



AI Data Engine Dokumentation

AI Data Engine

NetApp
March 13, 2026

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

Inhalt

AI Data Engine Dokumentation	1
Versionshinweise	2
Was ist neu in AI Data Engine	2
Was ist neu in der ersten Version von AIDE 9.18.1	2
Bekannte Einschränkungen der AI Data Engine	3
Bekannte Einschränkungen für die anfängliche Veröffentlichung von AIDE 9.18.1	3
Los geht's	6
Erfahren Sie mehr über Ihr AI Data Engine System	6
Erfahren Sie mehr über AI Data Engine	6
AI Data Engine Architektur	7
AIDE-Komponenten und Verantwortlichkeiten nach Rolle	11
Schnellstart für AI Data Engine	17
Installieren Sie AIDE	18
Voraussetzungen für die Installation von AI Data Engine	18
Installieren Sie das AFX-Speichersystem für AI Data Engine	19
Installieren Sie Ihre Datenrechenknoten	19
Richten Sie Ihr AIDE System ein	29
AI Data Engine einrichten	29
Installieren Sie eine AIDE-Lizenz in Ihrem AFX-System	32
OpenID Connect für AIDE in ONTAP konfigurieren	33
Arbeitsbereiche einrichten	36
Daten für AI Data Engine vorbereiten	36
Erstellen Sie einen Arbeitsbereich in AI Data Engine	37
Einen Arbeitsbereich erstellen	38
Arbeitsbereichsdetails überprüfen	38
Aktualisierungen und Versionierung von Arbeitsbereichen	39
Benutzerzugriff auf AI Data Engine Arbeitsbereiche zuweisen	39
Verwalten und überwachen	41
Clusterprozesse überwachen	41
AIDE-System- und Clusterstatus anzeigen	41
Insights anzeigen, um Ihr AIDE System zu optimieren	43
AIDE-Systemereignisse, Jobs und Revisionsprotokoll	44
AI Data Engine Arbeitsbereiche verwalten	46
Arbeitsbereichsstatus überprüfen	46
Arbeitsbereichseigenschaften bearbeiten und Aktualisierungsintervall bearbeiten	47
Fügen Sie einem bestehenden Arbeitsbereich Datencontainer hinzu	47
Datencontainer aus einem Arbeitsbereich entfernen	48
Workspace-Benutzer verwalten	48
Einen Arbeitsbereich löschen	48
Aktualisieren und warten Sie Ihr AIDE-System	49
AI Data Engine systemaktualisierungen und Kompatibilität	49
Aktualisieren Sie die AI Data Engine software	51
Fügen Sie Ihrem AIDE-Cluster Datenrechenknoten hinzu	52

Ersetzen Sie einen Knoten in Ihrem AIDE Cluster	54
Vektorisierung und Datensammlungen verwalten	56
Data-to-RAG-Schnellstart für AI Data Engine	56
Erkunden Sie Arbeitsbereichs-Metadaten in der AI Data Engine Console	57
Sign in to AIDE Console als Dateningenieur oder Datenwissenschaftler	57
Ihre zugänglichen Arbeitsbereiche anzeigen	57
Was kommt als Nächstes?	58
Erstellen Sie Datensammlungen in AI Data Engine Console	58
Erstellen einer Datensammlung aus Workspace-Metadaten	58
Veröffentlichen Sie eine Datensammlung	59
Aktualisieren oder Löschen einer Datensammlung	59
Was kommt als Nächstes?	60
Datensammlungen in AI Data Engine anzeigen	60
Clusterweite Datensammlungen anzeigen	61
Überwachen Sie sammlungsbezogene Jobs und Ereignisse	61
Datensammlungen aus AIDE Console anzeigen	62
Was kommt als Nächstes?	63
Leitplanken implementieren	64
Definieren Sie Ihre Data Guardrails-Richtlinien in AI Data Engine für Ihren Datenbestand	64
Richtlinientypen verstehen	64
Klassifikatoren aktivieren	64
Klassifiziererkategorien verwalten	65
Guardrail-Richtlinien erstellen und verwalten	66
Wie Richtlinien mit Workspaces interagieren	67
Verwandte Informationen	67
FAQ für NetApp AI Data Engine	68
AIDE-Grundlagen	68
Benutzer und Rollen	68
Anforderungen und Bereitstellung	68
Verwaltung und Schnittstellen	69
Funktionen und Fähigkeiten	69
Integration und Interoperabilität	70
Bereitstellung und Lizenzierung	70
Rechtliche Hinweise	72
Copyright	72
Marken	72
Patente	72
Datenschutzrichtlinie	72
Open Source	72
AI Data Engine	72

