



Documentation AI Data Engine

AI Data Engine

NetApp

March 13, 2026

This PDF was generated from <https://docs.netapp.com/fr-fr/ai-data-engine/index.html> on March 13, 2026.
Always check docs.netapp.com for the latest.

Sommaire

Documentation AI Data Engine	1
Notes de version	2
Nouveautés dans AI Data Engine	2
Nouveautés de la version initiale de AIDE 9.18.1	2
Limitations connues de AI Data Engine	3
Limitations connues pour la version initiale d'AIDE 9.18.1	3
Commencer	6
Découvrez votre système AI Data Engine	6
Découvrez AI Data Engine	6
Architecture de l'AI Data Engine	7
Composantes et responsabilités d'AIDE par rôle	11
Démarrage rapide pour AI Data Engine	17
Installer AIDE	18
Exigences pour l'installation de AI Data Engine	18
Installez le système de stockage AFX pour AI Data Engine	19
Installez vos nœuds de calcul de données	19
Configurez votre système AIDE	29
Configurer AI Data Engine	29
Installez une licence AIDE dans votre système AFX	31
Configurer OpenID Connect pour AIDE dans ONTAP	33
Configurer des espaces de travail	36
Préparez les données pour AI Data Engine	36
Créez un espace de travail dans AI Data Engine	37
Créer un espace de travail	38
Examiner les détails de l'espace de travail	38
Actualisation et versionnage de l'espace de travail	39
Attribuer l'accès utilisateur aux espaces de travail AI Data Engine	39
Administrer et surveiller	41
Surveiller les processus du cluster	41
Afficher l'état du système AIDE et l'état du cluster	41
Consultez les informations pour optimiser votre système AIDE	43
Afficher les événements système AIDE, les tâches et le journal des audits	44
Gérer les espaces de travail AI Data Engine	46
Vérifier l'état de l'espace de travail	46
Modifier les propriétés de l'espace de travail et le calendrier d'actualisation	47
Ajouter des containers de données à un espace de travail existant	47
Supprimer les containers de données d'un workspace	48
Gérer les utilisateurs de l'espace de travail	48
Supprimer un espace de travail	48
Mettez à niveau et entretenez votre système AIDE	49
Mises à jour et compatibilité du système AI Data Engine	49
Mettre à jour AI Data Engine software	51
Ajoutez des nœuds de calcul de données à votre cluster AIDE	52

Remplacez un nœud de votre cluster AIDE	54
Gérer la vectorisation et les collections de données	56
Démarrage rapide Data-to-RAG pour AI Data Engine	56
Explorez les métadonnées de l'espace de travail dans AI Data Engine Console	57
Sign in à AI Data Engine Console en tant qu'ingénieur de données ou data scientist	57
Consultez vos espaces de travail accessibles	57
Et ensuite ?	58
Créez des collectes de données dans AI Data Engine Console	58
Créer une collecte de données à partir des métadonnées de workspace	58
Publier une collecte de données	59
Mettre à jour ou supprimer une collecte de données	59
Et ensuite ?	60
Consultez les collections de données dans AI Data Engine	60
Afficher les collectes de données à l'échelle du cluster	61
Surveiller les tâches et les événements liés à la collecte	61
Afficher les collectes de données depuis la console AIDE	62
Et ensuite ?	63
Mettre en place des garde-fous	64
Définissez vos politiques de Data Guardrails dans AI Data Engine pour votre domaine de données	64
Comprendre les types de stratégies	64
Activer les classificateurs	64
Gérer les catégories de classificateur	65
Créer et gérer des politiques de Data Guardrails	66
Comment les politiques interagissent avec les espaces de travail	67
Informations connexes	67
FAQ pour NetApp AI Data Engine	68
Principes de base d'AIDE	68
Utilisateurs et rôles	68
Exigences et déploiement	68
Gestion et interfaces	69
Fonctionnalités et capacités	69
Intégration et interopérabilité	70
Déploiement et licences	70
Mentions légales	72
Copyright	72
Marques déposées	72
Brevets	72
Politique de confidentialité	72
Open source	72
AI Data Engine	72

Documentation AI Data Engine

Notes de version

Nouveautés dans AI Data Engine

AI Data Engine (AIDE) 9.18.1 est la version initiale de la plateforme de gestion des données AI de NetApp. Cette version introduit un moteur de métadonnées et des flux de travail de gestion permettant aux organisations de cataloguer et d'organiser les données non structurées pour les charges de travail d'IA, fournissant la base pour des capacités avancées de gouvernance et de vectorisation. La gouvernance avancée (garde-fous) et la vectorisation sont disponibles pour les clients disposant des licences AI Data Engine appropriées.

Nouveautés de la version initiale de AIDE 9.18.1

AIDE 9.18.1 introduit les fonctionnalités fondamentales suivantes :

"Metadata Engine pour le catalogage des données d'IA"

La version initiale comprend un Metadata Engine qui catalogue les fichiers et les objets sur les clusters ONTAP.

Les principales caractéristiques sont les suivantes :

- Extraction automatisée des métadonnées (attributs de base et étendus, balises d'objet) à partir de volumes ONTAP locaux et distants sur des clusters appariés.
- API REST centralisées pour l'interrogation et le filtrage, destinées aux applications nécessitant une vue globale des données d'entreprise.
- Stockage de métadonnées évolutif.
- Extraction automatique des métadonnées déclenchée lors de la création de l'espace de travail.

"Gestion de l'espace de travail"

Les espaces de travail permettent un regroupement logique des sources de données (volumes) pour les projets d'IA.

La version initiale prend en charge :

- Création d'espaces de travail couvrant des volumes ONTAP locaux et distants (à l'aide du peering de clusters).
- Attribution de contrôles d'accès aux espaces de travail, prenant en charge les environnements multi-utilisateurs et multi-locataires.
- Extraction automatique des métadonnées et remplissage du catalogue lors de la création de workspace.

"Synchronisation des données pour une actualisation automatisée des données"

Data Sync permet de maintenir à jour les catalogues de métadonnées et les collections de données au fur et à mesure que les données sources changent, sans intervention manuelle.

Les principales caractéristiques sont les suivantes :

- Synchronisation automatisée des données à partir de clusters ONTAP distants ou locaux à l'aide de la réplication SnapMirror pilotée par des politiques.

- Des mises à jour incrémentales qui ne propagent que les données modifiées, réduisant la surcharge.
- Intervalles d'actualisation configurables par espace de travail.
- Surveillance au niveau de l'espace de travail de l'état de synchronisation et de l'activité.

"Configuration et gestion du cluster"

La version initiale comprend les workflows suivants :

- Découverte et ajout de nœuds de calcul de données (DCN) lors de la configuration du cluster.
- Création de machines virtuelles de stockage de métadonnées dédiées pour le Metadata Engine.
- Configuration des interfaces de service AI Data Engine pour l'accès aux métadonnées à l'échelle du cluster.
- Interconnexion avec d'autres ONTAP clusters pour étendre le catalogage des métadonnées à l'ensemble du patrimoine de données.

"Authentification OpenID Connect (OIDC)"

- Authentification basée sur OIDC/OAuth pour un accès sécurisé à ONTAP System Manager et Data Engine Console avec Microsoft Entra ID et Active Directory Federation Services (ADFS).
- Contrôles d'accès basés sur les rôles pour la gestion de l'espace de travail et des métadonnées.

"Capacités avancées de curation et de gouvernance des données"

Les fonctionnalités suivantes sont disponibles pour les clients disposant des licences AI Data Engine appropriées :

- **Vectorisation et RAG** : Création de collectes de données, d'embeddings et de points de terminaison de récupération dans la AI Data Engine Console, à l'aide des métadonnées des espaces de travail AIDE.
- **Gouvernance basée sur des garde-fous** : Définition des politiques de garde-fous dans l'AI Data Engine Console et association de ces politiques avec les espaces de travail dans ONTAP System Manager.

Matériel et plateformes pris en charge

AI Data Engine 9.18.1 s'exécute sur des clusters de la plateforme de données ONTAP AI qui combinent :

- Nœuds de stockage AFX 1K
- Nœuds de calcul de données NetApp

Informations connexes

- ["Limitations connues de AI Data Engine"](#)
- ["Découvrez l'architecture et les composants d'AI Data Engine"](#)

Limitations connues de AI Data Engine

Les limitations connues indiquent les plateformes, les appareils ou les fonctions non pris en charge par cette version du produit, ou qui ne fonctionnent pas correctement avec celle-ci. Examinez attentivement ces limitations.

Limitations connues pour la version initiale d'AIDE 9.18.1

Ces limitations s'appliquent au Metadata Engine, aux nœuds de calcul de données et aux flux de travail de

Exigences et gestion des nœuds de calcul

- **Exigence minimale pour les nœuds de calcul de données**

Les clusters AIDE nécessitent un minimum et un maximum de 3 nœuds de calcul de données (DCN) pour la fonctionnalité Metadata Engine. Les clusters comportant moins de 3 DCN ne peuvent pas activer les fonctionnalités Metadata Engine.

- **Aucune prise en charge de l'ajout de nœuds DCN via la NetApp Console**

Les mises à niveau et les ajouts de nœuds DCN doivent être effectués à l'aide de ONTAP System Manager, pas via NetApp Console.

Sources de données prises en charge

- **Aucune prise en charge des compartiments ONTAP S3 ni de StorageGRID comme sources de données**

Seuls les volumes ONTAP (locaux ou distants) sont pris en charge comme sources de données pour les espaces de travail et le catalogage des métadonnées. Les compartiments ONTAP S3 et les objets StorageGRID ne peuvent pas être ajoutés aux espaces de travail et ne sont pas indexés par le Metadata Engine dans cette version.

- **Aucune prise en charge de la création d'espace de travail avec les volumes FlexCache®**

FlexCache volumes ne peuvent pas être ajoutés en tant que sources de données aux espaces de travail.

Limitations de mise à jour et de restauration du logiciel

- **Mises à jour logicielles manuelles pour les nœuds DCN uniquement**

Les mises à jour logicielles automatiques pour DCN clusters ne sont pas prises en charge dans AIDE 9.18.1. Le logiciel des nœuds DCN ne peut être mis à jour que par le chargement de l'image depuis un client local. Le téléchargement d'images depuis des serveurs externes (HTTP/FTP) n'est pas pris en charge.

- **Aucune restauration du logiciel du cluster DCN**

Le logiciel du cluster DCN ne peut pas être restauré à une version antérieure. Seules les mises à niveau vers des versions ultérieures sont autorisées.

- **Aucune restauration ONTAP des clusters de stockage AFX**

Les clusters de stockage AFX ne peuvent pas être restaurés vers des versions antérieures d'ONTAP. Seules les mises à niveau vers des versions ultérieures sont autorisées.

Cycle de vie de l'espace de travail et configuration d'accès

- **Aucune suppression réversible ni restauration pour les espaces de travail**

La suppression d'un espace de travail est définitive. Il n'y a aucune option pour restaurer les espaces de travail supprimés.

- **Aucune prise en charge de la configuration OIDC lors de la configuration initiale du cluster**

La configuration OIDC/OAuth doit être effectuée après la création du cluster à l'aide de ONTAP System Manager.

Informations connexes

- ["Nouveautés dans AI Data Engine"](#)

Commencer

Découvrez votre système AI Data Engine

Découvrez AI Data Engine

La NetApp AI Data Engine (AIDE) est une plateforme d'entreprise conçue pour accélérer et simplifier le traitement des données, la gestion et la gouvernance des données pilotées par l'IA. AIDE peut aider à transformer de grandes quantités de données non structurées en ensembles de données structurés et prêts pour l'IA. Elle est conçue pour répondre aux exigences des charges de travail modernes de machine learning (ML) et d'intelligence artificielle générative (GenAI), prenant en charge à la fois les opérations informatiques traditionnelles et les nouveaux rôles axés sur l'IA.

AIDE relève les défis de l'IA

AIDE est conçu pour aider les organisations à gérer les données pour les charges de travail d'IA et offre les capacités clés suivantes :

- **Gestion centralisée des métadonnées** : AIDE collecte et catalogue les métadonnées des volumes ONTAP, permettant ainsi de rechercher, de classer et d'appliquer des politiques de gouvernance aux ensembles de données.
- **Traitement automatisé des données** : AIDE prend en charge la création de pipelines de données pour les charges de travail d'IA et de ML, y compris la possibilité de générer des vector embeddings pour la recherche sémantique (avec les licences appropriées).
- **Isolation des données et contrôle d'accès** : AIDE applique des contrôles d'accès et une isolation de base des données pour plusieurs équipes ou projets.
- **Intégration avec les outils NetApp** : AIDE fonctionne avec ONTAP System Manager pour l'administration du stockage et fournit une interface dédiée (AI Data Engine Console) permettant aux ingénieurs et scientifiques des données de gérer les collectes de données et les flux de travail.

Caractéristiques de conception de haut niveau

Les caractéristiques de conception suivantes définissent la manière dont AI Data Engine est construit pour répondre aux besoins des charges de travail d'IA :

- **Services basés sur des microservices** : Utilise Kubernetes pour orchestrer des services modulaires et résilients pour le catalogage des métadonnées, la recherche vectorielle et la gestion de l'infrastructure.
- **Sécurité de niveau entreprise** : Met en œuvre le chiffrement, le contrôle d'accès basé sur les rôles (RBAC) et l'audit de toutes les données et métadonnées.
- **Accès aux données multi-protocoles** : Prend en charge NFS et SMB pour une ingestion et une récupération flexibles des données.
- **Pipelines de données automatisées** : Suit les modifications de données, crée des embeddings et gère les bases de données vectorielles pour les applications d'IA.

Comment les données circulent dans AIDE

Comprendre comment les données circulent dans AIDE aide à illustrer la valeur de la plateforme pour les

équipes d'AI/ML :

1. **Ingestion des données** : Les fichiers sont stockés dans des volumes ONTAP à l'aide des protocoles standard (NFS et SMB). Les données peuvent résider sur le stockage local AIDE (le cluster AFX au sein de votre déploiement AIDE) ou sur des clusters ONTAP distants. Les données provenant des clusters distants sont synchronisées avec le cluster AFX local à l'aide d'ONTAP SnapMirror, de sorte que toutes les données traitées par AIDE sont finalement stockées et accessibles localement.



Les compartiments S3 ne sont pas pris en charge comme sources de données pour les espaces de travail ou la collecte de données.

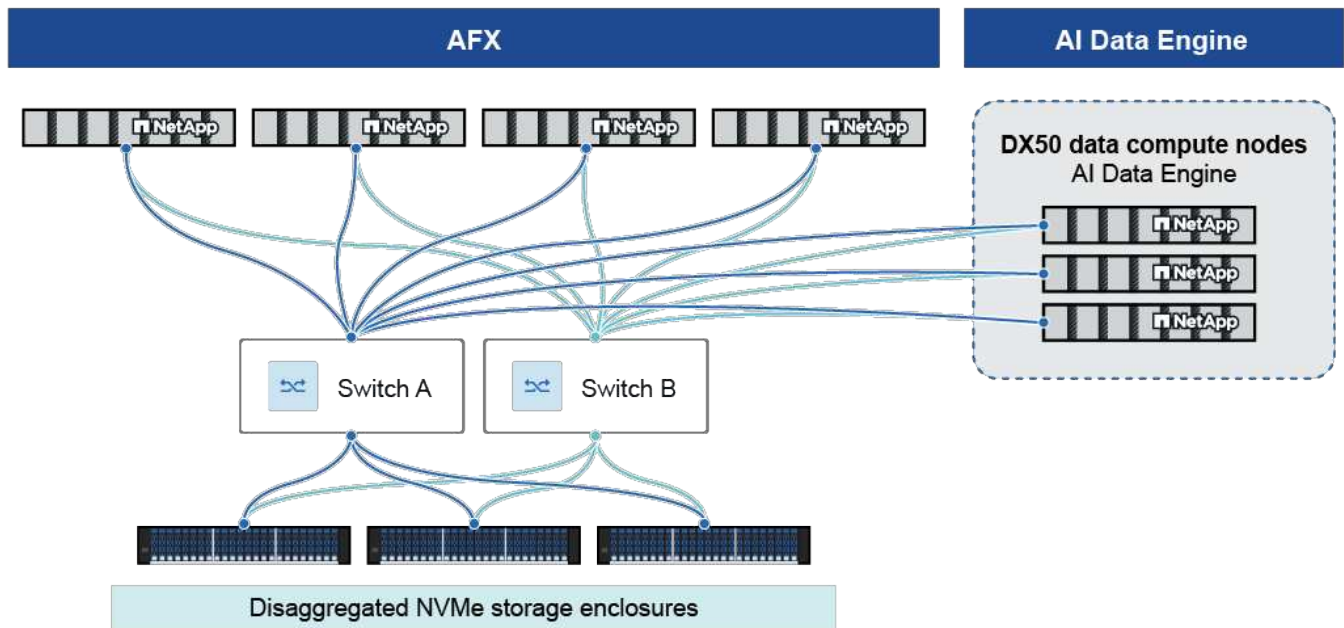
1. **Création d'espaces de travail** : Les administrateurs de stockage définissent des espaces de travail dans ONTAP System Manager, regroupant les volumes ONTAP associés pour des projets, des équipes ou des flux de travail spécifiques. Les autorisations d'accès et les politiques de gouvernance sont attribuées au niveau de l'espace de travail.
2. **Extraction des métadonnées** : AIDE analyse automatiquement les fichiers et les objets dans les espaces de travail, extrait les métadonnées (type de fichier, taille, horodatages, attributs personnalisés) et les stocke dans un catalogue centralisé. Cela se produit en continu à mesure que les données changent.
3. **Classification et gouvernance** : Les classificateurs analysent les données à la recherche d'informations sensibles (PII, données financières) ou de types de documents (juridiques, RH). Les politiques de Data Guardrails appliquent automatiquement la rédaction ou les restrictions d'accès.
4. **Création de collecte de données** : Les ingénieurs et les scientifiques des données utilisent l'AI Data Engine Console pour interroger le catalogue de métadonnées, filtrer les résultats et assembler des collectes de données organisées pour des tâches d'IA spécifiques.
5. **Vectorisation** : Pour les collections nécessitant une recherche sémantique, AIDE génère des embeddings à l'aide de modèles d'IA sélectionnés. Les vecteurs sont stockés dans la base de données de vecteurs pour une récupération haute performance.
6. **Consommation d'AI/ML** : Les applications accèdent aux données par de multiples voies :
 - Accès direct aux fichiers/objets à l'aide de NFS ou SMB
 - Requêtes de recherche sémantique sur la base de données vectorielle
 - Points d'accès RAG qui combinent la récupération de données avec l'intégration de modèles GenAI
 - Accès à l'API REST pour les flux de travail programmatiques

Ce flux de travail automatisé et basé sur des politiques réduit le temps et les efforts manuels nécessaires à la préparation des données pour l'IA, permettant aux équipes de se concentrer sur le développement de modèles et les insights plutôt que sur la manipulation des données.

Architecture de l'AI Data Engine

AIDE repose sur une architecture évolutive et tolérante aux pannes qui sépare le stockage et le calcul, permettant des performances et une flexibilité élevées pour les charges de travail AI.

Composants physiques



nœuds de contrôleur AFX

Les nœuds de contrôle AFX exécutent une version spécialisée du logiciel ONTAP conçue pour répondre aux exigences de l'environnement AFX. Les clients accèdent aux nœuds via plusieurs protocoles, notamment NFS et SMB. Chaque nœud dispose d'une vue complète du stockage, à laquelle il peut accéder en fonction des requêtes des clients. Les nœuds sont dotés d'une mémoire non volatile permettant de conserver les informations d'état critiques et intègrent des améliorations supplémentaires spécifiques aux charges de travail cibles.

Au moins quatre nœuds de contrôleur AFX sont nécessaires pour les déploiements AIDE afin de garantir une haute disponibilité et des performances élevées.

Nœuds de calcul de données

Les nœuds de calcul de données (DCN) sont des serveurs Linux dotés de ressources CPU, RAM et GPU élevées, dédiés aux tâches de traitement des données d'IA. Ils hébergent des services spécifiques à l'IA tels que le catalogue des métadonnées, la recherche vectorielle et les pipelines d'embedding.

Exactement *trois* DCN sont nécessaires pour les déploiements AIDE.

