pour améliorer les résultats rapportés par Data Classification sur les propriétaires de fichiers et sur les utilisateurs et groupes ayant accès à vos fichiers.

Lorsque vous configurez certaines sources de données (répertoriées ci-dessous), vous devez saisir les informations d'identification Active Directory pour que la classification des données analyse les volumes CIFS. Cette intégration fournit une classification des données avec les détails du propriétaire du fichier et des autorisations pour les données qui résident dans ces sources de données. L'Active Directory saisi pour ces sources de données peut différer des informations d'identification Active Directory globales que vous saisissez ici. La classification des données recherchera dans tous les annuaires Active Directory intégrés les détails des utilisateurs et des autorisations.

Cette intégration fournit des informations supplémentaires aux emplacements suivants dans la classification des données :

- Vous pouvez utiliser le « Propriétaire du fichier »"filtre" et voir les résultats dans les métadonnées du fichier dans le volet Investigation. Au lieu du propriétaire du fichier contenant le SID (Security IDentifier), il est renseigné avec le nom d'utilisateur réel.

  Vous pouvez également afficher plus de détails sur le propriétaire du fichier : nom du compte, adresse e-mail et nom du compte SAM, ou afficher les éléments appartenant à cet utilisateur.

- Tu peux voir"autorisations complètes du fichier" pour chaque fichier et répertoire lorsque vous cliquez sur le bouton « Afficher toutes les autorisations ».
- Dans le"Tableau de bord de gouvernance" , le panneau Autorisations d'ouverture affichera un niveau de détail plus élevé sur vos données.

Les SID des utilisateurs locaux et les SID des domaines inconnus ne sont pas traduits en nom d'utilisateur réel.

Sources de données prises en charge

Une intégration Active Directory avec la classification des données peut identifier les données à partir des sources de données suivantes :

- Systèmes ONTAP sur site
- Cloud Volumes ONTAP
- Azure NetApp Files
- FSx pour ONTAP

Connectez-vous à votre serveur Active Directory

Après avoir déployé la classification des données et activé l'analyse sur vos sources de données, vous pouvez intégrer la classification des données à votre Active Directory. Active Directory est accessible à l'aide d'une adresse IP de serveur DNS ou d'une adresse IP de serveur LDAP.

Les informations d'identification Active Directory peuvent être en lecture seule, mais la fourniture d'informations d'identification d'administrateur garantit que la classification des données peut lire toutes les données nécessitant des autorisations élevées. Les informations d'identification sont stockées sur l'instance de classification des données.

Pour les volumes/partages de fichiers CIFS, si vous souhaitez vous assurer que les « heures de dernier accès » de vos fichiers ne sont pas modifiées par les analyses de classification des données, l'utilisateur doit disposer de l'autorisation d'écriture des attributs. Si possible, nous vous recommandons de faire en sorte que l'utilisateur configuré par Active Directory fasse partie d'un groupe parent de l'organisation disposant d'autorisations sur tous les fichiers.

Exigences

- Vous devez déjà avoir un Active Directory configuré pour les utilisateurs de votre entreprise.
- Vous devez disposer des informations pour Active Directory :
  - Adresse IP du serveur DNS ou plusieurs adresses IP

    ou

  Adresse IP du serveur LDAP ou plusieurs adresses IP
  - Nom d'utilisateur et mot de passe pour accéder au serveur
  - Nom de domaine (nom Active Directory)
  - Que vous utilisiez ou non un LDAP sécurisé (LDAPS)
  - Port du serveur LDAP (généralement 389 pour LDAP et 636 pour LDAP sécurisé)
- Les ports suivants doivent être ouverts pour la communication sortante par l'instance de classification des données :