AI Data Engine Dokumentation

Versionshinweise

Was ist neu in AI Data Engine

AI Data Engine (AIDE) 9.18.1 ist die erste Version der NetApp-Plattform für AI Data Engine Datenmanagement. Diese Version führt eine Metadata Engine und Management-Workflows ein, mit denen Organisationen unstrukturierte Daten für AI-Workloads katalogisieren und organisieren können und so die Grundlage für erweiterte Governance- und Vektorisierungsfunktionen schaffen. Erweiterte Governance (Guardrails) und Vektorisierung stehen Kunden mit den entsprechenden AI Data Engine Lizenzen zur Verfügung.

Was ist neu in der ersten Version von AIDE 9.18.1

AIDE 9.18.1 führt die folgenden grundlegenden Funktionen ein:

"Metadata Engine für die KI-Katalogisierung"

Die erste Version beinhaltet eine Metadata Engine, die Dateien und Objekte über ONTAP Cluster hinweg katalogisiert.

Zu den wichtigsten Merkmalen gehören:

- Automatisierte Extraktion von Metadaten (Kern- und erweiterten Attributen, Objekt-Tags) aus lokalen und entfernten ONTAP Volumes auf Peered-Clustern.
- Zentralisierte Abfrage- und Filter-REST-APIs für Anwendungen, die einen globalen Überblick über Unternehmensdaten benötigen.
- Skalierbarer Metadatenpeicher.
- Automatische Metadatenextraktion, die während der Workspace-Erstellung ausgelöst wird.

"Arbeitsbereichsverwaltung"

Arbeitsbereiche ermöglichen die logische Gruppierung von Datenquellen (Volumes) für AI-Projekte.

Die erste Version unterstützt Folgendes:

- Erstellung von Arbeitsbereichen, die sich über lokale und entfernte ONTAP Volumes erstrecken (mithilfe von Cluster Peering).
- Zuordnung von Zugriffskontrollen zu Arbeitsbereichen, Unterstützung von Mehrbenutzer- und mandantenfähigen Umgebungen.
- Automatische Metadatenextraktion und Katalogbefüllung bei der Erstellung des Workspace.

"Data Sync für automatisierte Datenaktualität"

Data Sync hält Metadatenkataloge und Datensammlungen aktuell, wenn sich Quelldaten ändern, ohne manuelle Eingriffe.

Zu den wichtigsten Merkmalen gehören:

- Automatisierte Synchronisierung von Daten aus entfernten oder lokalen ONTAP Clustern mittels richtliniengesteuerter SnapMirror-Replikation.

- Inkrementelle Aktualisierungen, die nur geänderte Daten weitergeben und den Aufwand reduzieren.
- Konfigurierbare Aktualisierungsintervalle pro Arbeitsbereich.
- Überwachung des Synchronisierungsstatus und der Aktivität auf Workspace-Ebene.

"Cluster-Einrichtung und -Verwaltung"

Die erste Version umfasst die folgenden Arbeitsabläufe:

- Erkennung und Hinzufügung von Data Compute Nodes (DCNs) während der Cluster-Einrichtung.
- Erstellung dedizierter Metadatenpeicher-VMs für die Metadata Engine.
- Konfiguration der Data Engine Service-Schnittstellen für den clusterweiten Metadatenzugriff.
- Peering mit anderen ONTAP Clustern, um die Katalogisierung von Metadaten über den gesamten Datenbestand hinweg zu erweitern.