Commutateurs de cluster/stockage

Des commutateurs redondants à haut débit (100GbE ou plus) connectent ONTAP et les DCN pour un transfert de données à faible latence et une haute disponibilité.

Étagères de stockage

Les baies NVMe-oF avec des SSD haute densité offrent une latence ultra-faible et une redondance, prenant en charge le stockage à l'échelle du pétaoctet.

Réseautique

Tous les nœuds DCN et les nœuds de stockage ONTAP sont connectés par des commutateurs de cluster redondants à haut débit (minimum 100GbE). Cette architecture sépare les ressources de calcul et de stockage, permettant à chacune d'évoluer indépendamment et en optimisant à la fois les performances et

l'utilisation des ressources.

La mise en réseau entre les DCN et les nœuds ONTAP est isolée à l'aide de VLAN dédiés et d'IPspaces sur les commutateurs du cluster. Cela garantit que toutes les communications, telles que l'accès aux données, les API de gestion et le trafic des services internes, restent sécurisées, efficaces et n'interfèrent pas avec les autres opérations réseau.

Fonctionnalités principales de l'AI Data Engine

Les principales fonctionnalités de l'AI Data Engine (AIDE) fonctionnent ensemble pour automatiser, sécuriser et accélérer le cycle de vie des données d'IA. Chaque fonctionnalité est implémentée sous forme d'un ensemble de microservices s'exécutant sur des DCN, intégrés au stockage ONTAP et exposés via des API REST et des interfaces de gestion.

Metadata Engine

Le Metadata Engine génère automatiquement une vue structurée, à jour et interactive de votre patrimoine de données NetApp.

Licence et accès

Le Metadata Engine est inclus avec la licence de base ONTAP One et est disponible lors de l'installation d'AIDE.

Vous pouvez y accéder via ONTAP System Manager.

Capacités

- Catalogue les métadonnées de toutes les sources de données, y compris les volumes stockés localement sur le cluster AFX et ceux synchronisés à partir de clusters ONTAP distants.
- Extrait automatiquement les métadonnées et alimente le catalogue au fur et à mesure que les données sont ingérées ou modifiées.
- Fournit un accès à l'API REST pour interroger les métadonnées, permettant aux spécialistes des données et aux administrateurs de stockage de découvrir, classer et comprendre les données.
- Décharge les requêtes de métadonnées du chemin d'accès aux données, réduisant la charge de trafic NFS sur les systèmes de stockage.
- Prend en charge les enregistrements de métadonnées volumineux avec des fonctionnalités d'indexation et de recherche.
- S'intègre aux abstractions d'espace de travail et de collecte de données pour appliquer le contrôle d'accès et la gouvernance.

Synchronisation des données

La synchronisation des données est un service automatisé en arrière-plan qui garantit que le catalogue de métadonnées et les collections de données restent à jour et cohérents avec les sources de données sous-jacentes, même lorsque les données sources changent.

Licence et accès

La fonctionnalité Data Sync est incluse avec la licence de base ONTAP One et est disponible dès l'installation d'AIDE.

Capacités

- Synchronise les données des clusters ONTAP distants ou locaux à l'aide de la réplication SnapMirror basée sur des politiques. Les données des clusters distants sont copiées vers le cluster AFX local pour le

traitement AIDE.

- Mise à jour progressive basée sur les modifications détectées, en ne propageant que les données modifiées.
- Assure une mobilité des données et une synchronisation sécurisées et progressives à travers l'ensemble du parc de données.
- Planifie et surveille les intervalles de synchronisation avec des taux d'actualisation configurables par espace de travail.
- S'intègre aux flux de travail de création de workspace pour extraire et mettre à jour les métadonnées à mesure que de nouvelles sources de données sont ajoutées.

Data Guardrails

Le service Data Guardrails assure une gouvernance et une protection continues et automatisées des données sensibles tout au long du cycle de vie de l'IA.

Licence et accès

La fonctionnalité Data Guardrails n'est pas incluse avec la licence de base ONTAP One et nécessite une licence AIDE distincte.

Vous pouvez accéder à la fonctionnalité Data Guardrails via l'AI Data Engine Console.

Capacités

- Analyse, classe et catégorise les données en continu.
- Identifie les données sensibles et les risques grâce à des classificateurs intégrés et personnalisables pour des tâches telles que la détection de PII.
- Automatise la gestion des données sensibles grâce à la rédaction, au masquage et aux restrictions d'accès pilotés par des règles.
- Fait respecter les normes de l'entreprise et les réglementations grâce à des politiques de Data Guardrails appliquées aux espaces de travail.
- Restreint l'accès aux fichiers ou volumes sensibles selon la configuration, avec journalisation des audits et rapports de conformité.
- S'intègre à la gestion des espaces de travail et de la collecte de données pour appliquer des Data Guardrails de manière cohérente aux workflows de données d'IA.

Data Curator

Le service Data Curator permet une découverte, une recherche, une vectorisation et une récupération rapides des données pour les applications d'IA et de GenAI.

Licence et accès

La fonctionnalité Data Curator n'est pas incluse avec la licence de base ONTAP One et nécessite une licence AIDE distincte.

Vous pouvez accéder à Data Curator via la AI Data Engine Console.

Capacités

- Recherche les données pertinentes dans le stockage à l'aide du catalogue de métadonnées centralisé.
- Fournit aux data scientists des outils pour créer des collections de données organisées.
- Génère automatiquement des plongements vectoriels au niveau de la couche de stockage.

- Fournit un point de terminaison de récupération sécurisé pour les applications d'IA, prenant en charge la recherche sémantique vectorielle et le réordonnement.
- S'intègre aux outils et technologies d'IA, y compris les pipelines de Retrieval-Augmented Generation (RAG) et les frameworks d'IA agentiques.
- Fournit des API REST pour l'accès programmatique aux collections de données, la recherche vectorielle et les points de terminaison de récupération.

Sécurité et multi-location

La plateforme applique à la fois le contrôle d'accès basé sur les rôles (RBAC) et les listes de contrôle d'accès (ACL) au niveau des ressources. Toutes les actions des API et des utilisateurs sont auditées, et toutes les données sont chiffrées au repos et en transit. Chaque locataire est isolé pour les données et les métadonnées.

Informations connexes

- ["Installer une licence AIDE"](#)
- ["Démarrage rapide Data-to-RAG"](#)

Composantes et responsabilités d'AIDE par rôle

Composants de l'AI Data Engine et interactions basées sur les rôles

AI Data Engine (AIDE) se compose de nombreux composants principaux qui travaillent ensemble pour fournir une plateforme complète de gestion et de traitement des données pour les charges de travail d'IA. Ces composants incluent des espaces de travail, des collectes de données, des bases de données vectorielles, des garde-fous, des catalogues de métadonnées, des points de terminaison de récupération et des classificateurs. Chaque composant joue un rôle spécifique pour permettre une découverte, une curation, une gouvernance et une intégration efficaces des données avec les applications d'IA/ML.

Chaque utilisateur d'AIDE interagit avec les composants d'AIDE différemment selon son rôle.

Rôles des utilisateurs axés sur le stockage et les données

AIDE introduit de nouveaux rôles d'utilisateur tout en prenant toujours en charge les rôles d'administration système ONTAP traditionnels :

Utilisateurs de stockage

- **Administrateur de stockage** : Gère la configuration du cluster AFX et AIDE, le réseau, l'approvisionnement en stockage et l'accès des utilisateurs.

Utilisateurs de données

- **Ingénieur de données** : Conçoit et optimise les pipelines d'AI/ML, gère les collectes de données et intègre les modèles d'AI.
- **Scientifique des données** : Découvre, organise et analyse des ensembles de données, crée des collections de données et exploite les points de terminaison de récupération pour les applications GenAI.

Rôle (nom RBAC)	Description
Administrateur de stockage (<code>admin</code>)	Gère la configuration des clusters AFX et AIDE, le réseau, le provisionnement du stockage et l'accès utilisateur. Attribue des rôles RBAC aux utilisateurs qui déterminent le niveau de contrôle d'accès aux interfaces et fonctionnalités d'AIDE. Ce rôle d'administrateur dispose d'un accès complet à la gestion via ONTAP System Manager et AI Data Engine Console.
Ingénieur de données (<code>data-engineer</code>)	Conçoit et optimise des pipelines d'IA/ML, gère la collecte de données et intègre des modèles d'IA. Ce rôle a accès à la AI Data Engine Console pour les workflows d'ingénierie des données.
scientifique des données (<code>data-scientist</code>)	Découvre, organise et analyse des ensembles de données, crée des collections de données et exploite les points de récupération pour les applications GenAI. Ce rôle a accès à l'AI Data Engine Console pour les workflows de science des données.

Composants du système AIDE

Chaque utilisateur d'AIDE (administrateurs de stockage, ingénieurs de données et scientifiques des données) interagit avec les composants d'AIDE en fonction de son rôle.

Espaces de travail

Un espace de travail est un segment logique de données au sein du cluster, regroupant les volumes pour un projet, une équipe ou un flux de travail spécifique. Les espaces de travail définissent la portée de la visibilité, de l'accès et de la gouvernance des données dans AIDE.

Catalogue de métadonnées

Une base de données centralisée et évolutive stockant les enregistrements de métadonnées pour tous les fichiers et objets du cluster local, y compris les données synchronisées depuis des clusters ONTAP distants à l'aide d'ONTAP SnapMirror ou de l'appairage de clusters. Elle permet une recherche et un filtrage riches et interactifs.

Classificateurs

Les classificateurs sont des outils (intégrés ou personnalisés) qui analysent et étiquettent les fichiers pour des types spécifiques de données sensibles (par exemple, PII, financières, domaine de la santé) ou catégorisent les documents par type (par exemple, juridiques, RH, ventes).

Collecte de données

Une collecte de données est un groupe organisé de fichiers ou d'objets liés provenant d'un espace de travail, défini par une requête spécifiée par l'utilisateur pour une utilisation dans les workflows GenAI. Le contenu des fichiers dans la collecte de données, après publication, est disponible pour la recherche sémantique par les API des applications GenAI.

base de données vectorielle

La base de données vectorielle stocke les embeddings générés à partir de collections de données, permettant une recherche et une récupération sémantiques hautes performances pour les applications d'IA et de GenAI.

Garde-fous

Les garde-fous sont des mécanismes fondés sur des politiques qui appliquent la gouvernance des données, la classification et la protection (telles que la rédaction ou les restrictions de contrôle d'accès) tout au long du cycle de vie des données.

Point de terminaison de récupération (RAG endpoint)

Un point de terminaison de récupération (parfois appelé point de terminaison de génération augmentée par récupération ou « RAG ») est une API sécurisée qui permet aux applications d'IA et de GenAI d'accéder aux données, au contexte ou aux embeddings pertinents à partir de collections organisées et de la base de données vectorielle.

Les points d'accès RAG sont conçus pour prendre en charge les flux de travail d'IA avancés, tels que la recherche sémantique et les réponses contextuelles dans les modèles d'IA génératifs. En connectant vos applications d'IA à un point d'accès de récupération, vous pouvez améliorer la précision et la pertinence des modèles en fournissant un accès en temps réel à des ensembles de données organisés et prêts pour l'IA, gérés par AIDE.

Informations connexes

- ["Comment les administrateurs de stockage AIDE travaillent avec les composants AIDE"](#)
- ["Comment les ingénieurs de données AIDE travaillent avec les composants AIDE"](#)
- ["Comment les data scientists d'AIDE travaillent avec les composants d'AIDE"](#)

Interfaces AI Data Engine

AI Data Engine (AIDE) propose trois interfaces principales pour l'interaction utilisateur et l'automatisation. Chaque rôle, comme les administrateurs de stockage, les ingénieurs de données et les data scientists, utilise ces interfaces en fonction de ses tâches et responsabilités spécifiques.

ONTAP System Manager

ONTAP System Manager est une interface web conçue pour les administrateurs de stockage. Elle propose des flux de travail pour la configuration des clusters, la gestion des espaces de travail, la surveillance des DCN et l'application de politiques de garde-fous.

AI Data Engine Console

AI Data Engine Console est une interface dédiée aux ingénieurs des données et aux data scientists. Elle permet aux utilisateurs d'explorer des sources de données, de créer et de gérer des collections de données, de configurer des pipelines de données, d'appliquer des classificateurs et d'interagir avec les garde-fous et les fonctionnalités de recherche vectorielle. La console fournit des outils avancés pour la découverte, la curation et l'intégration des données aux workflows d'IA/ML.

API REST

AIDE expose l'API REST ONTAP pour l'automatisation, l'intégration et l'accès programmatique. L'API prend en charge la configuration des clusters, la gestion des espaces de travail et des collections, les requêtes de métadonnées, la recherche vectorielle et les points de terminaison de récupération.

Découvrez comment les administrateurs de stockage AI Data Engine travaillent avec les composants AIDE

En tant qu'administrateur de stockage, vous gérez l'infrastructure AIDE via ONTAP et la AI Data Engine Console, en provisionnant des espaces de travail, en appliquant des politiques de garde-fou et en surveillant l'état du système. Votre rôle consiste à garantir un stockage des données fiable, sécurisé et conforme pour les charges de travail d'IA.

Accès au composant d'administrateur de stockage

Composant	Niveau d'accès	Flux de travail de l'administrateur de stockage
ONTAP System Manager	Gérer (créer, modifier, supprimer)	Vous utilisez ONTAP System Manager comme interface principale pour l'administration du cluster, le provisionnement de l'espace de travail, la gestion des politiques de garde-fou et la surveillance de l'état du système.
AI Data Engine Console	Gérer (créer, modifier, supprimer)	Vous utilisez l'AI Data Engine Console pour surveiller les espaces de travail, consulter l'état des collections et superviser l'activité du système dans l'environnement AIDE.
API REST ONTAP	Gérer (créer, modifier, supprimer)	Vous utilisez l'API REST pour automatiser les tâches d'infrastructure, gérer les espaces de travail et les politiques de protection par programmation, et intégrer l'administration d'AIDE avec des outils et des flux de travail externes.
Espaces de travail	Gérer (créer, modifier, supprimer)	Vous créez et gérez des espaces de travail à l'aide de ONTAP System Manager. Vous sélectionnez les sources de données incluses, attribuez des autorisations aux data engineers et aux data scientists, et appliquez des politiques de garde-fou pour garantir la gouvernance et la conformité. Vous surveillez également l'état et l'accès aux espaces de travail.
Collecte de données	Afficher (lecture seule)	Vous consultez l'état et l'intégrité des collections de données dans chaque espace de travail à l'aide de System Manager. Vous vous assurez que les sources de données sous-jacentes sont disponibles et protégées, mais vous ne créez ni ne modifiez les collections.
Garde-fous	Gérer (créer, modifier, supprimer)	Vous définissez et associez des politiques de garde-fou aux espaces de travail à l'aide de System Manager. Vous surveillez l'état des garde-fous et les rapports de conformité. Vous veillez à ce que les politiques soient appliquées et mises à jour selon les besoins.

Composant	Niveau d'accès	Flux de travail de l'administrateur de stockage
Catalogue de métadonnées	Surveiller (afficher la santé, l'état, l'activité)	Vous veillez à ce que le catalogue de métadonnées soit rempli et à jour. Vous surveillez l'état du catalogue et soutenez le contrôle d'accès.
base de données vectorielle	Provisionnement/Surveillance (déployer, configurer, afficher l'état)	Vous provisionnez et supervisez l'infrastructure de base de données vectorielle, en veillant à ce que les nœuds de calcul de données disposent de ressources GPU et des licences appropriées. Vous assurez le support de l'environnement, mais ne gérez pas directement les embeddings ni les requêtes.
Classificateurs	Gérer (créer, modifier, supprimer)	Vous créez, configurez et gérez les classificateurs et leurs catégories. Vous appliquez les classificateurs aux espaces de travail et surveillez leur efficacité.

Découvrez comment les ingénieurs des données et les scientifiques des données d'AI Data Engine travaillent avec les composants d'AIDE

En tant qu'ingénieur des données ou data scientist, vous utilisez la AI Data Engine Console pour explorer les espaces de travail auxquels vous avez obtenu l'accès, créer et gérer des data collections, effectuer des recherches sémantiques et intégrer des retrieval endpoints dans les workflows d'AI/ML.

Les ingénieurs de données transforment les données brutes en jeux de données prêts pour l'IA en créant des collections, en configurant des pipelines d'intégration et en contrôlant quels utilisateurs peuvent accéder aux collections publiées. Les data scientists se concentrent sur l'exploitation de jeux de données organisés pour l'analyse, l'entraînement de modèles et les applications GenAI, sans gérer le contrôle d'accès ni l'infrastructure.

Accès des utilisateurs aux composants

Composant	Niveau d'accès	Flux de travail de l'ingénieur de données	Flux de travail du data scientist
AI Data Engine Console	Gérer (créer, modifier, supprimer)	La AI Data Engine Console est votre interface principale pour les tâches quotidiennes, notamment la découverte des données, la gestion de la collecte, la configuration des pipelines et la publication des points de terminaison RAG ou de récupération, pour les espaces de travail auxquels vous êtes autorisé à accéder.	La AI Data Engine Console est votre interface principale pour l'exploration des données, l'affinage et le versionnage des collections dans les espaces de travail auxquels vous avez accès, ainsi que pour la connexion des ensembles de données organisés et des points de terminaison de récupération aux flux de travail d'analyse, de modélisation et de GenAI.

Composant	Niveau d'accès	Flux de travail de l'ingénieur de données	Flux de travail du data scientist
API REST ONTAP	Gérer (créer, modifier, supprimer)	Vous utilisez l'API REST pour automatiser les opérations du cycle de vie des collections, déclencher et surveiller les pipelines d'intégration d'embedding, et intégrer par programmation les flux de données avec des outils externes.	Vous utilisez l'API REST pour accéder par programmation aux collections de données, exécuter des requêtes de recherche vectorielle et intégrer des points de terminaison de récupération dans des applications AI/ML et des frameworks agentiques.
Espaces de travail	Afficher/utiliser (lecture seule)	Vous explorez vos espaces de travail assignés pour identifier et comprendre les sources de données disponibles avant de créer des collections.	Vous parcourez vos espaces de travail attribués pour localiser les fichiers et objets pertinents pour des tâches de recherche ou de modélisation spécifiques.
Collecte de données	Gérer (créer, modifier, supprimer)	Vous créez des collections de données en sélectionnant et en filtrant les données sources à l'aide de balises, de classifications et d'autres attributs, et vous gérez l'intégralité du cycle de vie de la collection, de la création et du versionnage jusqu'à la publication en tant que points de terminaison RAG pour une utilisation par l'IA. Vous gérez également quels data scientists et autres utilisateurs peuvent accéder à chaque collection.	Vous créez, sélectionnez, annotez, versionnez et affinez des collections de données dans les espaces de travail auxquels vous avez accès. Vous utilisez ces collections comme base pour la recherche sémantique et les workflows GenAI.
Catalogue de métadonnées	Requête/utilisation (consommation pour les workflows)	Vous utilisez le catalogue de métadonnées pour évaluer et sélectionner les sources de données à ingérer, en exécutant des requêtes pour localiser les fichiers pertinents et confirmer qu'ils répondent aux exigences des collections que vous constituez dans vos espaces de travail attribués.	Vous recherchez et filtrez les métadonnées dans les espaces de travail auxquels vous avez accès pour localiser les fichiers et objets nécessaires à l'analyse ou à l'entraînement du modèle, en vous appuyant sur la structure du catalogue qui a été construite et maintenue par les ingénieurs de données.

Composant	Niveau d'accès	Flux de travail de l'ingénieur de données	Flux de travail du data scientist
base de données vectorielle	- Gérer les embeddings/recherche (data engineer) - Utiliser/rechercher (data scientist)	Vous déclenchez les pipelines d'intégration, surveillez l'état de la vectorisation, configurez les paramètres de segmentation et d'intégration, et exposez des points de terminaison de récupération basés sur la recherche vectorielle. Les applications et agents interrogent ensuite ces points de terminaison via l'API pour la recherche sémantique et les workflows RAG.	Vous exécutez des requêtes de recherche sémantique sur des embeddings générés par des pipelines gérés par des data engineers et intégrez les résultats de récupération dans des workflows GenAI ou RAG pour des réponses de modèles tenant compte du contexte. Vous ne configurez pas le découpage, les embeddings ou les paramètres du pipeline.
Classificateurs	Utiliser (consommer des données classifiées)	Vous utilisez les résultats de la classification pour annoter et étiqueter les données sources lors de la préparation de la collecte, garantissant que le contenu entrant dans vos pipelines est correctement étiqueté pour les workflows d'IA en aval.	Vous consommez des données pré-classées pour garantir que seul un contenu conforme et pertinent est utilisé dans votre analyse et votre modélisation.

