Déployer la classification des données

NetApp Data Classification

NetApp
February 11, 2026

Sommaire

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.

1 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.

2 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

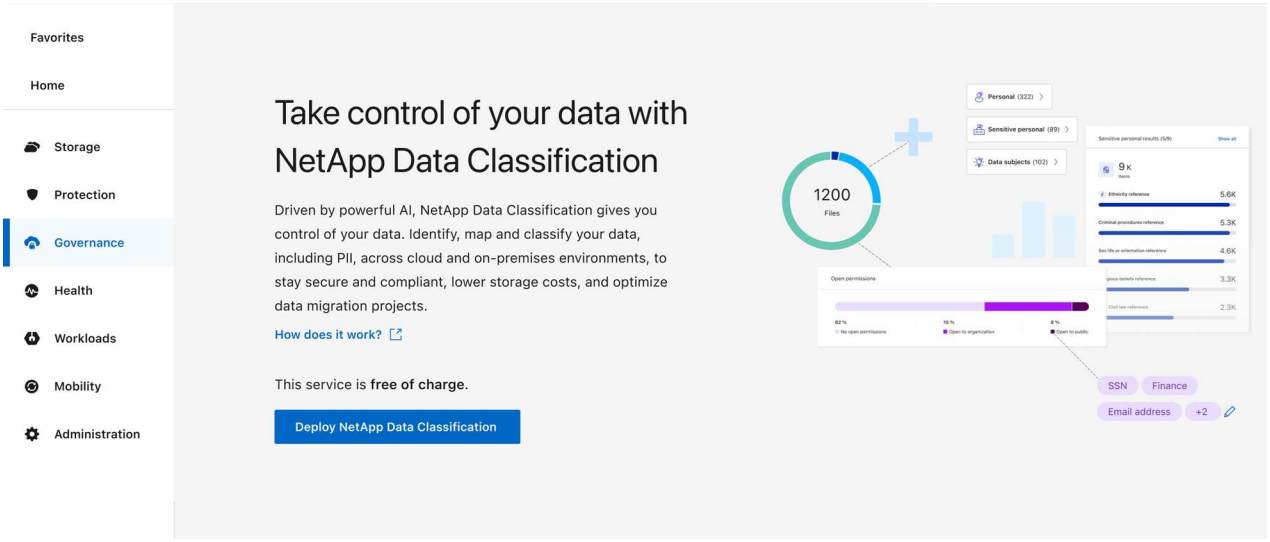
Suivez ces étapes pour déployer une instance de classification des données dans le cloud. L'agent de console déploiera l'instance dans le cloud, puis installera le logiciel de classification des données sur cette instance.

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**.