"OpenID Connect (OIDC) Authentifizierung"

- OIDC/OAuth-basierte Authentifizierung für sicheren Zugriff auf ONTAP System Manager und Data Engine Console mit Microsoft Entra ID und Active Directory Federation Services (ADFS).
- Rollenbasierte Zugriffskontrollen für Arbeitsbereichs- und Metadatenverwaltung.

"Erweiterte Datenkuratierungs- und Governance-Funktionen"

Folgende Funktionen stehen Kunden mit den entsprechenden AI Data Engine Lizenzen zur Verfügung:

- **Vektorisierung und RAG:** Erstellung von Datensammlungen, Einbettungen und Abrufendpunkten in der AI Data Engine Console unter Verwendung von Metadaten aus AIDE Workspaces.
- **Leitplankenbasierte Governance:** Definition von Leitplankenrichtlinien in der AI Data Engine Console und Zuordnung dieser Richtlinien zu Arbeitsbereichen im ONTAP System Manager.

Unterstützte Hardware und Plattformen

AI Data Engine 9.18.1 läuft auf ONTAP AI Datenplattform-Clustern, die Folgendes kombinieren:

- AFX 1K Speicherknoten
- NetApp Datenverarbeitungsknoten

Verwandte Informationen

- ["Bekannte Einschränkungen der AI Data Engine"](#)
- ["Erfahren Sie mehr über die Architektur und Komponenten der AI Data Engine"](#)

Bekannte Einschränkungen der AI Data Engine

Bekannte Einschränkungen kennzeichnen Plattformen, Geräte oder Funktionen, die von dieser Produktversion nicht unterstützt werden oder nicht ordnungsgemäß mit ihr zusammenarbeiten. Überprüfen Sie diese Einschränkungen sorgfältig.

Bekannte Einschränkungen für die anfängliche Veröffentlichung von AIDE 9.18.1

Diese Einschränkungen gelten für die Metadata Engine, Data Compute Nodes und Management-Workflows in AIDE 9.18.1.

Anforderungen und Management von Compute Nodes

- **Mindestanforderung an Datenrechenknoten**

AIDE-Cluster benötigen mindestens und maximal 3 Datenrechenknoten (DCNs) für Metadata Engine-Funktionalität. Cluster mit weniger als 3 DCNs können Metadata Engine-Funktionen nicht aktivieren.

- **Keine Unterstützung für das Hinzufügen von DCN-Knoten über die NetApp Console**

DCN-Knoten-Upgrades und -Erweiterungen müssen mit ONTAP System Manager und nicht über die NetApp Console durchgeführt werden.

Unterstützte Datenquellen

- **Keine Unterstützung für ONTAP S3-Buckets oder StorageGRID als Datenquellen**

Nur ONTAP Volumes (lokal oder remote) werden als Datenquellen für Workspaces und die Metadaten-Katalogisierung unterstützt. ONTAP S3 Buckets und StorageGRID Objekte können nicht zu Workspaces hinzugefügt werden und werden in dieser Version nicht von der Metadata Engine indiziert.

- **Keine Unterstützung für die Erstellung von Arbeitsbereichen mit FlexCache-Volumes**

FlexCache-Volumes können nicht als Datenquellen zu Arbeitsbereichen hinzugefügt werden.

Einschränkungen bei Software-Updates und Rücksetzungen

- **Manuelle Software-Updates nur für DCN-Knoten**

Automatische Software-Updates für DCN-Cluster werden in AIDE 9.18.1 nicht unterstützt. Die DCN-Knotensoftware kann nur durch Hochladen des Images von einem lokalen Client aktualisiert werden. Das Herunterladen von Images von externen Servern (HTTP/FTP) wird nicht unterstützt.

- **Keine Rücksetzung der DCN Cluster software**

DCN cluster software kann nicht auf eine frühere Version zurückgesetzt werden. Nur Upgrades auf spätere Versionen sind erlaubt.

- **Keine ONTAP-Revertierung von AFX-Speicherclustern**

AFX storage clusters können nicht auf frühere ONTAP Versionen zurückgesetzt werden. Nur Upgrades auf spätere Versionen sind zulässig.