Démarrage rapide pour AI Data Engine

Pour mettre en service votre AI Data Engine system, vous devez installer vos composants matériels, configurer votre cluster, configurer l'accès aux données depuis vos hôtes vers le système de stockage et provisionner votre stockage.

1

AIDE sera-t-il installé avec un cluster AFX nouveau ou existant ?

Vous devez décider si AIDE et AFX seront installés en même temps ou si AIDE sera intégré à un cluster AFX existant.

2

Installez et configurez votre matériel

"[Installer et configurer](#)" vos nœuds de calcul du cluster AIDE. Selon l'environnement d'installation, assurez-vous également de "[installer](#)" le matériel AFX.

3

Configurez votre cluster

Utilisez ONTAP System Manager pour vous guider tout au long d'un processus rapide et facile pour "[Configurer AIDE avec un cluster AFX](#)".

4

Configurer les espaces de travail et l'accès aux données

"Configurez un espace de travail et des utilisateurs pour cet espace de travail qui peuvent accéder aux données AI Data Engine".

Et ensuite ?

Vous pouvez désormais utiliser ONTAP System Manager pour gérer votre AI Data Engine et permettre à vos data engineers et data scientists de démarrer avec leurs espaces de travail et leurs configurations.

Installer AIDE

Exigences pour l'installation de AI Data Engine

Veuillez consulter la configuration requise pour l'installation du AI Data Engine. AIDE nécessite un système de stockage AFX, un minimum de trois Data Compute Node, des commutateurs réseau et des câbles.

Exigences matérielles

Le AI Data Engine nécessite un système de stockage AFX et au minimum trois Data Compute Node. Le système AFX fournit l'infrastructure de stockage, tandis que les Data Compute Node hébergent les composants logiciels AIDE qui permettent la gestion des données, la curation et les capacités d'IA.

- **Système de stockage AFX** : comprend un contrôleur AFX, un tiroir disque et un commutateur réseau. Le système de stockage AFX est requis pour le déploiement d'AIDE.
- **Nœuds de calcul de données** : Un minimum de trois nœuds de calcul de données est requis. Les nœuds de calcul de données sont des nœuds matériels fournis par NetApp qui hébergent le logiciel AIDE, y compris le Metadata Engine, Data Sync, Data Curator et Data Guardrails.
 - Chaque Data Compute Node dispose des emplacements d'E/S 4 et 5 disponibles pour la connectivité. L'emplacement 3 est réservé au GPU. Les emplacements 1 et 2 ne sont ni utilisés ni accessibles.
 - Les ports e4a et e5a sont utilisés pour les connexions au cluster.
 - Les ports e4b et e5b sont utilisés pour les connexions réseau hôte.

Exigences relatives aux commutateurs réseau

AI Data Engine nécessite des commutateurs réseau pour permettre la connectivité réseau de l'hôte et la communication inter-nœuds pour les Data Compute Node.

- Commutateurs clients pour la connectivité réseau hôte (Cisco Nexus 9332D-GX2B ou Cisco Nexus 9364D-GX2A)
- Commutateurs de cluster pour la communication inter-nœuds (Cisco Nexus 9332D-GX2B ou Cisco Nexus 9364D-GX2A)
- Commutateurs de gestion optionnels pour la gestion du réseau

Exigences de câblage

Les câbles suivants sont nécessaires pour connecter les Data Compute Node aux commutateurs réseau et au réseau de gestion.

- Câbles de dérivation 400-GbE-to-100-GbE (4x100GbE) pour connecter les nœuds aux commutateurs clients et de cluster

Prise en charge multi-cluster

Une fois AIDE déployé avec AFX, il peut se connecter et gérer les données d'autres clusters ONTAP 9.18.1 et versions ultérieures en utilisant SnapMirror et le peering de clusters.

Installez le système de stockage AFX pour AI Data Engine

Installez le système de stockage AFX en première étape du déploiement de l'AI Data Engine. Le système de stockage AFX constitue la base de l'infrastructure de stockage et est requis avant l'installation des Data Compute Node.

Suivez les ["Documentation d'installation AFX 1K"](#) pour installer le système de stockage AFX.

Et ensuite

Après avoir terminé l'installation du système de stockage AFX, ["installer les Data Compute Node"](#).

Installez vos nœuds de calcul de données

Flux de travail d'installation et de configuration des nœuds de calcul de données pour AI Data Engine

Pour installer et configurer vos nœuds de calcul de données (DCN), vous examinez les exigences matérielles, préparez votre site, installez et câblez les composants matériels, mettez le système sous tension et configurez votre cluster ONTAP.

1

["Examinez les exigences d'installation du matériel"](#)

Assurez-vous de disposer d'un système de stockage AFX 1K installé, puis consultez la configuration matérielle requise pour l'installation des nœuds de calcul de données pour AIDE. Pour plus d'informations sur l'installation du système de stockage AFX 1K, consultez ["Documentation d'installation du système de stockage AFX 1K"](#).

2

["Préparez-vous à installer vos nœuds de calcul de données"](#)

Pour préparer l'installation de vos nœuds de calcul de données, vous devez aménager le site, vérifier les exigences environnementales et électriques et vous assurer de disposer de suffisamment d'espace dans les armoires. Ensuite, déballiez le matériel, comparez son contenu au bon de livraison et enregistrez le matériel pour accéder aux avantages du support. Un minimum de trois nœuds de calcul de données est requis pour déployer l'AI Data Engine.

3

["Installez le matériel pour vos nœuds de calcul de données"](#)

Installez les kits de rails pour vos nœuds de calcul de données. Fixez solidement vos nœuds de calcul de données à l'intérieur de l'armoire. Enfin, attachez les dispositifs de gestion des câbles à l'arrière du système pour un acheminement organisé des câbles.

4

["Câblez vos nœuds de calcul de données"](#)

Pour câbler le matériel, connectez d'abord les nœuds de calcul de données à votre réseau de données, puis connectez les nœuds de calcul de données aux commutateurs du cluster.

"Mettez sous tension vos nœuds de calcul de données"

Après avoir installé le matériel du rack et câblé vos nœuds de calcul de données, vous devez mettre sous tension vos DCN et vos nœuds de contrôle pour le système de stockage AFX si ceux-ci ne sont pas déjà sous tension.

Configuration requise pour les nœuds de calcul de données pour AI Data Engine

Examinez l'équipement nécessaire et les précautions de levage pour vos nœuds de calcul de données pour AI Data Engine.

Prérequis

Avant d'installer les nœuds de calcul de données pour AIDE, assurez-vous de disposer des éléments suivants :

- Un système de stockage AFX 1K.



Pour plus d'informations sur l'installation du système de stockage AFX 1K, consultez ["Documentation d'installation du système de stockage AFX 1K"](#).

Matériel nécessaire à l'installation

Pour installer les nœuds de calcul de données pour AIDE, vous avez besoin des équipements et outils suivants.

- Accès à un navigateur web pour configurer vos nœuds de calcul de données
- Bracelet antistatique (ESD)
- Lampe de poche
- Ordinateur portable ou console avec une connexion USB/série
- Tournevis Phillips n° 2

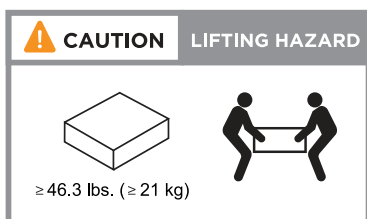
Précautions de levage

Les nœuds de calcul de données sont lourds. Faites preuve de prudence lors du levage et du déplacement de ces éléments.

Poids des nœuds de calcul de données

Prenez les précautions nécessaires lors du déplacement ou du levage de votre nœud de calcul de données.

Un nœud de calcul de données peut peser jusqu'à 46,3 lbs (21 kg). Pour soulever le nœud de calcul de données, utilisez deux personnes ou un élévateur hydraulique.



Informations connexes

- ["Informations de sécurité et avis réglementaires"](#)

Et ensuite ?

Après avoir examiné les exigences matérielles, ["Préparez-vous à installer vos nœuds de calcul de données"](#).

Préparez-vous à installer vos nœuds de calcul de données pour AI Data Engine

Préparez-vous à installer vos nœuds de calcul de données pour AI Data Engine en préparant le site, en déballant les boîtes et en comparant le contenu des boîtes au bon de livraison, puis en enregistrant le système pour accéder aux avantages du support.

Étape 1: Préparer le site

Pour installer vos nœuds de calcul de données, assurez-vous que le site et l'armoire ou le rack que vous prévoyez d'utiliser répondent aux spécifications de votre configuration.

Étapes

1. Utilisez ["NetApp Hardware Universe"](#) pour confirmer que votre site répond aux exigences environnementales et électriques de vos nœuds de calcul de données.
2. Assurez-vous de disposer d'un espace suffisant dans l'armoire pour vos nœuds de calcul de données au sein de l'installation du système de stockage AFX 1K existant.
 - 1U pour chaque nœud de calcul de données
 - 2U pour chaque nœud de contrôleur AFX 1K
 - 2U pour chaque shelf NX224
 - 1U ou 2U par commutateur, selon le modèle de commutateur.

Étape 2 : Déballer les cartons

Après avoir vérifié que le site et l'armoire ou le rack que vous prévoyez d'utiliser pour vos nœuds de calcul de données répondent aux spécifications requises, déballer tous les cartons et comparez leur contenu aux éléments figurant sur le bon de livraison.

Étapes

1. Ouvrez soigneusement toutes les boîtes et disposez le contenu de manière organisée.
2. Comparez les articles déballés avec le bon de livraison et notez toute anomalie.

Vous pouvez obtenir votre liste de colisage en scannant le code QR sur le côté du carton d'expédition.

Les éléments suivants sont quelques-uns des contenus que vous pourriez voir dans les boîtes.

Matériel	Câbles	
• panneau • Kits de rails avec instructions • Nœud(s) de calcul de données	• cordons d'alimentation	

Étape 3 : Enregistrez vos nœuds de calcul de données

Après avoir vérifié que votre site répond aux exigences des spécifications de votre nœud de calcul de données et que vous avez confirmé que vous disposez de toutes les pièces commandées, vous devez enregistrer votre système.

Étapes

- 1. Repérez les numéros de série de vos nœuds de calcul de données.

Vous trouverez les numéros de série aux emplacements suivants :

- Sur le bordereau d’expédition
- Dans votre courriel de confirmation
- Sur chaque nœud de calcul de données, ou sur certains systèmes, sur le module de gestion système de chaque nœud de calcul de données.



- 2. Accédez à "[Site de support NetApp](#)".
- 3. Déterminez si vous devez enregistrer votre système de stockage :

Si vous êtes un...	Suivez ces étapes...
Client NetApp existant	a. Sign in avec votre nom d'utilisateur et votre mot de passe. b. Sélectionnez Systems > My Systems. c. Confirmez que le nouveau numéro de série est indiqué. d. Si le numéro de série n'est pas indiqué, suivez les instructions destinées aux nouveaux NetApp clients.
Nouveau NetApp client	a. Cliquez sur Register Now pour créer un compte. b. Sélectionnez Systems > Register Systems. c. Saisissez le numéro de série du système de stockage et les détails demandés. Une fois NetApp a approuvé votre inscription, vous pouvez télécharger le logiciel requis. L'approbation peut prendre jusqu'à 24 heures.

Et ensuite ?

Après avoir préparé l'installation de vos nœuds de calcul de données, "[installer les Data Compute Node](#)".

Installez vos nœuds de calcul de données pour AI Data Engine

Installez et sécurisez vos nœuds de calcul de données dans l'armoire.

Avant de commencer

- Assurez-vous d'avoir les instructions emballées avec le kit de rails.

- Comprenez les préoccupations de sécurité liées au poids du nœud de calcul de données, du système de stockage et de l'étagère de stockage.
- Il faut comprendre que le flux d'air à travers le système de stockage entre par l'avant, là où le panneau ou les embouts sont installés, et s'évacue par l'arrière, là où se trouvent les ports.



En général, les commutateurs doivent être installés au centre de l'armoire. Les étagères de stockage doivent être installées sous le commutateur et au-dessus d'un second commutateur installé. Les nœuds de contrôleur peuvent être installés au-dessus ou en dessous des commutateurs dans l'armoire. Les nœuds de calcul de données peuvent être installés au-dessus ou en dessous des nœuds de contrôleur dans l'armoire.

Étapes

1. Installez les kits de rails pour vos nœuds de calcul de données, selon les besoins, en utilisant les instructions incluses avec les kits.
2. Installez et sécurisez vos nœuds de calcul de données dans l'armoire :
 - a. Positionnez le nœud de calcul de données sur les rails au milieu de l'armoire, puis soutenez l'apppliance par le bas et faites-la glisser en place.
 - b. Fixez le nœud de calcul de données à l'armoire à l'aide des vis de montage fournies.
3. Fixez le panneau à l'avant du nœud de calcul de données.

Et ensuite ?

Après avoir installé les nœuds de calcul de données, "[câbler les nœuds de calcul de données pour AIDE](#)".

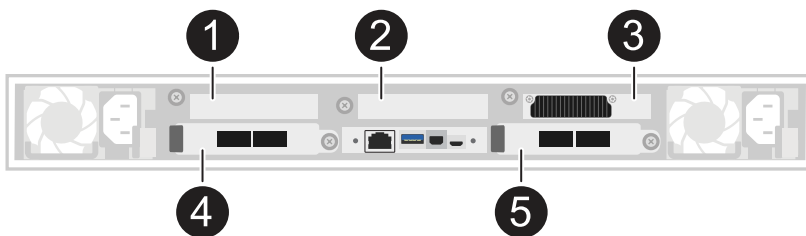
Exigences relatives au câblage des nœuds de calcul de données pour AI Data Engine

Les nœuds de calcul de données s'intègrent à votre système de stockage AFX 1K via les connexions réseau hôte et réseau du cluster. Vérifiez la configuration des emplacements d'E/S, les types de câbles et les exigences de connexion pour votre déploiement.

Configuration du câblage

Les nœuds Data Compute Node se connectent aux mêmes commutateurs de cluster que les nœuds de contrôleur AFX 1K, étendant ainsi votre système de stockage avec des ressources de calcul optimisées pour les charges de travail d'AI et de machine learning.

La configuration initiale de AI Data Engine (AIDE) prend en charge un minimum de trois Data Compute Node. Pour des détails complets sur la configuration et les priorités des emplacements, voir "[NetApp Hardware Universe](#)".



1	Emplacement inutilisé sur le Data Compute Node.
---	---

2	Emplacement inutilisé sur le Data Compute Node.
3	Emplacement GPU sur le Data Compute Node.
4	Emplacement d'E/S sur le Data Compute Node.
5	Emplacement d'E/S sur le Data Compute Node.

Configuration des emplacements d'E/S

Le Data Compute Node utilise un système de numérotation des emplacements spécifique, différent des configurations de serveurs standard. Comprendre la disposition des emplacements est essentiel pour un câblage correct.

- **Emplacement 3** : Réservé au GPU (non accessible pour le câblage d'E/S)
- **Emplacements 4 et 5** : emplacements d'E/S utilisés pour les connexions réseau
 - Port a: Connexions réseau du cluster
 - Port b : Connexions réseau de l'hôte
- **Emplacements 1 et 2** : vides et inaccessibles pour utilisation

Connexions réseau

Les nœuds Data Compute Node nécessitent deux types de connexions réseau pour s'intégrer au système de stockage AFX 1K.

• Connexions réseau de l'hôte

Les connexions au réseau hôte permettent d'accéder aux données client et permettent aux Data Compute Node de traiter les charges de travail. Chaque Data Compute Node utilise les ports e4b et e5b pour des connexions redondantes à des commutateurs réseau hôte distincts.

Affectations de ports :

- e4b: Se connecte au commutateur réseau hôte A
- e5b: Se connecte au commutateur réseau hôte B

• Connexions réseau en cluster

Les connexions réseau du cluster permettent la communication entre les Data Compute Node et les nœuds de contrôleur AFX 1K au sein du cluster de stockage. Chaque Data Compute Node utilise les ports e4a et e5a pour des connexions redondantes à des commutateurs réseau distincts du cluster.

Affectations de ports :

- e4a: Se connecte au commutateur réseau du cluster A
- e5a: Se connecte au commutateur réseau du cluster B

Composants matériels pris en charge

Les Data Compute Node nécessitent des câbles et des commutateurs spécifiques pour assurer une connectivité et des performances appropriées avec le système de stockage AFX 1K.

Data Compute Node	Commutateurs compatibles	Câbles pris en charge
Data compute nodes (minimum de trois requis)	• Cisco Nexus 9332D-GX2B (400GbE) • Cisco Nexus 9364D-GX2A (400GbE)	• Câbles de dérivation 400GbE QSFP-DD vers 4x100GbE QSFP56 pour les connexions aux Data Compute Node : ◦ 100GbE vers les ports réseau du cluster Data Compute Node (e4a, e5a) ◦ 100GbE vers les ports réseau hôtes du Data Compute Node (e4b, e5b) • Câbles RJ-45 pour les connexions de gestion

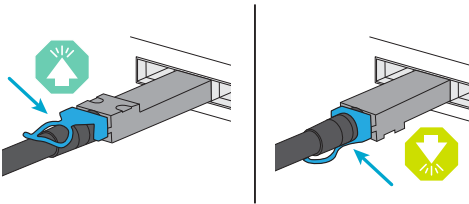


Les câbles de dérivation fournissent quatre connexions 100GbE à partir de chaque port de commutateur 400GbE. Connectez l'extrémité 400GbE aux commutateurs et l'extrémité 100GbE aux ports d'E/S du Data Compute Node.

Orientation du câble

Lors du raccordement des câbles aux Data Compute Node, une orientation correcte des câbles garantit des connexions fiables.

Les schémas de câblage dans les procédures d'installation montrent des icônes de flèches indiquant la bonne orientation (haut ou bas) de la languette de traction du connecteur de câble lors de l'insertion d'un connecteur dans un port. Lorsque vous insérez le connecteur, vous devriez le sentir s'enclencher. Si vous ne le sentez pas s'enclencher, retirez-le, retournez-le et réessayez.



Manipulez les composants délicats du connecteur avec précaution lors de leur mise en place.

Et ensuite ?

Après avoir examiné la configuration du câblage, "[câbler le matériel](#)" pour vos Data Compute Node.

Câblez vos Data Compute Node pour AI Data Engine

Connectez vos nœuds de calcul aux commutateurs du réseau hôte et du réseau de cluster pour activer le traitement des charges de travail d'IA et l'intégration avec votre système de stockage AFX 1K. Cette procédure utilise des connexions 100GbE pour l'accès au réseau hôte et la communication au sein du cluster, permettant aux nœuds de tirer parti de l'infrastructure de cluster existante sans éteindre le système AFX.

À propos de cette tâche

Ces procédures présentent des configurations courantes. Le câblage spécifique dépend des composants commandés pour votre système de stockage. Pour des détails complets sur la configuration et les priorités des emplacements, consultez "[NetApp Hardware Universe](#)".



Il n'est pas nécessaire de mettre hors tension le système de stockage AFX 1K lors du câblage des Data Compute Node. Vous pouvez ajouter les Data Compute Node à un système de stockage AFX 1K existant qui est déjà sous tension et configuré.

Avant de commencer

- Vous disposez déjà d'un système de stockage AFX 1K installé. Pour plus d'informations sur l'installation du système de stockage AFX 1K, consultez "[Documentation d'installation du système de stockage AFX 1K](#)".
- Vous avez installé et configuré les commutateurs réseau requis. Contactez votre administrateur réseau pour obtenir des informations sur la connexion du système à vos commutateurs réseau.
- Vous avez examiné le "[exigences de câblage pour les Data Compute Node](#)".



Un minimum de trois nœuds de calcul de données est requis pour déployer l'AI Data Engine.