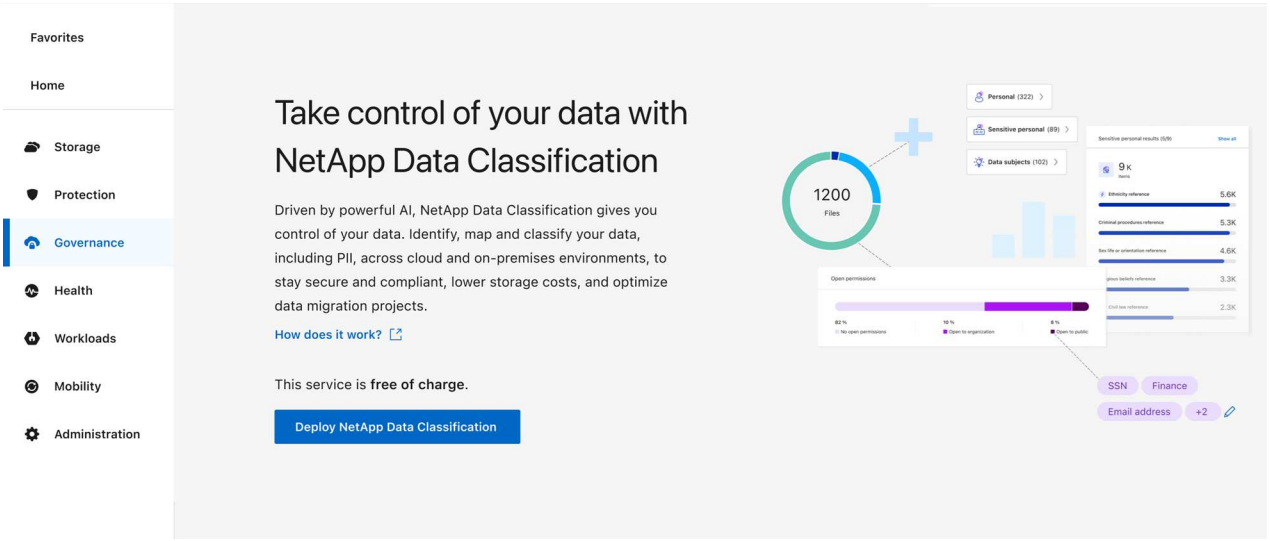


2. Depuis la page *Installation*, sélectionnez **Déployer > Déployer** pour utiliser la taille d'instance « Grande » et démarrer l'assistant de déploiement cloud.
3. L'assistant affiche la progression au fur et à mesure des étapes de déploiement. Lorsque des entrées sont requises ou si vous rencontrez 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**.



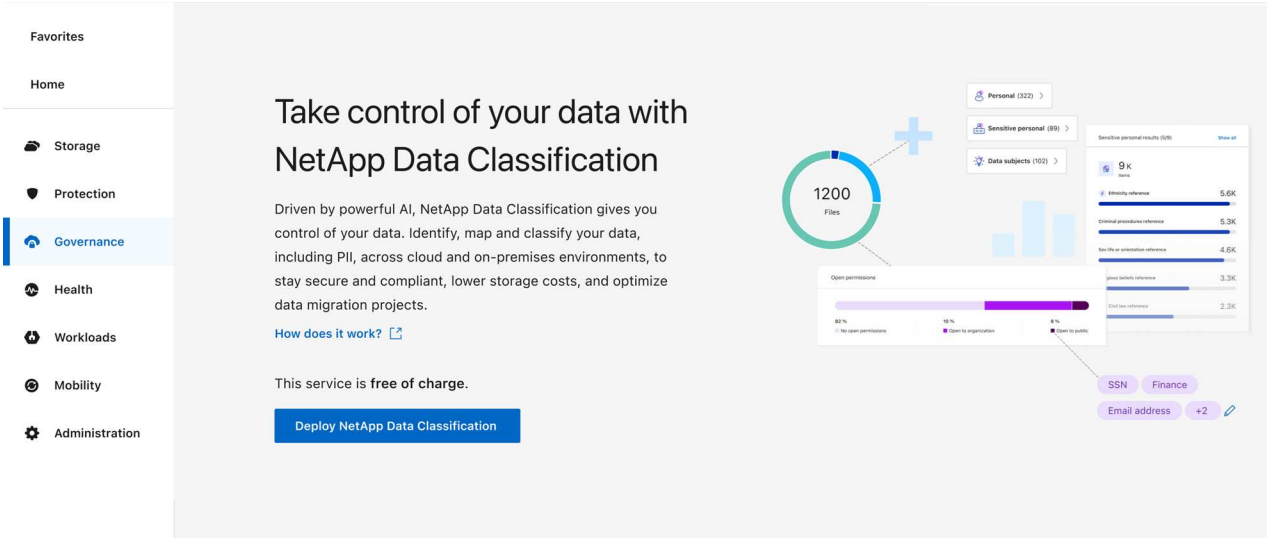
2. Sélectionnez **Déployer** pour démarrer l'assistant de déploiement cloud.

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.