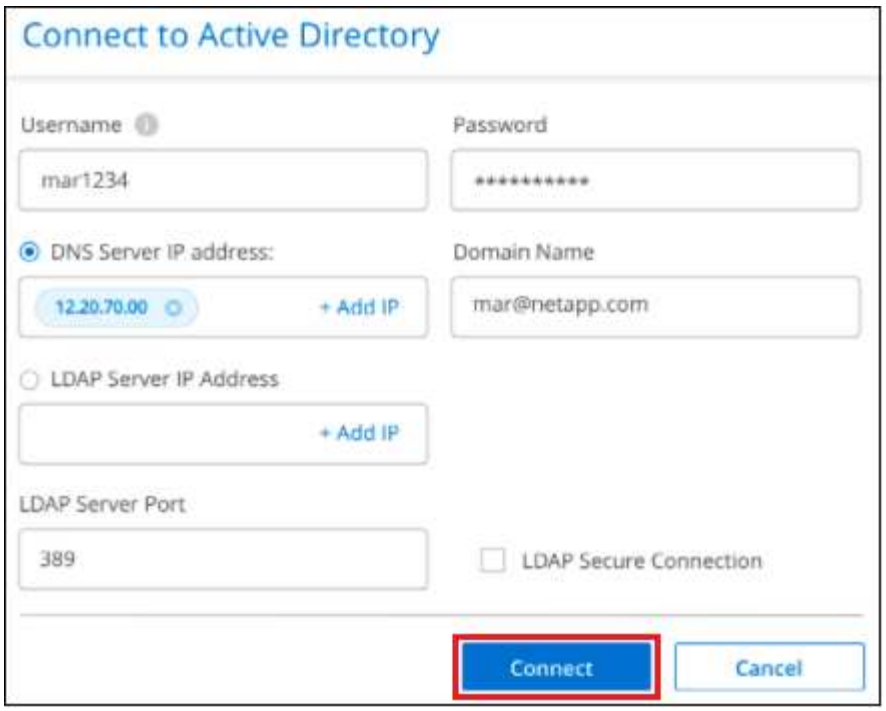
Protocole	Port	Destination	But
TCP et UDP	389	Active Directory	LDAP
TCP	636	Active Directory	LDAP sur SSL
TCP	3268	Active Directory	Catalogue mondial
TCP	3269	Active Directory	Catalogue global via SSL

Étapes

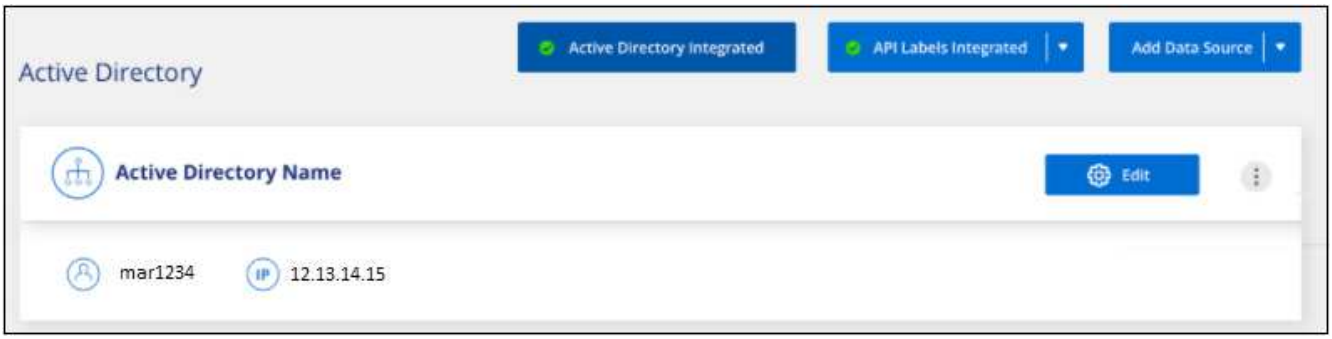
1. Depuis la page Configuration de la classification des données, cliquez sur **Ajouter Active Directory**.

2. Dans la boîte de dialogue Se connecter à Active Directory, entrez les détails d'Active Directory et cliquez sur **Connecter**.

   Vous pouvez ajouter plusieurs adresses IP, si nécessaire, en sélectionnant **Ajouter une IP**.



La classification des données s'intègre à Active Directory et une nouvelle section est ajoutée à la page de configuration.



Gérez votre intégration Active Directory

Si vous devez modifier des valeurs dans votre intégration Active Directory, cliquez sur le bouton **Modifier** et effectuez les modifications.

Vous pouvez également supprimer l'intégration en sélectionnant l'option bouton puis **Supprimer Active Directory**.

Informations sur le copyright

Informations sur les marques commerciales