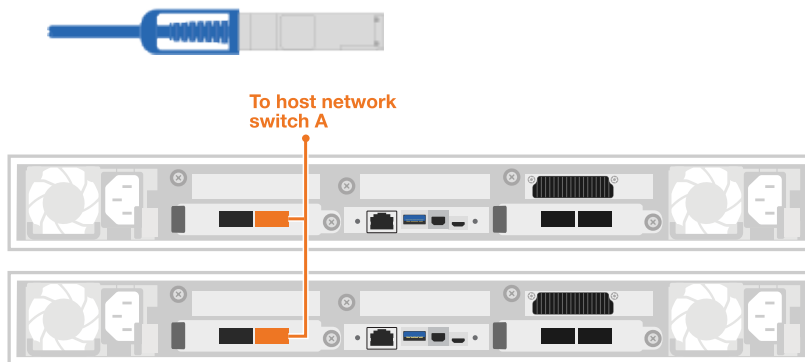
Étape 1 : Connectez les Data Compute Node au réseau hôte

Vous pouvez connecter les ports du Data Compute Node à votre réseau hôte.

Étapes

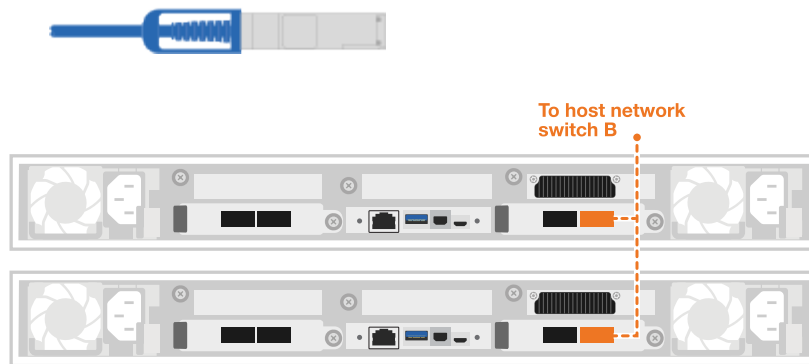
1. Connectez le port e4b des Data Compute Node suivants au commutateur de réseau de données Ethernet A :
 - Data Compute Node 1, port e4b
 - Data Compute Node 2, port e4b

Câbles 100GbE



2. Connectez le port e5b des Data Compute Node suivants au commutateur de réseau de données Ethernet B :
 - Data Compute Node 1, port e5b
 - Data Compute Node 2, port e5b

Câbles 100GbE



Étape 2 : Câbler les connexions du cluster de Data Compute Node

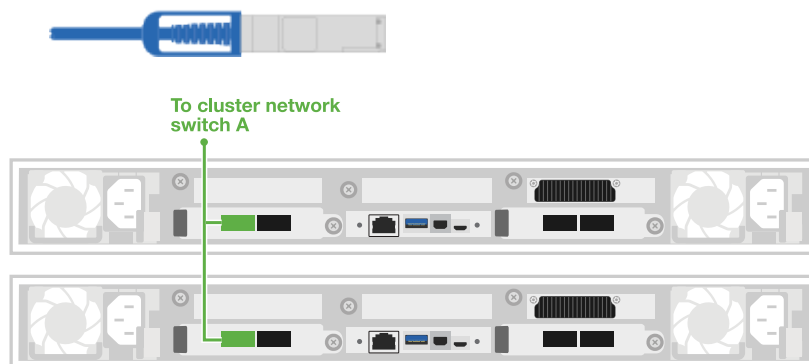
Pour les nœuds de calcul de données, utilisez des câbles de dérivation 4x100GbE pour connecter les ports e4a/e5a pour les connexions du cluster.

Étapes

1. Connectez le port e4a des Data Compute Node suivants à un port non-ISL sur le commutateur réseau du cluster A :

- Data Compute Node 1, port e4a
- Data Compute Node 2, port e4a

Câbles de dérivation 4x100GbE

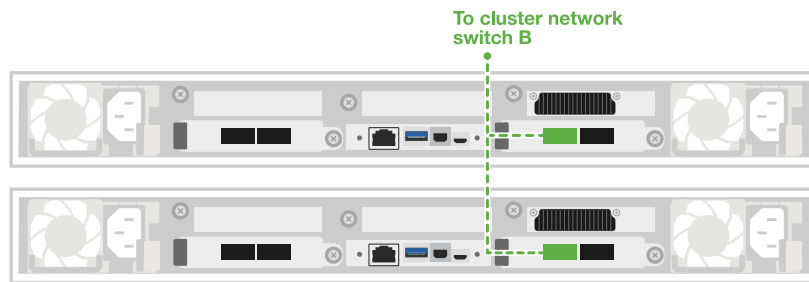


2. Connectez le port e5a des Data Compute Node suivants à un port non-ISL du commutateur réseau du cluster B :

- Data Compute Node 1, port e5a
- Data Compute Node 2, port e5a

Câbles de dérivation 4x100GbE





Et ensuite ?

Une fois le matériel câblé, "[Mettez sous tension vos Data Compute Node](#)".

Mettez sous tension vos nœuds de calcul de données pour AI Data Engine

Après avoir installé le matériel du rack et câblé vos nœuds de calcul de données, vous devez mettre sous tension vos DCN et vos nœuds de contrôle pour le système de stockage AFX si ceux-ci ne sont pas déjà sous tension.

Avant de commencer

- Assurez-vous que vos étagères sont alimentées et que chacune possède un identifiant unique. Pour plus d'informations sur l'attribution des identifiants d'étagère pour le système de stockage AFX, consultez la "[documentation relative à l'attribution d'identifiants uniques pour les étagères](#)".

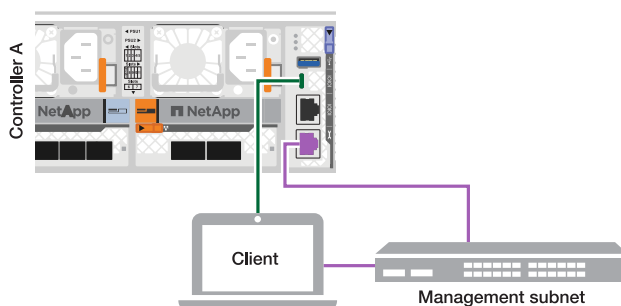
Étapes

Après avoir allumé vos étagères de stockage et attribué les identifiants uniques, mettez sous tension vos DCN et mettez sous tension les nœuds de contrôleur de stockage s'ils ne sont pas déjà allumés.

- Connectez votre ordinateur portable au port console série. Cela vous permet de surveiller la séquence de démarrage lorsque les contrôleurs sont mis sous tension.
 - Configurez le port de console série de l'ordinateur portable à 115 200 bauds avec N-8-1.

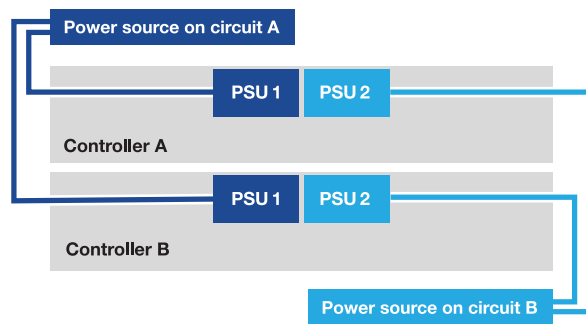
Consultez l'aide en ligne de votre ordinateur portable pour obtenir des instructions sur la façon de configurer le port de console série.

- Connectez le câble de la console à l'ordinateur portable, puis connectez le port de console série du contrôleur à l'aide du câble de console fourni avec votre système de stockage.
- Connectez l'ordinateur portable au commutateur du sous-réseau de gestion.



- Attribuez une adresse TCP/IP à l'ordinateur portable, en utilisant une adresse qui se trouve sur le sous-réseau de gestion.
- Branchez les cordons d'alimentation aux blocs d'alimentation du contrôleur, puis connectez-les à des

sources d'alimentation sur des circuits différents.



- Le système commence à démarrer. Le démarrage initial peut prendre jusqu'à huit minutes.
 - Les LED clignotent et les ventilateurs se mettent en marche, ce qui indique que les contrôleurs sont en cours de mise sous tension.
 - Les ventilateurs peuvent être bruyants au démarrage, ce qui est normal.
4. Branchez les cordons d'alimentation aux alimentations des nœuds de calcul de données, puis connectez-les à des sources d'alimentation sur des circuits différents.
 5. Fixez les cordons d'alimentation à l'aide du dispositif de fixation présent sur chaque alimentation.
 6. Mettez sous tension les nœuds de calcul de données.

Il vous faudra peut-être retirer le panneau pour accéder à l'interrupteur d'alimentation ; si c'est le cas, n'oubliez pas de le remettre en place ensuite.

Et ensuite ?

Après avoir activé vos nœuds de calcul de données, "[configurer un cluster ONTAP AIDE](#)".

Configurez votre système AIDE

Configurer AI Data Engine

En tant qu'administrateur de stockage ONTAP, vous pouvez intégrer AI Data Engine (AIDE) à votre déploiement du système AFX. Vous pouvez soit configurer AIDE dans le cadre du déploiement initial du système de stockage AFX, soit ajouter AIDE à un cluster AFX déjà déployé et en service. Dans les deux cas, vous devrez également finaliser la configuration d'AIDE après avoir terminé la configuration de base.

1. Configurez AIDE et intégrez-le à un système de stockage AFX

Il existe deux façons de configurer AIDE et de l'intégrer à un système de stockage AFX. Choisissez l'option qui convient à votre environnement.



Avant de configurer AIDE, assurez-vous que les Data Compute Node sont installés et connectés aux commutateurs du cluster. Voir "[Installer AIDE](#)" pour plus d'informations.

Configurez AIDE avec un nouveau cluster AFX

Si vous configurez AIDE dans le cadre du déploiement initial d'un système AFX, le processus de configuration d'AIDE fait partie du flux de travail standard de configuration du cluster AFX. Lors de la configuration du cluster

ONTAP, System Manager détecte automatiquement les DCN connectés et les inclut dans la configuration du cluster. Après cela, vous êtes prêt à finaliser la configuration d'AIDE, y compris la configuration de la licence requise.

Configurez AIDE avec un cluster AFX existant

Si vous ajoutez un cluster de Data Compute Node à un système de stockage AFX existant, vous devez configurer les DCN et terminer la configuration de votre cluster AIDE.

Ajoutez les Data Compute Node au cluster

ONTAP détecte dynamiquement les nouveaux Data Compute Node connectés à votre réseau de cluster et les affiche dans System Manager. Vous pouvez ["Ajoutez des nœuds de calcul de données à votre cluster AIDE"](#). Notez également ce qui suit :

- Les Data Compute Node doivent être physiquement connectés aux commutateurs du cluster. Exactement trois DCN sont requis lors de la création d'un cluster AIDE.
- Vous devez disposer d'un nombre suffisant d'adresses IP pour le réseau DCN et d'autres informations de configuration réseau requises.

Terminez la configuration de votre AIDE cluster

Après avoir intégré les Data Compute Node et terminé la configuration réseau initiale, vous pouvez finaliser la configuration de votre nouveau AIDE cluster.

À propos de cette tâche

Une fenêtre d'aide à la configuration s'affiche à droite de la page avec les étapes de configuration nécessaires. Une coche s'affiche à gauche de chaque étape ou action déjà effectuée. Une flèche (→) s'affiche pour les éléments qui n'ont pas encore été complétés. Pour chaque action incomplète :

- Sélectionnez le lien du titre activé et complétez l'action de configuration pour votre environnement.
- Si le lien est désactivé, survolez le titre pour afficher l'action requise pour l'activer et continuez avec l'action de configuration.

Chaque étape comporte un lien vers une documentation complémentaire lorsque cela est possible. Une partie de la configuration peut être incluse dans la configuration du cluster AFX. Voir la ["Documentation AFX"](#) pour plus d'informations.

Étapes

1. Sélectionnez **Configurer les groupes d'agrégation de liens et les VLAN**.

Si votre environnement l'exige, vous pouvez configurer des LAG et des VLAN au niveau du cluster. La sélection de cette option ouvre la page des ports Ethernet.

2. Sélectionnez **Configurer les protocoles réseau**.

Vous pouvez configurer les protocoles réseau de la machine virtuelle de stockage (SVM).

3. Sélectionnez **Update data compute software**.

Si une image logicielle de Data Compute Node est disponible pour une mise à jour, cette option sera activée. Sélectionnez-la pour lancer la page de mise à jour du logiciel de Data Compute Node. Pour plus d'informations, consultez ["Mettre à jour AI Data Engine software"](#).

4. Sélectionnez **Configurer la mise en réseau de l'AI Data Engine**.

Cette option affiche la page permettant de configurer l'adresse réseau du data engine.

5. Sélectionnez **Créer des interfaces réseau intercluster**

Cette option ouvre la page pour configurer les interfaces réseau inter-clusters dans le cadre d'une procédure en deux étapes si vous souhaitez cataloguer des données distantes.

6. Sélectionnez **Peer avec d'autres systèmes de stockage ONTAP**.

Après avoir créé les interfaces réseau inter-clusters, vous pouvez établir des relations de peering avec d'autres systèmes de stockage ONTAP.

7. Sélectionnez **Ajouter un container de données**.

Vous pouvez ajouter un volume destiné à être utilisé par AIDE et l'associer à une SVM spécifique.

8. Sélectionnez **Add Workspaces**.

Cela ouvre la page des espaces de travail de Data Engine où vous pouvez ["créer un espace de travail"](#).

9. Sélectionnez **Configure OpenID Connect**.

Vous devrez ["configurer l'authentification OpenID"](#) pour activer l'accès à la AI Data Engine Console.

Après avoir terminé

Une fois toutes les actions terminées, la fenêtre d'aide est masquée. Vous pouvez la fermer et la rouvrir manuellement selon vos besoins.

2. Configuration complète d'AIDE

Une fois que votre cluster AIDE est configuré et intégré au système de stockage AFX, vous devez terminer la configuration AIDE.

Étapes

1. ["Ajouter les licences AIDE requises"](#).

Il existe une licence AI Data Engine que vous devez installer pour bénéficier de la fonctionnalité complète d'AIDE, y compris la vectorisation et les garde-fous.

2. ["Configurer l'authentification OIDC"](#).

Veillez à configurer OIDC pour tous les déploiements. Vous devez configurer OIDC pour accéder à la AI Data Engine Console.

Installez une licence AIDE dans votre système AFX

Vous devrez installer une licence NetApp pour accéder à toutes les fonctionnalités de l'AI Data Engine (AIDE). En tant qu'administrateur du cluster ONTAP, vous pouvez effectuer la gestion des licences à l'aide d'ONTAP System Manager.

Préparez-vous à acquérir une licence pour AI Data Engine

Une licence est un enregistrement d'un ou plusieurs droits d'utilisation de logiciels. Toutes les licences AFX sont livrées sous forme de fichier de licence NetApp (NLF), qui est un fichier unique permettant d'activer plusieurs fonctionnalités. Il y a plusieurs éléments à prendre en compte avant d'installer une licence sur votre système de stockage AFX pour la prise en charge d'AIDE.

Types de licences

Il existe deux principaux types de licences nécessaires pour commencer à utiliser AIDE via votre système de stockage AFX.

Licence ONTAP One

La licence Metadata Engine Basic est généralement installée en usine dans le cadre de la licence ONTAP One. Elle permet d'accéder à la fonctionnalité Metadata Engine qui constitue la base essentielle pour les opérations AIDE. Toutes les capacités principales nécessaires à l'administration de votre système ONTAP sont également incluses.

AI Data Engine

Vous devez acheter et installer la licence AI Data Engine pour accéder aux services premium nécessaires à l'activation de toutes les capacités d'AIDE. La licence débloque vos nœuds de calcul de données, activant des fonctionnalités d'IA telles que la vectorisation, les garde-fous de gouvernance, l'inférence et une expérience UI intégrée. La licence inclut un nombre de GPU ainsi qu'une date d'expiration.

Exigences d'installation de la licence

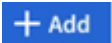
Vous devrez acheter la licence AIDE et télécharger le fichier NLF associé sur votre système local. Vous pourrez ensuite transférer le fichier vers votre système de stockage AFX via System Manager. Assurez-vous également de disposer des éléments suivants :

- Identifiants d'administrateur pour se connecter à ONTAP System Manager
- Un cluster AFX exécutant ONTAP 9.18.1 et versions ultérieures

Installez une licence sur votre système AFX

Vous pouvez installer une licence AIDE pour activer les fonctionnalités AIDE supplémentaires nécessaires à votre système de stockage AFX.

Étapes

1. Dans le System Manager, sélectionnez **Cluster** puis **Settings**.
2. À côté de **Licences**, sélectionnez ➔.
3. Sélectionnez l'onglet **Fonctionnalités** pour afficher les fonctionnalités ONTAP disponibles.
4. Pour installer une licence, sélectionnez l'onglet **Licences installées**.
5. Sélectionnez  **Add**.
6. Sélectionnez un fichier de licence local et sélectionnez **Add**.

Informations connexes

- ["Aperçu des licences ONTAP"](#)
- ["Comment télécharger les licences NLF depuis le site d'assistance NetApp"](#)

- "ONTAP CLI : system license add"

Configurer OpenID Connect pour AIDE dans ONTAP

En tant qu'administrateur du cluster ONTAP, vous pouvez utiliser ONTAP System Manager pour configurer l'authentification OpenID Connect (OIDC) pour un cluster AI Data Engine (AIDE). Cela permet une connexion sécurisée et centralisée via un fournisseur d'identité externe (IdP).



Vous devez configurer OIDC pour accéder à la AI Data Engine Console. Une fois configuré, toute l'authentification transite par OIDC. Si OIDC n'est pas configuré, la console sera inaccessible aux administrateurs ainsi qu'aux ingénieurs de données et aux data scientists. Dans ce cas, la connexion à System Manager revient à l'authentification locale.

Notez également les points suivants concernant la configuration OIDC pour l'accès AIDE :

- Vous ne pouvez pas modifier une configuration OIDC existante. Si vous devez apporter une modification, supprimez d'abord la configuration et créez-en une nouvelle avec les paramètres souhaités.
- Si vous désactivez ou supprimez OIDC, System Manager reviendra à l'authentification utilisateur ONTAP locale.

Aperçu de l'OIDC

OpenID Connect (OIDC) est un protocole d'authentification construit sur le framework OAuth 2.0. Il étend OAuth 2.0, qui est utilisé principalement pour l'autorisation, en ajoutant une couche d'identité. OIDC introduit le concept de jeton d'identité, qui est un JSON Web Token (JWT) contenant des revendications sur l'événement d'authentification et l'identité de l'utilisateur.

Vous devez sélectionner et configurer un fournisseur d'identité externe (IdP) pris en charge par AFX avec AIDE. L'IdP authentifie les utilisateurs et émet des jetons qu'AFX, via System Manager, peut utiliser pour autoriser l'accès à la AI Data Engine Console.

Configurer les fournisseurs d'identité tiers

Pour vous authentifier à l'aide d'OIDC, vous devez d'abord configurer un fournisseur d'identité externe. L'implémentation OIDC d'ONTAP utilise les revendications de rôle dans les jetons pour appliquer le RBAC. Lors de la configuration d'un fournisseur d'identité, assurez-vous qu'il est configuré pour renvoyer les revendications de rôle dans le jeton d'identité et le jeton d'accès. ONTAP prend en charge deux IdPs pour l'authentification OIDC : Entra ID et Active Directory Federation Services (AD FS).

Entra ID

Vous pouvez configurer Entra ID en utilisant les étapes générales suivantes :

1. Créez un nouvel enregistrement d'application sur la page de configuration Entra ID.
2. Définissez la valeur URI de redirection (Web) sur `https://$CLUSTER_MGMT_IP/oidc/callback`, en remplaçant l'adresse IP de gestion du cluster ou le FQDN approprié.
3. Créez les rôles requis sous App Roles et attribuez-les à vos utilisateurs.
4. Mettez à jour les revendications du jeton sous Configuration du jeton pour renvoyer les rôles dans id-token et access-token.

Voir "[Configurez un fournisseur OpenID Connect avec Entra ID](#)" pour plus d'informations.