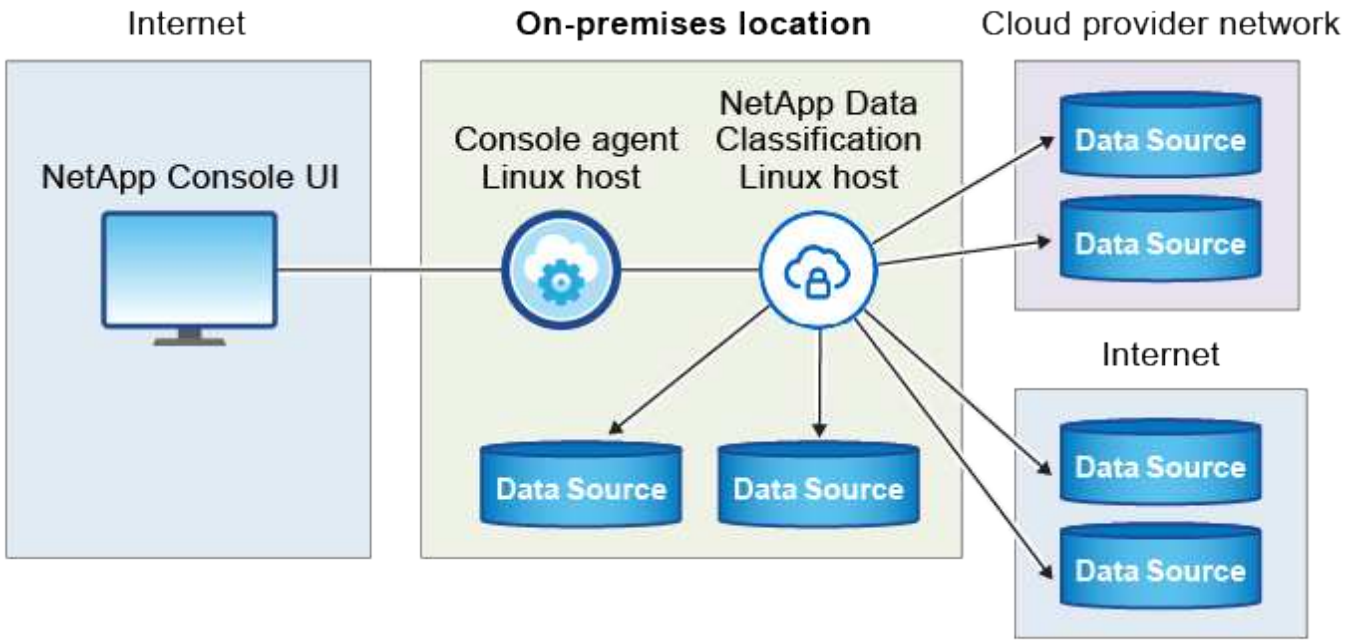
Installer NetApp Data Classification sur un hôte disposant d'un accès Internet

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.

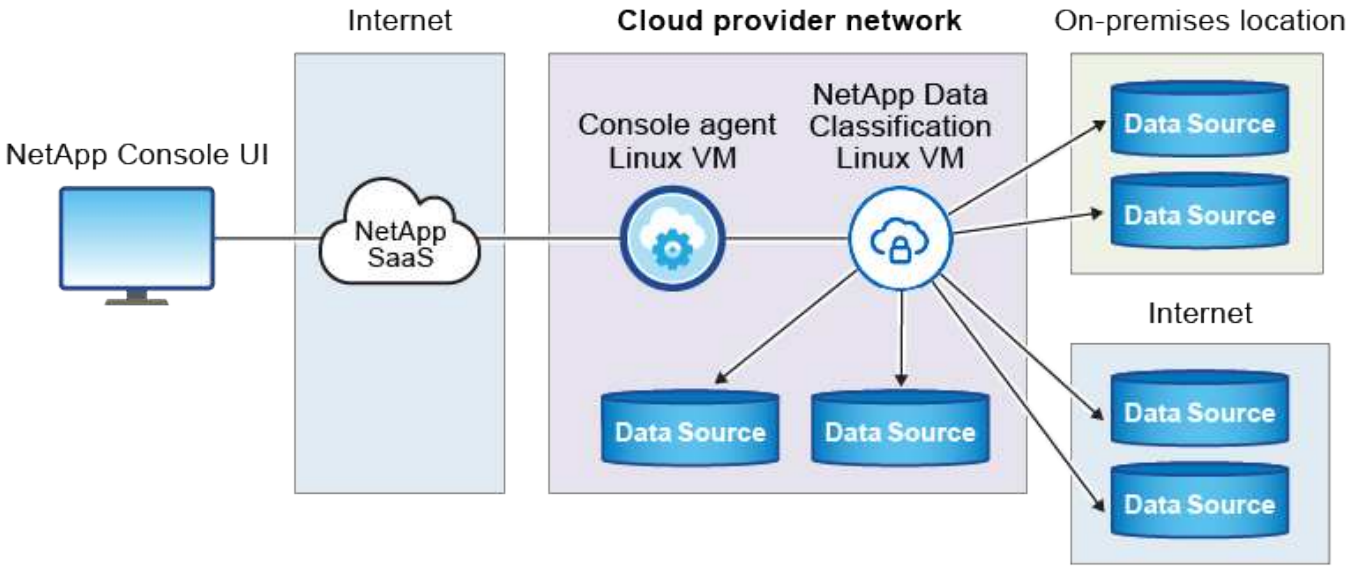
L'installation sur site est une bonne option si vous préférez analyser les systèmes ONTAP sur site à l'aide d'une instance de classification des données également située sur site. Ce n'est pas une exigence. Le logiciel fonctionne de la même manière quelle que soit la méthode d'installation choisie.

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.