Workspace-Lebenszyklus und Zugriffskonfiguration

- **Kein Soft Delete oder Wiederherstellen für Arbeitsbereiche**

Das Löschen eines Arbeitsbereichs ist endgültig. Es gibt keine Option, gelöschte Arbeitsbereiche wiederherzustellen.

- **Keine Unterstützung für die OIDC-Konfiguration während der initialen Cluster-Einrichtung**

Die OIDC/OAuth-Konfiguration muss nach der Clustererstellung mit ONTAP System Manager durchgeführt werden.

Verwandte Informationen

- ["Was ist neu in AI Data Engine"](#)

Los geht's

Erfahren Sie mehr über Ihr AI Data Engine System

Erfahren Sie mehr über AI Data Engine

Die NetApp AI Data Engine (AIDE) ist eine Plattform der Enterprise-Klasse, die entwickelt wurde, um die KI-gesteuerte Datenverarbeitung, Verwaltung und Governance zu beschleunigen und zu vereinfachen. AIDE kann dabei helfen, große Mengen unstrukturierter Daten in strukturierte, KI-bereite Datensätze zu transformieren. Sie ist darauf ausgelegt, die Anforderungen moderner Machine-Learning- (ML) und generativer KI- (GenAI) Workloads zu erfüllen und sowohl traditionelle IT-Operationen als auch neue KI-zentrierte Rollen zu unterstützen.

AIDE begegnet KI-Herausforderungen

AIDE wurde entwickelt, um Organisationen bei der Verwaltung von Daten für AI-Workloads zu unterstützen und bietet die folgenden Hauptfunktionen:

- **Zentrale Metadatenverwaltung:** AIDE sammelt und katalogisiert Metadaten aus ONTAP Volumes, wodurch die Suche, Klassifizierung und Anwendung von Governance-Richtlinien auf Datensätze möglich wird.
- **Automatisierte Datenverarbeitung:** AIDE unterstützt die Erstellung von Datenpipelines für AI- und ML-Workloads, einschließlich der Möglichkeit, Vektoreinbettungen für semantische Suche zu generieren (mit entsprechender Lizenzierung).
- **Datenisolation und Zugriffskontrolle:** AIDE erzwingt Zugriffskontrollen und grundlegende Datenisolation für mehrere Teams oder Projekte.
- **Integration mit NetApp Tools:** AIDE arbeitet mit ONTAP System Manager für die Speicherverwaltung zusammen und bietet eine dedizierte Schnittstelle (AI Data Engine Console) für Dateningenieure und Wissenschaftler zur Verwaltung von Datensammlungen und Arbeitsabläufen.

High-Level-Designmerkmale

Die folgenden Designmerkmale definieren, wie AI Data Engine aufgebaut ist, um den Anforderungen von AI-Workloads gerecht zu werden:

- **Microservices-basierte Dienste:** Nutzt Kubernetes zur Orchestrierung modularer, robuster Dienste für Katalogisierung von Metadaten, Vektorsuche und Infrastrukturmanagement.
- **Sicherheit der Enterprise-Klasse:** Implementiert Verschlüsselung, rollenbasierte Zugriffssteuerung (RBAC) und Auditing für alle Daten und Metadaten.
- **Datenzugriff über mehrere Protokolle:** Unterstützt NFS und SMB für flexible Datenaufnahme und -abfrage.
- **Automatisierte Datenpipelines:** Verfolgt Datenänderungen, erstellt Embeddings und verwaltet Vektordatenbanken für AI-Anwendungen.

Wie Daten durch AIDE fließen

Das Verständnis des Datenflusses durch AIDE verdeutlicht den Wert der Plattform für KI/ML-Teams:

1. **Datenerfassung:** Dateien werden in ONTAP-Volumes mithilfe von Standardprotokollen (NFS und SMB) gespeichert. Daten können sich auf lokalem AIDE-Speicher (dem AFX-Cluster innerhalb Ihrer AIDE-Bereitstellung) oder auf entfernten ONTAP-Clustern befinden. Daten von entfernten Clustern werden mit ONTAP SnapMirror mit dem lokalen AFX-Cluster synchronisiert, sodass alle von AIDE verarbeiteten Daten letztendlich lokal gespeichert und abgerufen werden.