Active Directory Federation Services

Vous pouvez configurer AD FS en suivant les étapes générales suivantes :

1. Créez un nouveau groupe d'applications et sélectionnez **Server application accessing a web API**.
2. Définissez la valeur URI de redirection (Web) sur `https://$CLUSTER_MGMT_IP/oidc/callback`, en remplaçant l'adresse IP de gestion du cluster ou le FQDN approprié.
3. Configurez les revendications pour renvoyer les rôles dans les jetons.

Voir "[Ajouter AD FS en tant que fournisseur d'identité OpenID Connect](#)" pour plus d'informations.

Configurer OIDC dans System Manager

Après avoir configuré votre IdP, vous pouvez configurer l'authentification OIDC dans System Manager pour permettre un accès sécurisé à l'AI Data Engine Console.

Avant de commencer

- Vous devez disposer d'un accès administrateur à System Manager.
- Votre fournisseur d'identité OIDC doit être configuré et accessible.

Étapes

1. Dans le Gestionnaire système, sélectionnez **Cluster** puis **Paramètres** et localisez la carte OpenID Connect.
2. Si OIDC est déjà configuré, vous pouvez modifier ou désactiver la configuration. Si OIDC n'est pas configuré, sélectionnez  pour démarrer le processus de configuration.
3. Sous Configurer OpenID Connect, indiquez les valeurs des champs suivants :
 - Fournisseur
 - Émetteur
 - URI de l'ensemble de clés Web JSON
 - Point de terminaison d'autorisation
 - point de terminaison du token
 - Point de terminaison de fin de session
 - Émetteur de jetons d'accès (facultatif)
4. Sous Configuration du client, fournissez des valeurs pour les champs suivants :
 - ID client
 - Réclamation de l'utilisateur distant
 - Intervalle de rafraîchissement
5. Sous Détails de la connexion, fournissez des valeurs pour les champs suivants :
 - Adresse IP du cluster ou FQDN
 - Proxy sortant (facultatif)
6. Sous Mappage des rôles externes, sélectionnez un mappage de rôle existant ou définissez un nouveau rôle pour l'ONTAP `admin user`.

7. Sélectionnez **Activer maintenant** puis **Enregistrer**. System Manager s'actualisera pour appliquer les nouveaux paramètres d'authentification.
8. Connectez-vous avec vos identifiants IdP ; après une authentification réussie, vous serez redirigé vers System Manager.

Informations connexes

- ["OpenID Foundation"](#)
- ["Implémentation ONTAP OAuth 2.0"](#)

Configurer des espaces de travail

Préparez les données pour AI Data Engine

Après avoir créé un cluster, établissez un container de données contenant les données que vous prévoyez d'utiliser avec AI Data Engine (AIDE). Ce container de données doit être un volume ONTAP, soit un volume local, soit un volume provenant d'un cluster ONTAP apparié exécutant ONTAP 9.

Au lieu de charger manuellement les données du cluster ONTAP dans AIDE, vous devrez appairer les clusters et SVM ciblés avec le cluster AIDE, puis décider quels volumes NFS vous souhaitez utiliser avec le catalogue de métadonnées AIDE. Après avoir créé le container de données, vous pouvez créer un espace de travail et associer le container de données à cet espace de travail. Les utilisateurs de cet espace de travail peuvent alors accéder et interagir avec les collections de données et les ressources affiliées à l'espace de travail pour leurs charges de travail d'IA.

À propos de cette tâche

Vous devez connecter chaque SVM contenant les données que vous souhaitez utiliser avec AI Data Engine. La connexion du cluster seul ne suffit pas. Cela garantit que AI Data Engine puisse accéder à vos données, les intégrer et les indexer comme prévu.

Vous devez connecter la SVM dans le cluster AIDE qui servira de destination SnapMirror.

Il n'est pas nécessaire de créer des relations SnapMirror entre les clusters ONTAP et le cluster AIDE. Ces relations seront créées automatiquement lors de la création de l'espace de travail.

Avant de commencer

- Vous avez besoin des privilèges d'administrateur de stockage pour appairer les clusters et les SVM et sélectionner les container de données.
- Vous avez identifié les ONTAP clusters et les SVM qui contiennent les données que vous souhaitez utiliser avec AI Data Engine.
- Vous avez confirmé que les volumes des sources de données répondent aux exigences suivantes :
 - Les volumes sont en ligne et accessibles.
 - Le protocole NFS est activé. Seuls les volumes compatibles NFS sont pris en charge comme container de données pour AI Data Engine. Les volumes SMB et CIFS ne sont pas pris en charge.
 - Les volumes ne sont pas des volumes FlexCache.
 - Les sources de données sont des volumes en lecture-écriture. Les volumes de protection des données ne sont pas pris en charge.

Étapes

1. "[Appairer chaque cluster ONTAP et SVM](#)" qui contient les données que vous souhaitez utiliser avec AI Data Engine.
2. "[Sélectionnez les volumes que vous souhaitez utiliser avec AI Data Engine](#)".

Pour chaque volume, notez les informations suivantes :

- Nom du volume
- UUID

- Nom et UUID de la SVM
- Nom du cluster et UUID

Et ensuite ?

"[Créer un espace de travail](#)" et associez les containers de données que vous avez créés à un workspace.

Informations connexes

- "[Options de migration de données](#)"

Créez un espace de travail dans AI Data Engine

Une fois votre cluster configuré, vous pouvez créer un espace de travail. Les espaces de travail vous permettent de segmenter les données du cluster, de contrôler l'accès aux données pour chaque utilisateur et d'exclure les données auxquelles AI Data Engine (AIDE) ne doit pas accéder.

Si vous administrez le stockage, vous utiliserez ONTAP System Manager pour créer et gérer des workspaces.

Les organisations créent des espaces de travail en fonction des équipes, des projets, des niveaux de sensibilité des données ou d'autres critères pertinents. Par exemple, si vous travaillez dans le domaine de la santé, vous pouvez segmenter les données cliniques dans un espace de travail, mais exclure les données relatives à l'informatique, au service juridique ou à d'autres départements.

À propos de cette tâche

Les limites de traitement du système affectent la création d'espaces de travail (généralement jusqu'à 15 Go par jour par cluster). Si vous créez plusieurs espaces de travail en parallèle ou en succession rapide, chaque espace de travail peut prendre plus de temps à traiter et vous pourriez rencontrer des retards importants.

Suivez l'état de la création des espaces de travail depuis la page d'inventaire des espaces de travail. Pour de meilleurs résultats, évitez de créer plusieurs espaces de travail à la fois si vous avez besoin d'accéder immédiatement à ces fonctionnalités.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage pour créer des espaces de travail et associer des collectes de données.
- Vous avez déterminé les sources de données distantes (pairées) et locales que vous avez l'intention d'utiliser avec l'espace de travail et avec AI Data Engine.
- Vous avez "[créé au moins un container de données](#)" que l'espace de travail peut utiliser, tel qu'un volume local ou un volume provenant d'un cluster apparié.



Ajoutez un volume à un espace de travail que vous ne supprimerez pas pendant la durée de vie prévue de cet espace de travail. Si vous supprimez un volume après l'avoir ajouté à un espace de travail, l'espace de travail passera en état d'échec. Confirmez la viabilité à plus long terme du volume avant de créer un espace de travail.

- Assurez-vous que NFS est activé sur le volume, mais que CIFS ne l'est pas. Les espaces de travail prennent uniquement en charge les volumes avec NFS. Les volumes avec CIFS (SMB) ne sont pas pris en charge.

Créer un espace de travail

Créez un espace de travail et associez des container de données contenant les données que vous souhaitez utiliser avec AI Data Engine.

Étapes

1. Dans ONTAP System Manager, accédez à **AI Data Engine > Workspaces**.
2. Sélectionnez **Add**.
3. Dans la boîte de dialogue **Ajouter un espace de travail**, sélectionnez au moins un container de données disponible à associer à l'espace de travail.
4. Configurez "[clusters appariés](#)" afin que les données de ces clusters puissent être accessibles dans l'espace de travail
5. Si vous souhaitez configurer l'accès des utilisateurs à l'espace de travail, vous pouvez le faire maintenant ou "[Attendez que l'espace de travail soit créé](#)".
6. Configurez un intervalle d'actualisation pour la fréquence à laquelle l'espace de travail se synchronise avec les data containers associés afin de capturer les données nouvelles ou mises à jour (par exemple, six heures).



Choisissez un intervalle qui assure un bon équilibre entre la fraîcheur des données et les performances du système. Si vous ajoutez un container de données à plusieurs espaces de travail, le système utilise automatiquement l'intervalle le plus agressif (le plus court). Pour en savoir plus, consultez la documentation sur [Actualisation et versionnage de l'espace de travail](#).

7. Sélectionnez **Continue**.
8. Dans la boîte de dialogue **Finaliser workspace**, saisissez un nom et une description pour l'espace de travail.
9. Sélectionnez **Add** pour créer l'espace de travail.

Résultat

Le processus de création de l'espace de travail prend de plusieurs minutes à plusieurs heures, selon l'ensemble de données associé, le nombre de fichiers, la taille des fichiers et d'autres facteurs.

Le système extrait automatiquement les métadonnées de toutes les sources de données et les stocke dans un catalogue de métadonnées que les utilisateurs peuvent utiliser pour localiser les fichiers dont ils ont besoin pour leurs projets. Après avoir affecté les utilisateurs à l'espace de travail, les ingénieurs de données peuvent configurer et interagir avec les composants affiliés à l'espace de travail depuis AI Data Engine Console.

Le nouvel espace de travail apparaît sur la page Espaces de travail dans l'état `Creating` jusqu'à ce que le processus soit terminé et que l'état passe à `ready`.

Examiner les détails de l'espace de travail

Après la création de l'espace de travail, vérifiez les détails de l'espace de travail.

Étapes

1. Examinez les détails de l'espace de travail, y compris la taille totale, le pourcentage de capacité du cluster utilisé et la date de la dernière actualisation de l'espace de travail.
2. Sélectionnez le nom de l'espace de travail pour ouvrir la page de détails.

3. Dans l'onglet Vue d'ensemble, consultez les détails de l'espace de travail, notamment les containers de données associés, les utilisateurs et l'activité.

Actualisation et versionnage de l'espace de travail

Chaque actualisation d'un espace de travail crée une version immuable qui capture l'état actuel de tous les fichiers et objets dans l'espace de travail. Les versions incluent des métadonnées complètes, des références aux instantanés utilisés lors de l'extraction et un identifiant de tâche pour la traçabilité. Cela prend en charge la traçabilité des données, la reproductibilité et l'audit.

Les actualisations s'effectuent soit selon la planification que vous configurez (par exemple, toutes les six heures), soit lorsque vous les déclenchez manuellement. L'intervalle d'actualisation minimal pris en charge est d'une heure ; le maximum est d'un an. Si un container de données est inclus dans plusieurs espaces de travail, le système utilise l'intervalle d'actualisation le plus fréquent et de plus courte durée pour planifier l'extraction des métadonnées.

Par défaut, le système conserve les versions précédentes, actuelles et suivantes (en cours). Le système conserve les versions plus anciennes conformément à la politique de votre organisation et peut les supprimer si nécessaire.

Vous pouvez lister toutes les versions d'un espace de travail et visualiser les différences entre versions afin d'identifier les fichiers ou objets ajoutés, modifiés ou supprimés. Cela vous permet de suivre les modifications au fil du temps et de comprendre l'évolution des données de votre espace de travail.

Attribuer l'accès utilisateur aux espaces de travail AI Data Engine

En tant qu'administrateur de stockage, vous affectez les utilisateurs à un espace de travail en fonction de leur rôle : data engineers, data scientists ou autres rôles selon la structure et les besoins de votre organisation. Les utilisateurs se connectent à la AI Data Engine Console à l'aide de leurs identifiants et accèdent aux ressources du data container au sein de l'espace de travail qui leur est attribué.

ONTAP System Manager vous permet de gérer quels utilisateurs ont accès aux espaces de travail AI Data Engine. Vous pouvez ajouter ou supprimer des utilisateurs pour contrôler qui peut consulter, modifier et interagir avec les données et les activités des espaces de travail.

Avant de commencer

- Vous devez disposer des privilèges de *storage administrator* pour gérer l'accès des utilisateurs aux espaces de travail.
- Confirmez que vous avez créé l'espace de travail et qu'il est actif dans l'inventaire des espaces de travail.
- Vérifiez que tous les container de données pertinents ont été ajoutés à l'espace de travail et sont accessibles.
- "[Confirmez que OIDC est activé et configuré pour le cluster](#)". Le mappage des rôles de l'IdP vers les rôles ONTAP doit être effectué pour chaque utilisateur ou groupe IdP concerné, ingénieur de données ou data scientist.

Étapes

- Ajouter un utilisateur à un espace de travail :
 - a. Dans ONTAP System Manager, accédez à **AI Data Engine > Workspaces**.

- b. Sélectionnez le nom de l'espace de travail pour ouvrir sa page de détails.
 - c. Accédez à l'onglet Utilisateurs.
 - d. Sélectionnez le bouton **Ajouter** pour ouvrir la boîte de dialogue d'ajout d'utilisateurs.
 - e. Saisissez les informations d'un ou plusieurs utilisateurs. Saisissez les informations sous forme de liste d'utilisateurs OIDC séparés par des virgules.
 - f. Sélectionnez **Ajouter** pour accorder à l'utilisateur l'accès à l'espace de travail.
- Supprimer un utilisateur d'un workspace :
 - a. Dans l'onglet Utilisateurs de la page de détails de l'espace de travail, localisez et sélectionnez l'utilisateur que vous souhaitez supprimer.
 - b. Sélectionnez le bouton **Remove**.
 - c. Confirmez la suppression dans la boîte de dialogue.
 - d. Le système supprime immédiatement l'utilisateur et révoque son accès à l'espace de travail.

Résultat

Seuls les utilisateurs listés dans l'onglet Utilisateurs de l'espace de travail peuvent accéder et interagir avec les données et les activités de l'espace de travail.

Administrer et surveiller

Surveiller les processus du cluster

Afficher l'état du système AIDE et l'état du cluster

En tant qu'administrateur de stockage, vous pouvez utiliser ONTAP System Manager pour accéder au tableau de bord et afficher l'état du cluster. Il s'agit d'une bonne première étape avant de commencer vos tâches d'administration AIDE ou si vous suspectez un problème opérationnel.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage pour effectuer des tâches administratives liées à AIDE ONTAP.

Surveillez la santé et la capacité d'AIDE depuis le tableau de bord

1. Connectez-vous à ONTAP System Manager en utilisant l'adresse de gestion du cluster :

`https://$FQDN_OR_IP/`

2. Sign in avec un compte administrateur.
3. Sélectionnez **Dashboard** dans le volet de navigation de gauche.
4. Consultez la vignette **Health** :
 - Confirmer l'état général du cluster.
 - Vérifiez le nombre et l'état des **Data compute nodes** et leur statut.
 - Vérifiez les alertes :
 - Problèmes de nœud DCN ou problèmes de connectivité
 - Espaces de travail ou collecte de données en erreur (par exemple, échecs de publication de collecte)
5. Consultez la tuile **Capacité** :
 - Notez la capacité totale du cluster et la capacité utilisée.
 - Pour les clusters AIDE, vérifiez :
 - Capacité utilisée par les métadonnées AIDE et les volumes d'applications (VM de stockage de métadonnées)
 - Capacité utilisée par les espaces de travail et les collectes de données (si disponible)
6. Vous pouvez également consulter les vignettes **Réseau** et **Performances** pour comprendre le comportement global du cluster qui pourrait avoir un impact sur les charges de travail AIDE (par exemple, la congestion du réseau ou le retard de protection).

Consulter les données DCN sur la santé et l'utilisation

1. Dans le volet de navigation, sélectionnez **Cluster** et ensuite **Overview**.
2. Sélectionnez l'onglet **Data compute**.

Cet onglet affiche tous les nœuds DCN du cluster avec :

- Nom du nœud, modèle, numéro de série et version logicielle
- État global du nœud
- Utilisation du CPU et de la mémoire
- Utilisation du GPU (si des GPU sont présents)
- Indicateurs d'erreur au niveau des nœuds

3. Développez un nœud DCN pour ouvrir la vue détaillée et vérifiez :

- Utilisation du processeur et de la mémoire du système
- Utilisation de la mémoire GPU
- Problèmes matériels ou de service signalés

4. Sélectionnez **Câblage** sur la page **Cluster > Vue d'ensemble** pour vérifier que les nœuds DCN sont correctement câblés aux commutateurs du cluster et pour identifier tout problème de port ou de liaison.

Surveiller les espaces de travail et l'encombrement des métadonnées

1. Dans le volet de navigation, sélectionnez **Data engine** puis **Workspaces**.

2. Consultez le résumé de l'espace de travail en haut de la page :

- Nombre d'espaces de travail et leurs états (par exemple, `Processing`, `Healthy`, `Error`).
- Taille totale de l'espace de travail.
- Pourcentage de la capacité du cluster consommée par tous les espaces de travail.

3. Examinez la grille de l'espace de travail :

- Vérifiez que les espaces de travail critiques affichent un état **Healthy**.
- Vérifiez les tailles des espaces de travail et la consommation de capacité.
- Recherchez les espaces de travail en `Error` ou en `Processing` état de longue durée.

4. Pour consulter les détails d'un espace de travail spécifique, sélectionnez son nom :

- Dans l'onglet **Aperçu**, confirmez :
 - État et taille de l'espace de travail
 - Conteneurs de données (volumes) inclus et leur nombre d'éléments
 - Dernière heure de mise à jour pour chaque source de données
- Dans l'onglet **Collecte de données**, confirmez :
 - Quelles collections de données existent pour cet espace de travail (les collections de données sont en lecture seule dans System Manager)
 - Leur état, leur taille et leur dernière date de mise à jour
- Dans l'onglet **Utilisateurs**, vérifiez quels utilisateurs AI Data Engine Console ont accès.

Surveillez les métadonnées Storage VM et la protection gérée par AIDE

1. Dans le volet de navigation, sélectionnez **Cluster** puis **Storage VMs**.

2. Localisez la Storage VM avec le sous-type `data-engine` (la metadata SVM) :

- Confirmez que le SVM de métadonnées est en ligne.

- Vous pouvez également ouvrir ses détails pour voir les comptes pour :
 - Volumes
 - LIFs avec type `Data compute network` (utilisées pour la communication DCN-ONTAP)
- 3. Sélectionnez **Protection** puis **Relationships** pour afficher la protection des sources de données distantes utilisées dans les espaces de travail :
 - Identifiez les relations SnapMirror créées par AIDE selon le modèle de dénomination :
 - Volume de destination `<source_volume_name>_dest_<source_volume_UUID>`
 - Politique: `<source_volume_name>_dest_aide_policy_<source_volume_UUID>`
 - Utilisez cette vue pour vérifier que les relations sont saines et que le délai de latence correspond aux attentes de rafraîchissement de l'espace de travail.



Ne modifiez pas la machine virtuelle de stockage des métadonnées, les relations SnapMirror créées par AIDE, ni les snapshots gérés par AIDE (ou leurs plannings) directement dans ONTAP. Les modifications peuvent perturber l'historique des versions d'AIDE. ["Ajuster les paramètres d'actualisation de l'espace de travail"](#) si vous devez ajuster le comportement d'actualisation.

Consultez les alertes et notifications relatives à AIDE

1. Dans le volet de navigation, sélectionnez **Events & Jobs** puis **System alerts**.
2. Passez en revue toutes les alertes actives liées à :
 - État ou connectivité du nœud DCN
 - Problèmes de réseau du data engine
 - Erreurs d'espace de travail ou de collecte de données
 - Incompatibilités de versions logicielles entre ONTAP et le cluster DCN
3. Au besoin, configurez les destinations de notification (par exemple, e-mail, syslog) dans **Cluster > Settings > Notification management** pour vous assurer que les alertes liées à AIDE sont transmises à vos outils d'exploitation.

Informations connexes

["Préparez-vous à administrer votre système de stockage AFX"](#)

Consultez les informations pour optimiser votre système AIDE

En tant qu'administrateur de stockage, vous pouvez utiliser la fonctionnalité *Insights* d'ONTAP System Manager pour afficher des suggestions de mise à jour de configuration conformes aux NetApp best practices. Ces modifications permettent d'optimiser la sécurité et les performances de votre AIDE cluster.