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

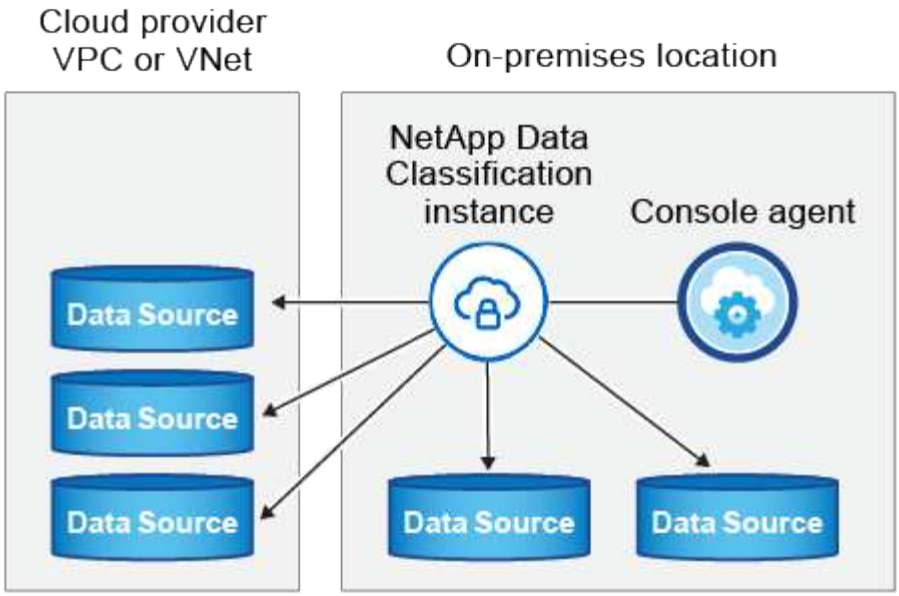
Vous devez vous assurer que tous les ports requis sont ouverts pour la communication entre l'agent de console, la classification des données, Active Directory et vos sources de données.

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, 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.

Type de connexion	Ports	Description
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.
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.

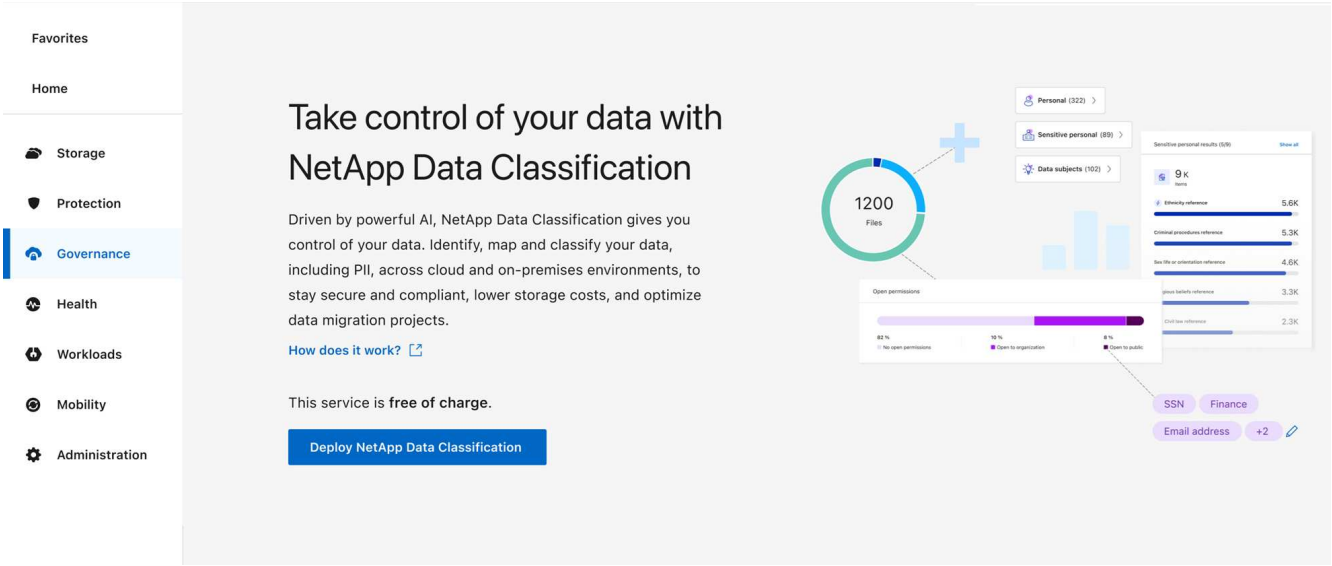
- Vérifiez que votre environnement hors ligne répond aux exigences requisesautorisations et connectivité .

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**

Taille du système	processeur	RAM (la mémoire d'échange doit être désactivée)	Disque
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.

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.
\ 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, 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.