À propos de cette tâche

Chaque information est présentée sous forme de vignette ou de fiche distincte sur la page que vous pouvez choisir d'appliquer ou d'ignorer. Vous pouvez également sélectionner le lien de documentation associé pour en savoir plus sur une technologie spécifique.

Étapes

1. Dans le Gestionnaire système, sélectionnez **Analyse** puis **Insights**.
2. Examinez les recommandations disponibles.

Et ensuite

Effectuez l'une des actions recommandées pour mettre en œuvre les meilleures pratiques de configuration.

Afficher les événements système AIDE, les tâches et le journal des audits

En tant qu'administrateur de stockage, vous pouvez consulter les événements, les tâches et les messages du journal des audits générés par AIDE afin de suivre le traitement interne et de diagnostiquer les problèmes potentiels. Le système AIDE peut être configuré pour transmettre ces informations, ainsi que d'autres données connexes, pour un traitement et un archivage ultérieurs.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage pour effectuer des tâches administratives liées à AIDE ONTAP.

Surveiller les activités, les événements et les tâches d'AIDE

Vous pouvez utiliser la vue centralisée **Activité** pour surveiller les événements et les tâches spécifiques à AIDE dans tous les espaces de travail, ou consulter l'activité limitée à des espaces de travail individuels.

Afficher l'activité AIDE à l'échelle du cluster

Surveillez les opérations de l'espace de travail, résolvez les problèmes d'extraction de métadonnées et suivez la publication de la collecte de données sur l'ensemble de votre déploiement AIDE.

1. Dans ONTAP System Manager, dans le volet de navigation, sélectionnez **Data engine** puis **Activity**.
2. Sélectionnez l'onglet **Events** :
 - Passez en revue les événements récents propres à AIDE, tels que les suivants :
 - Création, mise à jour ou suppression d'espace de travail
 - Opérations d'ajout/suppression de conteneurs de données
 - Publication de la collecte de données (le cas échéant)
 - Utilisez des filtres (par gravité, type d'objet, espace de travail ou plage horaire) pour vous concentrer sur les événements actifs ou critiques.
3. Sélectionnez un événement individuel à ouvrir et à examiner :
 - Description et horodatage
 - Espace de travail, collecte de données ou source de données concernés
 - Action recommandée, si elle est fournie
4. Sélectionnez l'onglet **Jobs** :
 - Surveillez les tâches de longue durée telles que :
 - Extraction initiale des métadonnées d'un espace de travail
 - Actualisations de l'espace de travail / mises à jour du catalogue
 - Tâches de publication ou de mise à jour de la collecte de données

- Vérifiez l'état et la progression du travail.
5. Sélectionnez un job pour ouvrir l'aperçu et le consulter :
- Heure de début et de fin
 - Pourcentage d'avancement et phase (par exemple, **Scanning, Publishing**)
 - Espace de travail, collecte de données ou source de données concernés
 - Messages d'erreur pour les tâches ayant échoué

Afficher l'activité spécifique à l'espace de travail

Pour dépanner un espace de travail spécifique, ouvrez les détails de l'espace de travail (**Data engine > Workspaces**, sélectionnez un espace de travail) et utilisez l'onglet **Activité** :

- Examinez les événements et les tâches limités à cet espace de travail uniquement.
- Utilisez cette vue pour isoler les problèmes tels qu'un espace de travail bloqué dans `Processing` un état.

Afficher les événements à l'échelle du cluster

Consultez les messages d'événements pour obtenir un précieux historique de l'activité du système. Chaque événement comprend une description et un identifiant unique ainsi qu'une action recommandée.

1. Dans ONTAP System Manager, sélectionnez **Events & jobs** puis **Events**.
2. Examinez et répondez aux actions recommandées en haut de la page, telles que l'activation de la mise à jour automatique.
3. Sélectionnez l'onglet **Events log** pour afficher une liste des messages.
4. Sélectionnez un message d'événement pour l'examiner plus en détail, y compris le numéro de séquence, la description, l'événement et l'action recommandée.
5. Vous pouvez également sélectionner l'onglet **Active IQ suggestions** et vous inscrire auprès de Active IQ pour obtenir des informations détaillées sur les risques du cluster.

Afficher les tâches à l'échelle du cluster

Consultez toutes les tâches en cours d'exécution sur le cluster AIDE, y compris les tâches spécifiques à AIDE et les tâches ONTAP générales.

1. Dans ONTAP System Manager, sélectionnez **Events & Jobs** puis **Jobs**.
2. Personnalisez l'affichage ainsi que la recherche et le téléchargement des informations sur les emplois selon vos besoins.

Consulter le journal des audits

Utilisez le journal des audits pour consulter un enregistrement de l'activité du système basé sur l'utilisation de protocoles d'accès tels que HTTP.

1. Dans ONTAP System Manager, sélectionnez **Events & jobs** puis **journal des audits**.
2. Sélectionnez **Paramètres** pour activer ou désactiver les opérations qui sont suivies.

Gérer les notifications

Configurez les destinations de notification pour transférer automatiquement les événements AIDE et les journaux des audits.

Étapes

1. Dans ONTAP System Manager, sélectionnez **Cluster** et ensuite **Settings**.
2. Accédez à **Gestion des notifications** et sélectionnez .
3. Sélectionnez l'action appropriée pour afficher ou configurer les destinations utilisées par AIDE :
 - a. Destinations des événements : sélectionnez **View event destinations**
 - b. Destinations du journal des audits : sélectionnez **Afficher les destinations du journal des audits**
4. Sélectionnez **Add** le cas échéant et indiquez les informations de destination.
5. Sélectionnez **Save**.

Informations connexes

- ["Surveillance des événements, des performances et de l'état de santé ONTAP"](#)

Gérer les espaces de travail AI Data Engine

Un espace de travail est un ensemble de sources de données (volumes) que l'AI Data Engine (AIDE) utilise pour créer et actualiser un catalogue de métadonnées pour un projet ou un cas d'utilisation particulier. En tant qu'administrateur de stockage, vous pouvez utiliser ONTAP System Manager pour surveiller l'état de l'espace de travail, ajuster la configuration, contrôler les sources de données, gérer les utilisateurs et supprimer les espaces de travail lorsqu'ils ne sont plus nécessaires.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage pour gérer les espaces de travail.

Vérifier l'état de l'espace de travail

Vérifiez l'état de l'espace de travail, l'utilisation de la capacité et l'état des métadonnées pour vous assurer que le Metadata Engine fonctionne comme prévu et ne consomme pas de ressources inattendues.

Étapes

1. Depuis ONTAP System Manager, dans le volet de navigation, sélectionnez **Data engine > Workspaces**.
2. Consultez le résumé en haut de la page pour le nombre total d'espaces de travail, l'état général des espaces de travail et l'utilisation de la capacité.
3. Pour obtenir des informations spécifiques à un espace de travail, sélectionnez un nom d'espace de travail. Dans l'onglet **Aperçu**, confirmez :
 - État et taille de l'espace de travail.
 - Conteneurs de données (volumes) inclus dans le workspace.
 - Nombre d'éléments et date de dernière mise à jour par source de données.
 - Avertissements au niveau de l'espace de travail.



Si un espace de travail ou une collecte de données affiche un état d'erreur, vérifiez que tous les volumes sources sont en ligne et accessibles.

4. Sélectionnez l'onglet **Data collections** pour voir :

- Toutes les collections de données associées à cet espace de travail.
- État (tel que `Published` ou `Error`), taille et date de dernière mise à jour.



Le System Manager est en lecture seule pour les collections de données. Les ingénieurs de données peuvent créer et gérer des collections de données dans l'AI Data Engine Console.

5. Sélectionnez l'onglet **Users** pour afficher :

- La liste des utilisateurs ayant accès à cet espace de travail.

6. Sélectionnez l'onglet **Activité** pour voir les événements et les tâches liés uniquement à cet espace de travail.

Modifier les propriétés de l'espace de travail et le calendrier d'actualisation

Vous pouvez modifier le nom, la description, l'intervalle d'actualisation d'un espace de travail et (si vous disposez d'une licence) sa politique de garde-fous.

Étapes

1. Dans **Data engine > Workspaces**, sélectionnez  à côté d'un workspace et sélectionnez **Edit**.
2. Modifier les propriétés de l'espace de travail :
 - Mettez à jour **Name** et **Description** si nécessaire.
 - Ajustez l'**intervalle d'actualisation** (fréquence de mise à jour des métadonnées) dans la plage autorisée (heures et jours).
 - Si une licence AIDE est installée, vous pouvez sélectionner **stratégie Guardrail**.
3. Sélectionnez **Save**.



Les modifications apportées à l'intervalle d'actualisation ou au traitement des métadonnées peuvent affecter la fréquence à laquelle les relations distantes SnapMirror sont mises à jour pour cet espace de travail.

Ajouter des containers de données à un espace de travail existant

Vous pouvez ajouter des volumes montés supplémentaires (locaux ou provenant d'un cluster distant apparié) afin que leurs métadonnées soient incluses dans le catalogue de l'espace de travail.

1. Dans **Data Engine > Workspaces**, effectuez l'une des opérations suivantes :
 - Sélectionnez  à côté de l'espace de travail et sélectionnez **Ajouter des containers de données**.
 - Ouvrez l'espace de travail, sélectionnez l'onglet **Aperçu**, puis sélectionnez **Ajouter** dans la section des containers de données.
2. Dans la boîte de dialogue **Ajouter des containers de données à l'espace de travail** :
 - Localisez les volumes locaux sur le cluster AIDE.

- Développez les clusters appariés pour sélectionner les volumes distants (les volumes distants nécessitent un peering de cluster et de SVM).



Seuls les volumes en ligne éligibles qui ne sont pas exclus globalement et qui ne font pas déjà partie de l'espace de travail peuvent être sélectionnés.

3. Si vous êtes invité à mapper un volume distant :
 - Sélectionnez la machine virtuelle de stockage cible sur le cluster AIDE pour recevoir les destinations SnapMirror pour les volumes distants choisis.
4. Sélectionnez **Add**.
5. Utilisez l'onglet **Activité** de l'espace de travail ou **Data engine > Activité** pour suivre l'extraction des métadonnées et toute initialisation SnapMirror pour les nouvelles sources de données.

Supprimer les containers de données d'un workspace

Vous pouvez supprimer un container de données lorsqu'il n'est plus pertinent pour l'objectif de l'espace de travail ou si vous souhaitez réduire la portée de la gestion des métadonnées pour cet espace de travail. La suppression d'un container de données interrompt l'actualisation des métadonnées pour ce volume et supprime ses métadonnées du catalogue de métadonnées.



Ne supprimez pas un volume source d'ONTAP qui a été ajouté à un espace de travail. Si vous supprimez le volume, l'espace de travail passera en état d'échec. Supprimez toujours le container de données de l'espace de travail avant de supprimer tout volume ONTAP sous-jacent.

Étapes

1. Accédez à **Data engine > Workspaces** et sélectionnez l'espace de travail qui contient le container de données.
2. Dans l'onglet **Aperçu**, repérez le container de données à supprimer.
3. Sélectionnez **Supprimer** à côté du container de données.
4. Examinez la boîte de dialogue de confirmation et sélectionnez **Remove**.



La suppression d'un container de données d'un espace de travail ne supprime pas le volume ONTAP sous-jacent ni sa relation SnapMirror. Elle affecte uniquement l'utilisation des métadonnées au sein de l'AI Data Engine.

Gérer les utilisateurs de l'espace de travail

Vous pouvez accorder ou révoquer l'accès pour les utilisateurs data engineer et data scientist à un workspace. Ces utilisateurs sont définis dans votre identity provider (OIDC) et associés à des rôles ONTAP. Voir la documentation "[affecter des utilisateurs à des espaces de travail](#)" pour savoir comment gérer l'accès des utilisateurs.

Supprimer un espace de travail

Vous pouvez supprimer un espace de travail pour supprimer la définition de l'espace de travail et les métadonnées AIDE associées. Toutes les collections de données et les plongements vectoriels liées à l'espace de travail sont également supprimées.



Les données ONTAP sous-jacentes (volumes, SnapMirror relationships) ne sont pas supprimées.

Étapes

1. Dans **Data Engine > Workspaces**, effectuez l'une des opérations suivantes :
 - Supprimez un seul espace de travail, sélectionnez  et sélectionnez **Supprimer**.
 - Supprimez plusieurs espaces de travail, cochez les cases des espaces de travail, puis sélectionnez **Supprimer**.
2. Dans la boîte de dialogue de confirmation, examinez les conséquences de l'action avant de poursuivre :
 - Les métadonnées de l'espace de travail sont définitivement supprimées.
 - Les collections de données et les embeddings associés à l'espace de travail sont définitivement supprimés.



Il n'existe aucune option de suppression ou de restauration logicielle.

3. Cochez la case pour confirmer votre compréhension, puis sélectionnez **Delete**.

Informations connexes

- ["Attribuer des utilisateurs aux espaces de travail"](#)

Mettez à niveau et entretenez votre système AIDE

Mises à jour et compatibilité du système AI Data Engine

Maintenez les composants du système AI Data Engine (AIDE) à jour pour garantir des performances optimales et l'accès aux nouvelles fonctionnalités. Mettez à jour les composants après le déploiement, lorsque de nouveaux logiciels ou micrologiciels sont disponibles, lorsque vous ajoutez ou remplacez des nœuds, ou périodiquement pour les mises à jour de fonctionnalités.

Composants du système AIDE

Il y a deux composants principaux qui nécessitent des mises à jour dans un système AIDE : le logiciel ONTAP et le logiciel AIDE, qui inclut le firmware DCN.

Logiciel ONTAP

ONTAP est le système d'exploitation qui fonctionne sur les systèmes de stockage NetApp, y compris ceux utilisés dans les déploiements AIDE. Maintenez ONTAP à jour pour préserver la stabilité, la sécurité et la compatibilité du système avec les composants AIDE. Les composants AIDE sont mis à jour séparément.

Mises à jour du logiciel AIDE

Les mises à jour du logiciel AIDE Console et du firmware DCN sont distribuées ensemble sous forme d'un seul package (.tgz) et ne sont pas intégrées dans les images ONTAP. Les mises à jour garantissent le bon fonctionnement des composants matériels au sein du système AIDE et apportent de nouvelles fonctionnalités, des améliorations de performance et des corrections de bogues.

Comprendre le processus de mise à jour

Les mises à jour du logiciel AIDE peuvent être gérées via ONTAP System Manager.

AIDE ne prend pas en charge la fonctionnalité de mises à jour logicielles automatiques ONTAP. Vous pouvez vous inscrire pour recevoir des notifications de "[NetApp Support Site téléchargements](#)", mais toutes les mises à jour du logiciel AIDE sont effectuées manuellement par l'administrateur.

Types et étendue des versions :

- Les versions majeures d'ONTAP (9.x.x) et d'AIDE (9.x.x U0) introduisent de nouvelles fonctionnalités, des API ou des modifications qui affectent l'intégration d'ONTAP.
- Les versions correctives ONTAP (9.x.x Px) et les versions de mise à jour AIDE (9.x.x Ux) contiennent des correctifs et des mises à jour qui n'ont pas d'impact sur l'intégration ONTAP.

Matrice de compatibilité

Assurez la compatibilité entre ONTAP et AI Data Engine software lors de la planification des mises à jour.

AIDE software est distribué sous forme de versions « U ». Une version majeure d'AIDE est une version « U0 » et les versions mineures suivantes seront « U1 » et ultérieures.

Compatibilité ONTAP et AIDE

Versions AIDE	Versions ONTAP prises en charge
9.18.1 U0	9.18.1 GA et tous les 9.18.1 Px
9.18.1 U1 et versions ultérieures	9.18.1 GA et tous les 9.18.1 Px



« Px » désigne toutes les versions correctives ONTAP dans une version majeure (par exemple, 9.18.1 P1, 9.18.2 et autres).

Voies de mise à niveau AIDE

Les exemples suivants utilisent des versions futures hypothétiques pour montrer les chemins de mise à niveau et de mise à jour autorisés à partir de AIDE 9.18.1 U0 et 9.18.1 U1.

Si votre version actuelle d'AIDE est...	Et votre version cible d'AIDE est...	Votre chemin de mise à niveau ou de mise à jour est...
9.18.1 U0	9.18.1 U1	Direct
9.18.1 U0	9.18.1 U3	Direct (vous pouvez effectuer une mise à jour depuis n'importe quelle version Ux vers n'importe quelle version Ux ultérieure dans 9.18.1)
9.18.1 U1	9.18.1 U3	Direct (vous pouvez effectuer une mise à jour depuis n'importe quelle version Ux vers n'importe quelle version Ux ultérieure dans 9.18.1)

Rétablir les limitations

Les systèmes AIDE ne prennent pas en charge les opérations de restauration pour le firmware DCN, les mises à jour logicielles AIDE ou ONTAP sur les systèmes de stockage AFX. Après avoir installé une mise à jour ou

une mise à niveau, vous ne pouvez pas revenir à une version précédente. Examinez [exigences de compatibilité](#) avant de procéder à la mise à jour ou à la mise à niveau.

Informations connexes

- ["Mettre à niveau le logiciel ONTAP pour les systèmes AFX"](#)
- ["Mettre à jour le logiciel AIDE"](#)

Mettre à jour AI Data Engine software

En tant qu'administrateur de stockage, vous pouvez mettre à jour le logiciel AI Data Engine (AIDE), le firmware du data compute node (DCN) et d'autres fichiers système sur votre système AIDE à l'aide de ONTAP System Manager.

AIDE ne prend pas en charge la fonctionnalité de mises à jour logicielles automatiques ONTAP. Vous pouvez vous inscrire pour recevoir des notifications de ["NetApp Support Site téléchargements"](#), mais toutes les mises à jour du logiciel AIDE sont effectuées manuellement par l'administrateur.

À propos de cette tâche

Le package logiciel AI Data Engine est nettement plus volumineux qu'un package de mise à jour ONTAP classique (le package AIDE pèse environ 40 Go). Prévoyez des temps de chargement et d'installation plus longs lors de la mise à jour du logiciel AIDE.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage pour mettre à jour le firmware DCN et AI Data Engine software.
- Vous avez besoin de vos identifiants du site de support NetApp pour votre compte actif.
- ["Assurez la compatibilité entre ONTAP, le firmware DCN et AI Data Engine software lors de la planification des mises à jour"](#).



La fonctionnalité de réversion n'est pas prise en charge pour les mises à jour du micrologiciel DCN ou du logiciel AI Data Engine. Après l'installation d'une mise à jour, il est impossible de revenir à une version précédente.

Étapes

1. ["Téléchargez le fichier de mise à jour combiné du firmware DCN et du logiciel AIDE sur un client local"](#).
2. Dans System Manager, sélectionnez **Cluster > Paramètres > Mises à jour logicielles**.
3. À côté de **Mises à jour logicielles**, sélectionnez ➔.
4. Sous AI Data Engine updates, sélectionnez **Add AI Data Engine software files** et sélectionnez le package de mise à jour.
5. Une fois le téléchargement du package terminé, sélectionnez **Mettre à jour** pour commencer l'installation des mises à jour sur les nœuds DCN.



AIDE Console n'est pas disponible ou accessible pendant la mise à jour du nœud DCN.

Résultat

Les DCN sont mis à jour avec le AI Data Engine software et la version mise à jour est affichée pour chaque nœud.

Informations connexes

- ["Mettre à niveau le logiciel ONTAP pour les systèmes AFX"](#)

Ajoutez des nœuds de calcul de données à votre cluster AIDE

Vous pouvez ajouter des nœuds de calcul de données (DCN) lors de la création d'un nouveau cluster AI Data Engine (AIDE) ou de l'extension d'un cluster existant. Le processus consiste à découvrir et à configurer les nœuds à l'aide de ONTAP System Manager.

Préparez-vous à ajouter des nœuds

Plusieurs éléments sont à prendre en compte lors de l'ajout de nœuds DCN.

Lors de la création d'un nouveau AIDE cluster

Il faut exactement trois nœuds DCN pour un nouveau cluster AIDE.

Installation et adressabilité du matériel

Assurez-vous que les conditions préalables suivantes sont remplies :

- Le nouveau matériel DCN est installé en rack, mis sous tension et câblé aux commutateurs du cluster.
- Vous disposez d'une plage d'adresses IP disponible pour le sous-réseau backend du DCN vers ONTAP.
- Le ONTAP cluster a été initialisé et est accessible depuis l'interface logique de gestion du cluster (LIF).

Identifiants System Manager

Vous devez disposer des privilèges d'administrateur de stockage pour effectuer des tâches de création ou d'extension de cluster AIDE.

Compatibilité logicielle

Consultez la documentation suivante pour confirmer que vos versions matérielles et logicielles DCN sont compatibles avec votre cluster ONTAP :

- ["Le logiciel DCN est compatible avec la version ONTAP"](#).

Lors de l'ajout de nœuds, System Manager vérifiera que les nouveaux nœuds exécutent une version logicielle compatible avec :

- La version effective du cluster ONTAP (ECV) si c'est la première jonction DCN.
- La version du cluster DCN existant si des DCN sont déjà présents.

Si un nœud est incompatible :

- Une erreur s'affiche à côté du DCN concerné dans la boîte de dialogue **Ajouter**.
- Vous devez d'abord mettre à jour le logiciel DCN (ou ONTAP, selon le cas) vers une version compatible.

Ajouter les nœuds de calcul de données

Vous ajoutez des nœuds DCN lors de la création d'un nouveau cluster AIDE ou de l'extension d'un cluster existant.

Étapes

1. Dans System Manager, sélectionnez **Dashboard** dans le volet de navigation, puis la carte **Health**.
2. Vérifiez qu'il y a des nœuds à ajouter et sélectionnez **Afficher les détails** pour afficher la liste.

La liste contient les nœuds découverts qui ne font pas encore partie du cluster AIDE

3. Vous pouvez également sélectionner **Cluster** et **Overview** et l'onglet **Data compute** pour afficher la liste.
4. Au bas de la page de calcul des données, sélectionnez **Ajouter** au-dessus de la liste des nœuds.
5. Dans la boîte de dialogue **Ajouter des nœuds de calcul de données**, sélectionnez les nœuds DCN que vous souhaitez ajouter.

Vous pouvez éventuellement renommer chaque nœud avant de les ajouter.

6. Si vous ajoutez des nœuds pour la première fois et qu'aucun sous-réseau backend n'existe, sélectionnez **Add subnet** et indiquez :
 - Nom du sous-réseau (à usage interne)
 - Adresse et masque du sous-réseau
 - Plage d'adresses IP pour les nœuds DCN et ONTAP sur ce réseau dorsal

System Manager valide que la plage comprend suffisamment d'adresses IP libres pour tous les DCN ajoutés, tous les nœuds ONTAP du cluster, ainsi que les adresses IP flottantes supplémentaires au niveau du cluster utilisées pour la communication entre DCN et ONTAP.

7. Que vous ayez ajouté le sous-réseau backend ou qu'il existe déjà :
 - a. Examinez les adresses IP disponibles.
 - b. Si nécessaire, sélectionnez **Modifier le sous-réseau** et étendez la plage IP.
 - Vous pouvez uniquement étendre la plage. La réduction ou la modification du sous-réseau n'est pas prise en charge.
 - Modifier le sous-réseau ou la plage d'adresses IP peut nécessiter la recreation du cluster Kubernetes sous-jacent sur les DCN et peut prendre plusieurs minutes.
8. Vous pouvez éventuellement configurer l'interface du service Data Engine en fournissant :
 - Adresse IP du service
 - Masque de réseau
 - Passerelle (si nécessaire pour votre environnement)

L'adresse IP unique sera répartie entre les DCN et utilisée comme adresse frontale pour l'AI Data Engine Console et les API associées.

9. Vérifiez les nœuds sélectionnés, le sous-réseau dorsal et les paramètres de l'interface du service Data Engine.
10. Sélectionnez **Ajouter** et attendez la fin de l'opération. System Manager effectuera les actions suivantes :
 - Ajouter les nœuds sélectionnés au cluster DCN
 - Provisionnez le réseau dorsal et joignez les nœuds au cluster DCN basé sur Kubernetes
 - Mise à jour des métadonnées internes pour la découverte DCN
11. Une fois terminé, sélectionnez **Cluster** et **Overview** et confirmez :
 - a. Sous **Data compute**, les nouveaux DCN apparaissent comme faisant partie du cluster

- b. Tous les nœuds sont **Healthy**
- c. Vérifiez que la carte **Santé** du tableau de bord affiche le nombre de nœuds mis à jour

Remplacez un nœud de votre cluster AIDE

Vous devez remplacer un nœud de calcul de données (DCN) dans votre AI Data Engine (AIDE) cluster s'il cesse de fonctionner ou s'il doit être remplacé en raison d'une défaillance matérielle, d'une mise à niveau ou d'une opération de maintenance. Cela garantit que le cluster AIDE reste sain et opérationnel. La procédure peut être effectuée sans perturber les services en cours.

Préparez à remplacer un nœud

Plusieurs éléments sont à prendre en compte avant de remplacer un nœud dans votre AIDE cluster.

Identifiants System Manager

Vous devez disposer des privilèges d'administrateur de stockage pour effectuer les tâches de remplacement de nœud de cluster AIDE.

Restrictions

Vous devez être conscient des restrictions suivantes lors du remplacement d'un nœud dans votre AIDE cluster :

- Le remplacement de nœud n'est pris en charge qu'à l'aide du CLI et, en option, de l'API REST.
- Vous ne pouvez pas effectuer de remplacement de nœud à l'aide de System Manager.
- Le nouveau nœud doit correspondre à la version logicielle du cluster's ; ONTAP le mettra à jour si nécessaire.
- Le nœud défaillant ne doit pas être mis sous tension lorsqu'il est connecté au réseau du cluster afin d'éviter les conflits d'adresses IP.

Exigences

Vous aurez besoin des éléments suivants :

- Numéro de série du nouveau nœud de remplacement

Remplacez un nœud DCN dans votre cluster AIDE

Vous pouvez remplacer un nœud DCN dans votre cluster AIDE en utilisant la procédure suivante.

Étapes

1. Retirez physiquement le nœud défaillant

Mettez le nœud hors tension et déconnectez-le du réseau du cluster. Assurez-vous que le nœud ne soit pas démarré sur le réseau pendant la procédure de remplacement.

2. Supprimez le nœud défaillant du cluster à l'aide de la commande suivante :

```
dcn cluster node delete -name <node_name> -force true
```

Indiquez le nom exact pour la valeur <node_name>.

3. Connectez physiquement le nouveau nœud au cluster

Assurez-vous que le nœud est connecté par câble, sous tension et détectable.

4. Consultez les nœuds détectables et non configurés pour vérifier que le nouveau nœud est en ligne :

```
dcn cluster node show -membership available
```

5. Ajoutez le nœud au cluster à l'aide de la commande suivante :

```
dcn cluster node create -serial-number <new_node_serial>
```

ONTAP attribuera une adresse IP au nouveau nœud. Si la version logicielle du nœud ne correspond pas à celle du cluster, ONTAP mettra automatiquement le nœud à jour.

6. Vérifiez l'état du cluster et l'intégration des nœuds à l'aide de l'une des commandes suivantes :

```
dcn cluster node show
```

```
dcn cluster node show -instance
```

Informations connexes

- ["Développez votre cluster de calcul"](#)

Gérer la vectorisation et les collections de données

Démarrage rapide Data-to-RAG pour AI Data Engine

Passez d'un système AI Data Engine (AIDE) nouvellement déployé à un point de terminaison retrieval-augmented generation (RAG) opérationnel grâce à ce flux de travail. Comprenez comment les administrateurs de stockage, les ingénieurs de données et les data scientists collaborent à l'aide de ONTAP System Manager et AIDE Console.

Avant de commencer

- Vous avez installé et ajouté des nœuds de calcul de données (DCN) au cluster ONTAP.
- Vous avez installé et licencié AI Data Engine software pour la vectorisation et les garde-fous.
- Vous avez configuré "[OpenID Connect \(OIDC\)](#)" et attribué des rôles aux fonctions d'administrateur, d'ingénieur de données et de scientifique des données.

1

Définir le périmètre et la gouvernance des données

En tant qu'administrateur de stockage ou administrateur de sécurité, vous souhaitez préparer l'environnement dans AIDE Console et ONTAP System Manager :

- "[Créez un ou plusieurs espaces de travail](#)" à partir de sources de données locales et distantes.
- "[Configurer les classificateurs et les politiques de guardrail](#)" dans AIDE Console.
- "[Attribuer l'accès aux espaces de travail aux data engineers et data scientists](#)".

2

Explorer les métadonnées de l'espace de travail

En tant qu'ingénieur des données ou data scientist, vous souhaitez explorer les métadonnées de l'espace de travail à l'aide de AI Data Engine Console :

- "[Explorer les métadonnées de l'espace de travail](#)" pour comprendre le contenu disponible.
- Définissez un ou plusieurs sous-ensembles logiques de données qui devraient alimenter RAG (par exemple, des articles de support, des manuels de produits ou des notes cliniques anonymisées).

3

Créer et publier une collecte de données

En tant qu'ingénieur des données ou data scientist, vous souhaitez transformer le sous-ensemble choisi en une collection prête pour RAG :

- "[Créer une collecte de données](#)" à partir de l'espace de travail en utilisant les filtres sélectionnés.
- "[Publier la collecte de données](#)" et surveillez l'indexation jusqu'à ce qu'elle atteigne l'état Ready.
- Copiez l'URI du point de terminaison de récupération pour la collection choisie et fournissez-la aux data scientists ou aux développeurs d'applications.
- "[Afficher l'état de la collecte de données et l'encombrement du vecteur](#)" au besoin.

Et ensuite ?

- ["Définissez votre patrimoine de données et vos politiques de garde-fous dans AI Data Engine"](#)
- ["Explorez les métadonnées de l'espace de travail dans AI Data Engine Console"](#)
- ["Créez des collectes de données dans AI Data Engine Console"](#)

Explorez les métadonnées de l'espace de travail dans AI Data Engine Console

En tant qu'ingénieur de données ou data scientist, votre première tâche dans AI Data Engine (AIDE) est de comprendre quelles données sont disponibles dans vos espaces de travail. Vous utilisez AI Data Engine Console pour interroger le catalogue de métadonnées, rechercher les fichiers pertinents et identifier les sous-ensembles de données à transformer en collecte de données.

Avant de commencer

- Vous devez disposer des privilèges de *data engineer* ou de *data scientist* dans AI Data Engine Console et d'un accès à au moins un workspace.
- Un administrateur de stockage a :
 - Créé un ou plusieurs espaces de travail dans ONTAP System Manager.
 - Attribuez à votre utilisateur ou groupe l'accès aux espaces de travail concernés.
- L'extraction des métadonnées de l'espace de travail est terminée et l'espace de travail est dans l'état Ready.
- Les classificateurs sont activés, de sorte que les métadonnées incluent des balises de classification (par exemple, des indicateurs PII).

Sign in à AI Data Engine Console en tant qu'ingénieur de données ou data scientist

Étapes

1. Dans un navigateur, accédez à l'URL de la AI Data Engine Console :

```
https://<cluster_management_ip>/console
```

2. Authentifiez-vous via le fournisseur OIDC de votre organisation.
3. Vérifiez que votre rôle est bien reconnu comme étant celui d'un data engineer ou d'un data scientist (par exemple, par l'espace de travail disponible et les actions de collecte de données). Pour plus d'informations, consultez ["Documentation relative au rôle d'AIDE"](#) pour savoir comment les data engineers et les data scientists travaillent avec les composants AIDE.

Résultat

Vous êtes connecté à AI Data Engine Console et vous ne voyez que les espaces de travail auxquels vous avez été autorisé à accéder.

Consultez vos espaces de travail accessibles

Étapes

1. Dans AI Data Engine Console, accédez à **Data Curator > Workspaces**.

2. Consultez la liste des espaces de travail auxquels vous avez accès.
3. Sélectionnez un espace de travail pour ouvrir ses détails.

Résultat

Vous disposez désormais d'une vue limitée à votre espace de travail de l'ensemble des données que les administrateurs de stockage ont mises à disposition pour vos projets.

Et ensuite ?

- ["Créer des collectes de données pour RAG à partir d'un workspace"](#)

Créez des collectes de données dans AI Data Engine Console

Les collections de données constituent les éléments de base RAG dans AI Data Engine (AIDE). En tant qu'ingénieur ou scientifique des données, vous définissez les fichiers qui doivent figurer dans une collection, configurez les options d'intégration et d'indexation, et publiez la collection afin que les applications puissent l'interroger via un point de terminaison de récupération.

Vous effectuerez toutes les tâches de collecte de données dans la AI Data Engine Console.

Avant de commencer

- Vous devez disposer des privilèges d'*data engineer* ou d'*data scientist* dans AI Data Engine Console (https://<cluster_management_ip>/console).
- Vous avez accès à au moins un espace de travail avec des métadonnées extraites et à l'état Ready.
- Vous avez exploré les métadonnées de l'espace de travail et identifié des requêtes ou des filtres qui définissent des sous-ensembles de données significatifs.
- La licence AI Data Engine software est installée et les fonctionnalités d'inférence sont activées.

Créer une collecte de données à partir des métadonnées de workspace

Étapes

1. Accédez à **Data Curator > Workspaces** et sélectionnez l'espace de travail qui contient vos données cibles.
2. Sélectionnez **Add data collection**.
3. Sur la page **Créer une nouvelle collecte de données**, procédez comme suit :
 - a. Saisissez un nom et une description pour la collecte (par exemple, `Support_KB_RAG_EN`).
 - b. Choisissez si la collecte de données doit être :
 - **Dynamique** : Les nouveaux fichiers sont automatiquement identifiés et ajoutés à la collecte de données en fonction des critères de filtrage que vous définissez. Cela se produit lors des actualisations de l'espace de travail.
 - **Statique** : Vous choisissez les fichiers à inclure dans la collecte de données. Vous pouvez modifier les fichiers si la collecte de données est en `draft` état. Après le passage de la collecte de données à l'`Published` état, elle ne peut plus être modifiée.

4. Spécifiez le sous-ensemble source :

- a. Utilisez des mots-clés et des filtres (type de fichier, horodatages et autres attributs) pour trouver les fichiers pertinents à inclure.



Vous pouvez sélectionner un nom de fichier pour ouvrir une fenêtre d'aperçu du contenu.

5. Ajoutez ces fichiers à la collecte de données.

6. Sélectionnez **Enregistrer** pour finaliser la collecte.

Résultat

Vous avez défini le périmètre de la collecte de données et y avez ajouté les fichiers requis. AIDE génère les embeddings et construit l'index vectoriel lorsque vous publiez la collecte.



Créez des collections restreintes et ciblées (par exemple, par cas d'utilisation ou domaine) plutôt qu'une seule collection "everything". Cela améliore la pertinence de la récupération et la gestion.

Publier une collecte de données

Publiez la collecte de données pour la rendre interrogeable par les applications d'IA via un point d'accès de récupération RAG. La publication génère des vector embeddings à partir des fichiers sélectionnés et les indexe pour la recherche sémantique. Après que la collection ait atteint l'état *Ready*, son point d'accès devient disponible pour que les data scientists puissent l'intégrer dans des notebooks, des pipelines et des applications d'IA pour la retrieval-augmented generation (RAG) et la recherche.



Pour les grandes collections, envisagez de programmer la publication initiale et les principales republications pendant les heures creuses afin de minimiser la concurrence pour les ressources.

Étapes

1. Accédez à **Data Curator > Data collections** et sélectionnez le menu des options (**...**) pour votre collecte de données.
2. Sélectionnez **Publier**.
3. Sélectionnez une configuration d'optimisation par défaut ou personnalisée.
4. Sélectionnez **Publier** pour lancer la transformation des données.
5. Dans AIDE Console, ouvrez la vue détaillée de la collecte (**Data Curator > Data collections**) pour les mises à jour de statut.

Résultat

La collecte atteint l'`Ready` état et est disponible pour une utilisation par les applications en aval et les data scientists.

Dans **Data Curator > Data collections**, vous pouvez sélectionner **Copy URI** pour obtenir les informations nécessaires pour accéder à la collecte de données à l'aide d'une API.

Mettre à jour ou supprimer une collecte de données

Au fil du temps, vous pourriez avoir besoin d'affiner ou de retirer des collections de données. L'affinage d'une

collection peut impliquer l'ajustement des filtres pour ajouter ou supprimer des fichiers, la modification des paramètres d'intégration ou la mise à jour de la description de la collection. La suppression d'une collection la retire définitivement et rend son point de terminaison de récupération indisponible.

Mettre à jour une collecte de données

Vous pouvez mettre à jour une collecte de données lorsqu'elle est dans `draft` un état donné.

Étapes

1. Accédez à **Data Curator > collecte de données**.
2. Sélectionnez la collection que vous souhaitez modifier.
3. Choisissez **Modifier**.
4. Modifiez l'un des éléments suivants :
 - Nom et description
 - Filtres (chemins, types de fichiers, classification tags).
 - Paramètres d'intégration et de segmentation.
5. Enregistrez vos modifications.
6. Republiez la collection afin que la nouvelle définition et les nouveaux embeddings prennent effet.

Résultat

Une nouvelle tâche d'indexation s'exécute avec la configuration mise à jour, et la collection revient à un `Ready` état une fois terminée.

Supprimer une collection

La suppression d'une collection est définitive. Assurez-vous qu'aucune application de production ne dépende encore du point de terminaison de récupération de la collection avant de la supprimer.

Étapes

1. Accédez à **Data Curator > Data collections**, et sélectionnez le menu des options (`...`) pour la collection.
2. Choisissez **Delete**.
3. Confirmez la suppression.

Résultat

La définition de la collection et ses représentations vectorielles ont été supprimées d'AI Data Engine. Les applications tentant d'interroger l'ancien point de terminaison de récupération échoueront après la suppression de la collection.

Et ensuite ?

- ["Afficher les collectes de données"](#)

Consultez les collections de données dans AI Data Engine

Après que les ingénieurs de données ou les data scientists créent et publient des collectes de données à partir des espaces de travail, vous avez besoin de visibilité sur leur statut, leur taille et leur impact sur le cluster AI Data Engine.

Si vous êtes administrateur de stockage, ingénieur de données ou data scientist, vous pouvez consulter les data collections via ONTAP System Manager et AIDE Console.

Avant de commencer

- Vous devez disposer des privilèges d'administrateur de stockage dans ONTAP System Manager ou des privilèges d'ingénieur de données ou de scientifique des données dans AI Data Engine Console (https://<cluster_management_ip>/console pour visualiser les collectes de données).
- Au moins un espace de travail existe avec des métadonnées extraites avec succès.
- Les ingénieurs ou scientifiques des données ont créé et publié au moins une collecte de données à partir de AI Data Engine Console.
- La licence AI Data Engine software est installée et les fonctionnalités d'inférence sont activées, de sorte que les points de terminaison de vectorisation et de récupération sont actifs.

Afficher les collectes de données à l'échelle du cluster

Pour les administrateurs de stockage, ONTAP System Manager offre une vue d'ensemble des collections de données et de leur encombrement mais ne permet pas aux administrateurs de les créer ou de les modifier.

Étapes

1. Dans System Manager, accédez à **Data Engine > Data collections**.
2. Consultez le récapitulatif de l'inventaire en haut de la page :
 - Nombre total de collecte de données par statut
 - Espace total occupé par la base de données vectorielle dans toutes les collections
 - Espace vectoriel en pourcentage de la capacité globale du cluster
3. Sélectionnez une collecte de données individuelle et examinez-la :
 - Nom et description de la collection
 - UUID
 - Espace de travail associé
 - Statut
 - Taille de la collection
 - Créateur
 - Dernière actualisation

Résultat

Vous disposez désormais d'une vue d'ensemble de toutes les collections de données du cluster et de leur impact sur le stockage. Utilisez cette vue pour identifier les collections volumineuses, obsolètes ou bloquées dans un état non prêt.

Vous pouvez également voir si une collecte de données est activement mise à jour et si des défaillances bloquent l'utilisation de RAG.

Surveiller les tâches et les événements liés à la collecte

En tant qu'administrateur de stockage, vous pouvez surveiller les tâches qui créent et mettent à jour les collections à partir de la page **Activity** à l'échelle du cluster et à partir des détails de l'espace de travail.

Étapes

1. Dans System Manager, accédez à **Data Engine > Activity**.
2. Dans l'onglet **Événements** :
 - a. Filtrer par type (par exemple, espace de travail, collecte de données) ou gravité.
 - b. Développez n'importe quel événement lié à la collecte de données (par exemple, « Data collection publish failed ») pour voir plus de détails.
3. Dans l'onglet **Jobs** :
 - a. Filtrer pour se concentrer sur l'indexation de la collecte de données et les tâches de publication.
 - b. Pour chaque tâche, ouvrez l'aperçu pour voir :
 - Pourcentage de progression.
 - Heures de début et de fin.
 - Tous les messages d'erreur ou avertissements signalés.
4. Vous pouvez également revenir à l'espace de travail concerné (**Data Engine > Workspaces**) et ouvrir son onglet **Activité** pour voir les événements et les tâches limités à cet espace de travail.

Résultat

Vous pouvez suivre le cycle de vie des collectes de données, identifier les tâches bloquées ou ayant échoué et recueillir des informations contextuelles à transmettre aux data engineers, data scientists ou au support.



Lorsqu'une collecte de données reste dans l'état `Publishing` pendant une période prolongée, vérifiez s'il existe une tâche de longue durée correspondante sur la page Activité avant de conclure à une panne.

Afficher les collectes de données depuis la console AIDE

Les ingénieurs et les scientifiques des données surveillent généralement les collectes de données directement depuis AI Data Engine Console, où elles sont créées et publiées.

Étapes

1. Connectez-vous à AI Data Engine Console en tant qu'ingénieur de données ou scientifique des données.
2. Accédez à **Data Collections** et sélectionnez la collecte de données souhaitée.
3. Pour chaque collection :
 - a. Vérifiez l'état (`Draft`, `Publishing`, `Ready`, ou `Failed`).
 - b. Sélectionnez le nom de la collecte de données pour consulter les détails de la définition (filtres, types de fichiers inclus, options du classificateur, paramètres d'intégration).
 - c. Vérifiez les horodatages de la dernière publication ou de la dernière mise à jour.
4. Si nécessaire, ouvrez les détails de la tâche ou les journaux (lorsqu'ils sont disponibles) pour comprendre les échecs ou les exécutions incomplètes.

Résultat

Les ingénieurs et les scientifiques des données peuvent itérer sur les définitions de collections et les republier tout en surveillant leur état et leur santé, sans impliquer les administrateurs de stockage.

Et ensuite ?

- ["Créer des collections de données pour RAG dans AI Data Engine Console"](#)

Mettre en place des garde-fous

Définissez vos politiques de Data Guardrails dans AI Data Engine pour votre domaine de données

En tant que propriétaire de données ou de plateforme, vous utilisez AI Data Engine Console pour définir quelles données sont concernées par l'IA, quelles données sont toujours interdites et quelles règles de sécurité s'appliquent lorsque ces données sont utilisées pour la classification et la génération augmentée par récupération (RAG).

Utilisez ces procédures pour définir ces politiques dans AI Data Engine Console afin que ONTAP System Manager puisse les appliquer à toutes les données des espaces de travail.

Avant de commencer

- Vous avez besoin de privilèges d'administrateur de stockage dans AI Data Engine Console (https://<cluster_management_ip>/console pour créer et gérer des politiques globales.
- Vous disposez d'un AIDE cluster avec des nœuds de calcul de données déployés et opérationnels.
- "OpenID Connect (OIDC)" est configuré, et votre rôle IdP est associé à un rôle d'administrateur AIDE qui permet la gestion des politiques de données.
- La licence AI Data Engine software est installée afin que les fonctionnalités de garde-fou et d'inférence soient activées.
- Au moins un espace de travail existe, ou vous vous êtes coordonné avec l'administrateur pour comprendre quelles sources de données (volumes) seront utilisées dans les espaces de travail.

Comprendre les types de stratégies

AIDE Console expose ces types de politiques qui structurent votre patrimoine de données :

- **Classificateurs** : Activez les classificateurs pour détecter les PII, les problèmes de sécurité ou d'autres modèles dans tous les espaces de travail.
- **Catégories de classificateurs** : Regroupez les classificateurs en catégories de conformité pour l'organisation et la gestion.
- **Politiques de Data Guardrails** : règles de sécurité et de rédaction appliquées au moment de la récupération ou de l'inférence.

Vous ne pouvez pas utiliser ONTAP System Manager pour créer ou gérer ces guardrail policies. Il ne fait que les lire et les appliquer lorsqu'un administrateur de stockage les applique aux espaces de travail. Toute la définition et la maintenance des politiques s'effectuent dans AIDE Console.

Activer les classificateurs

Les classificateurs analysent les métadonnées et le contenu pour annoter les fichiers et les objets (par exemple, en détectant les PII ou les catégories sensibles). Avant que les classificateurs puissent s'exécuter sur les données de l'espace de travail, vous devez les activer dans AI Data Engine Console.

À propos de cette tâche

Le comportement des classificateurs est contrôlé globalement dans AIDE Console. Tous les classificateurs activés s'exécutent sur chaque espace de travail. Étant donné qu'ils sont appliqués globalement, ils ne

peuvent pas être activés ou désactivés pour un espace de travail individuel. Ils peuvent être activés ou désactivés uniquement globalement.

Étapes

1. Dans AI Data Engine Console, accédez à **Data Guardrails > Classifieurs**.
2. Sélectionnez une catégorie de classificateur pour afficher les classificateurs qu'elle contient.
3. Cochez les cases correspondant aux classificateurs que vous souhaitez activer, ou sélectionnez toutes les lignes pour activer les classificateurs en masse.
4. Sélectionnez **Enable**.



Utilisez l'option de sélection groupée pour activer plusieurs classificateurs simultanément. Chaque fois que vous activez un classificateur, une actualisation de l'ensemble des espaces de travail est déclenchée. Pour minimiser les actualisations inutiles, activez plusieurs classificateurs à la fois plutôt qu'un par un.

Résultat

Tous les espaces de travail nouvellement créés et existants exécutent les classificateurs activés lors du traitement des métadonnées.

Les étiquettes de classification sont inscrites dans le catalogue de métadonnées et deviennent disponibles pour les ingénieurs de données pour le filtrage lors de la création de collections de données.

Gérer les catégories de classificateur

Les classificateurs sont organisés en catégories (telles que « PII » ou « Financial data »). Les catégories permettent de regrouper les classificateurs apparentés pour une gestion simplifiée et une meilleure visibilité de la conformité. Vous pouvez utiliser les catégories par défaut fournies par AIDE ou créer des catégories personnalisées pour répondre à vos exigences de conformité.

Étapes

1. Dans AI Data Engine Console, accédez à **Data Guardrails > Classifieurs**.
2. Consultez les catégories de classification existantes. Il existe deux grandes catégories de classification :
 - Contenu ou données : détecte des types de données particuliers dans les fichiers.
 - Document : Classifie le type de document en fonction du contenu.
3. Déterminez si les sous-catégories de classificateur par défaut sont suffisantes ou si vous souhaitez créer votre propre sous-catégorie.
 - Si vous utilisez une sous-catégorie de classificateur par défaut (par exemple, **General Privacy**) :
 - i. Sélectionnez le nom de la catégorie dans Classifier categories pour afficher les classificateurs associés.
 - ii. Examinez la liste des classificateurs.
 - iii. Sélectionnez **Ajouter** pour rechercher et ajouter des classificateurs non répertoriés à partir de la liste complète des classificateurs disponibles.
 - Si vous souhaitez créer une catégorie personnalisée, sélectionnez **+ Add** .
 - i. Ajoutez un nom unique, une description et attribuez les classificateurs disponibles à la catégorie.
 - ii. Sélectionnez **Add**
4. Pour désactiver un classificateur au sein d'une catégorie, sélectionnez **...** pour le classificateur et

choisissez **Désactiver**. Vous pouvez également sélectionner toutes les lignes pour modifier leur état en une seule opération.

Résultat

Les catégories organisent les classificateurs pour la visibilité de la conformité. Les ingénieurs de données peuvent utiliser les étiquettes de classification lors du filtrage et de la création de collectes de données.

Créer et gérer des politiques de Data Guardrails

Les politiques de Data Guardrails déterminent comment AIDE réagit lorsque les classificateurs détectent du contenu sensible ou lorsque les invites et les résultats de récupération violent les règles de contenu.

Les comportements typiques des Data Guardrails comprennent :

- Masquage ou suppression des PII dans les extraits récupérés.
- Blocage des réponses qui enfreignent les règles de conformité.
- Enregistrement ou étiquetage des violations à des fins d'audit.

À propos de cette tâche

Vous créez et gérez les politiques de garde-fous uniquement dans AI Data Engine Console.

Vous ne pouvez associer des espaces de travail dans ONTAP System Manager qu'à une seule stratégie de Data Guardrails à la fois.

Étapes

1. Dans AI Data Engine Console, accédez à **Data Guardrails > Guardrail policies**.
2. Sélectionnez **Add**.
3. Saisissez un nom et une description qui décrivent clairement la portée (par exemple, `Customer PII redaction for support KB`).
4. Configurer les conditions requises pour l'activation du garde-fou pilotées par le classificateur de données :
 - a. Définissez les conditions d'activation du garde-corps :
 - i. Choisissez la catégorie de classificateur ou le type de classificateur pour chaque condition.
 - ii. Ajoutez et définissez des conditions supplémentaires au besoin.
 - iii. Définissez des critères de recherche spécifiques dans **Search**, puis sélectionnez **Accept**.
 - b. Définissez les actions de la politique de Data Guardrails, telles que l'anonymisation du contenu ou le blocage et la suppression d'un fichier d'une collecte de données.
5. Sélectionnez l'espace de travail auquel le guardrail sera appliqué.
6. Définissez l'état de la politique :
 - **Activé** : Active la stratégie immédiatement.
 - **Mode test** : Permet de valider l'impact de la policy avant de l'activer.
 - **Désactivé** : Enregistre le Data Guardrails sans l'appliquer.
7. Sélectionnez **Add** pour enregistrer la stratégie et l'appliquer à l'espace de travail.



Utilisez le **mode test** avec un espace de travail pilote et une collecte de données non productive pour comprendre combien de réponses seraient affectées avant d'activer l'application stricte.

Résultat

La nouvelle politique de Data Guardrails est active et limitée à l'espace de travail sélectionné.

Comment les politiques interagissent avec les espaces de travail

Une fois les politiques définies :

- L'administrateur du stockage utilise ONTAP System Manager pour créer des espaces de travail, sélectionner des conteneurs de données et associer une politique de Data Guardrails.
- Les classificateurs s'exécutent automatiquement sur le contenu de l'espace de travail en fonction de ce que vous avez activé.
- Les Data Guardrails associés à l'espace de travail influencent le comportement des points de terminaison de récupération.

Pour les ingénieurs et les scientifiques des données :

- L'ensemble des données visibles (espaces de travail et collections de données) est déjà filtré par attribution de rôle.
- Les métadonnées que vous interrogez (par exemple, les balises PII) sont déterminées par les classificateurs activés.
- Les réponses reçues par vos pipelines RAG sont limitées par les Data Guardrails configurés au niveau de l'espace de travail.

Informations connexes

- ["Consultez les collections de données dans AI Data Engine"](#)
- ["Démarrage rapide Data-to-RAG pour AI Data Engine"](#)

FAQ pour NetApp AI Data Engine

Cette FAQ couvre les questions courantes concernant NetApp AI Data Engine (AIDE), y compris son architecture, son déploiement, les types d'utilisateurs, ses fonctionnalités techniques, son intégration et ses licences.

Principes de base d'AIDE

Qu'est-ce que NetApp AI Data Engine (AIDE) ?

NetApp AI Data Engine (AIDE) est un service de données d'IA intégré au stockage qui couvre l'intégralité du cycle de vie de l'IA, de la découverte et la préparation des données brutes à la fourniture de points d'accès pour alimenter l'IA générative (GenAI), la génération augmentée par la recherche (RAG), l'IA agentique et les fabriques d'IA. AIDE automatise la synchronisation et la détection des modifications, offrant une vue unifiée et actualisée des données sélectionnées pour la découverte et la curation des données.

Comment fonctionne AIDE ?

AIDE s'intègre directement aux systèmes de stockage NetApp ONTAP pour créer une vue globale et structurée de l'ensemble du patrimoine de données NetApp avec détection automatisée des modifications et synchronisation. AIDE offre une vectorisation en temps réel avec compression et déduplication, des garde-fous pilotés par des politiques et une intégration avec des outils d'IA.

Utilisateurs et rôles

Qui utilise l'AI Data Engine ?

Les principaux utilisateurs d'AIDE incluent :

1. **ONTAP storage administrators** : Gèrent l'infrastructure, les besoins de stockage spécifiques à l'IA, la sécurité et la conformité.
2. **Ingénieurs de données** : Gèrent le déplacement, la préparation et l'intégration des données entre les environnements.
3. **Scientifiques des données**: Préparez et transformez les données pertinentes pour la consommation par l'IA.

Exigences et déploiement

Quel matériel est requis?

AIDE nécessite des systèmes AFX pour le déploiement (y compris un contrôleur AFX, un tiroir disque et un commutateur réseau), mais peut utiliser les données de cluster provenant de clusters exécutant ONTAP 9 en utilisant SnapMirror et l'appairage de clusters. Au moins quatre nœuds de contrôleur AFX sont requis pour les déploiements AIDE afin de garantir une haute disponibilité et des performances.

AIDE s'exécute sur un nœud de calcul de données NetApp (DCN). Trois DCN sont nécessaires. Le DCN héberge le logiciel AIDE, qui comprend le Metadata Engine, Data Sync, Data Curator et Data Guardrails.

Puis-je utiliser mon propre DCN ?

Non. Le DCN est un nœud matériel de calcul de données fourni par NetApp et constitue le seul mécanisme de déploiement pour le AI Data Engine.

Quel est le nombre minimum de DCN requis ?

Exactement trois DCN sont requis.

Quel système d'exploitation fonctionne sur les DCN ?

Les DCN exécutent une pile logicielle fournie par NetApp avec AIDE.

AIDE peut-il être déployé sans AFX ?

Non. AIDE nécessite AFX pour le déploiement. AIDE utilise Trident pour consommer les volumes AFX comme stockage interne (volumes persistants). Le cluster AFX fournissant le stockage pour AIDE peut être apparié avec un système ou un cluster ONTAP 9. Il utilise l'appariement de clusters et SnapMirror pour synchroniser les données du cluster ONTAP distant vers le système AFX.

Gestion et interfaces

La console AIDE fait-elle partie de la console NetApp ou s'agit-il d'une interface séparée ?

La AIDE Console est une interface de gestion distincte qui s'exécute sur les DCN. Vous utilisez la AIDE Console pour gérer les services AIDE, tels que Data Guardrails et Data Curator. Vous pouvez également utiliser ONTAP System Manager pour surveiller le cluster AIDE.

Fonctionnalités et capacités

Quelles sont les principales caractéristiques d'AIDE ?

Il existe quatre fonctionnalités principales d'AIDE :

Metadata Engine

- Génère automatiquement une vue structurée, à jour et interactive de vos données.
- Fonctionne avec les données stockées sur ONTAP.
- Permet aux spécialistes des données de collaborer avec les administrateurs de stockage pour trouver et comprendre des données.
- Les API interrogent les métadonnées pour fournir des fonctionnalités tout en réduisant la charge de trafic NFS sur les systèmes de stockage.
- La capacité d'extraction et de catalogage des métadonnées est conçue spécifiquement pour AIDE et fonctionne en continu, en tirant parti des capacités d'ONTAP telles que les snapshots.

Synchronisation des données

- Maintient automatiquement la fraîcheur des données lorsque les données sources changent, sans intervention manuelle.
- Les administrateurs définissent l'intervalle d'actualisation des données en jours ou en heures.
- Fournit une mobilité des données incrémentielle et une synchronisation sur l'ensemble des données afin d'éliminer les copies redondantes des données d'IA.

Data Guardrails

- Identifie et protège automatiquement les données sensibles tout au long du cycle de vie de l'IA. Il est accessible via AI Data Engine Console.
- Analyse, classe et catégorise les données en continu.
- Identifie les données sensibles (telles que les PII) et les risques.
- Facilite la création de politiques pour le traitement automatique des données sensibles, conformément aux normes de l'entreprise et aux réglementations.

- Assure la suppression automatique des informations sensibles pour la protection des données.
- Restreint l'accès aux fichiers sensibles selon les besoins.

Data Curator

- Permet aux data scientists de rechercher des données pertinentes dans le stockage.
- Crée des collections de données organisées avec des données existant sur des volumes AFX.
- Génère des représentations vectorielles au niveau du stockage pour réduire la taille des données et augmenter les performances.
- Fournit un point de terminaison de récupération pour les applications d'IA avec recherche sémantique vectorielle et réordonnancement.

Intégration et interopérabilité

AIDE prend-il en charge les métadonnées fédérées sur plusieurs clusters ONTAP ?

AIDE peut se connecter à plusieurs clusters ONTAP en utilisant SnapMirror et le peering de clusters, permettant une visibilité centralisée des métadonnées.

Où sont stockées les métadonnées ?

AIDE stocke les métadonnées sur le cluster AFX connecté à l'aide d'un volume persistant fourni par AFX. Les DCNs utilisent un stockage local pour les opérations internes.

Le AIDE Metadata Engine classe-t-il les données ?

Non. Le Metadata Engine catalogue les métadonnées du système de fichiers et fournit des API pour interroger ces métadonnées cataloguées.

Quelles sources de données sont prises en charge ?

AIDE prend en charge les volumes ONTAP (locaux ou distants) comme sources de données. Les clusters ONTAP distants doivent exécuter ONTAP 9 et être connectés via l'appairage de cluster et SnapMirror.

Les compartiments ONTAP S3 et les objets StorageGRID ne sont pas pris en charge comme sources de données dans AIDE 9.18.1.

Quels types de fichiers AIDE peut-il traiter pour la classification, la vectorisation et la recherche sémantique ?

AIDE prend en charge un large éventail de types de fichiers, notamment PDF, DOCX, PPTX, TXT et les fichiers image avec des capacités OCR.

AIDE prend-il en charge la classification des données non anglaises ?

AIDE ne prend en charge que les données en anglais.

Quelles intégrations AIDE prend-il en charge ?

AIDE fournit un point de terminaison d'API RAG accessible par appels d'API directs ou via un serveur Model Context Protocol (MCP). Cela prend en charge l'intégration avec les frameworks et outils d'IA agentique.

Déploiement et licences

Quelles sont les options de déploiement ?

AIDE est déployé sur site sur une infrastructure AFX avec des DCN. Il s'intègre directement aux installations NetApp ONTAP AFX.

Comment AIDE est-il licencié ?

AIDE nécessite une licence logicielle pour exécuter Data Guardrails et Data Curator.

Si vous n'avez besoin que du Metadata Engine, la licence ONTAP One, incluse avec tous les systèmes AFX, vous donne droit aux fonctionnalités réservées au Metadata Engine.

Informations connexes

- ["Installez les licences AIDE dans ONTAP System Manager"](#)
- ["Découvrez l'architecture et les composants d'AIDE"](#)

Mentions légales

Les mentions légales donnent accès aux déclarations de droits d’auteur, aux marques de commerce, aux brevets et plus encore.

Copyright

["https://www.netapp.com/company/legal/copyright/"](https://www.netapp.com/company/legal/copyright/)

Marques déposées

NETAPP, le logo NETAPP et les marques figurant sur la page NetApp Trademarks sont des marques de NetApp, Inc. Les autres noms de sociétés et de produits peuvent être des marques de leurs propriétaires respectifs.

["https://www.netapp.com/company/legal/trademarks/"](https://www.netapp.com/company/legal/trademarks/)

Brevets

Une liste actuelle des brevets détenus par NetApp peut être trouvée à :

<https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf>

Politique de confidentialité

["https://www.netapp.com/company/legal/privacy-policy/"](https://www.netapp.com/company/legal/privacy-policy/)

Open source

Les fichiers de notification fournissent des informations sur les droits d’auteur et les licences de tiers utilisés dans les logiciels NetApp.

AI Data Engine

["Avis pour AIDE 9.18.1"](#)

Informations sur le copyright

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

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

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

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

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

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

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

Informations sur les marques commerciales

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