S3-Buckets werden nicht als Datenquellen für Workspaces oder Datensammlungen unterstützt.

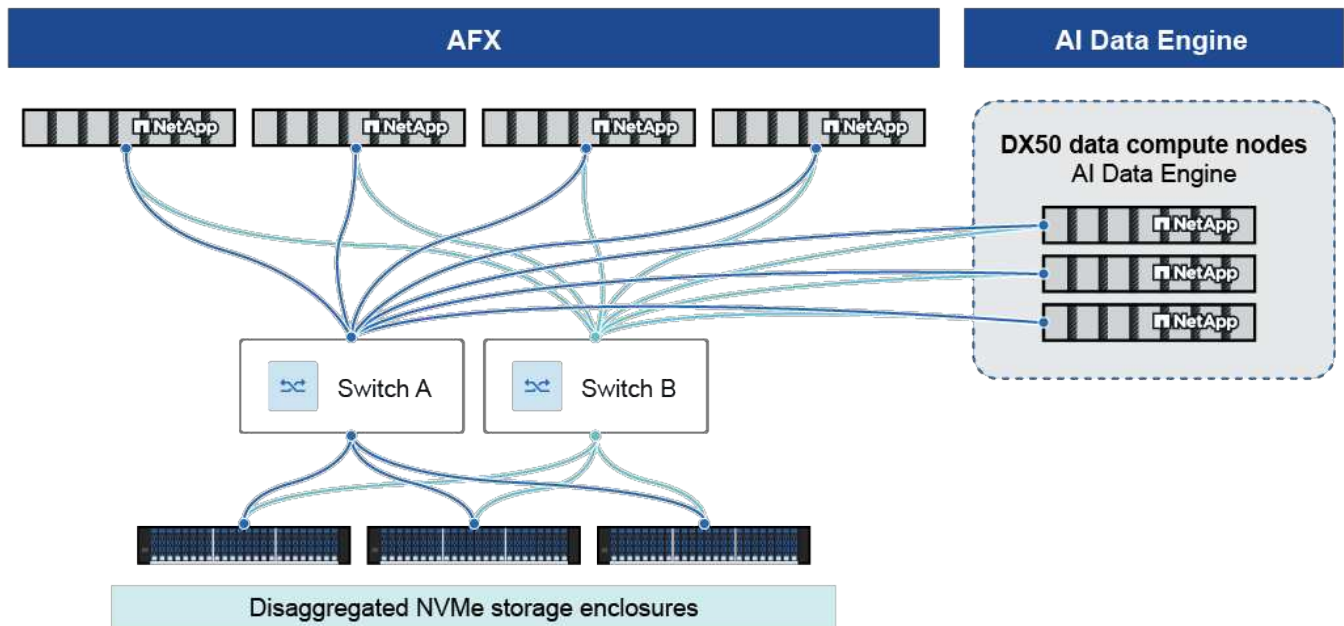
1. **Arbeitsbereichserstellung:** Speicheradministratoren definieren Arbeitsbereiche im ONTAP System Manager, gruppieren zusammengehörige ONTAP Volumes für bestimmte Projekte, Teams oder Workflows. Zugriffsberechtigungen und Governance-Richtlinien werden auf Arbeitsbereichsebene zugewiesen.
2. **Metadatenextraktion:** AIDE scannt automatisch Dateien und Objekte in Arbeitsbereichen, extrahiert Metadaten (Dateityp, Größe, Zeitstempel, benutzerdefinierte Attribute) und speichert sie in einem zentralen Katalog. Dies geschieht kontinuierlich, während sich Daten ändern.
3. **Klassifizierung und Governance:** Klassifizierer durchsuchen Daten nach sensiblen Informationen (PII, Finanzdaten) oder Dokumenttypen (rechtlich, HR). Guardrail-Richtlinien erzwingen automatisch Schwärzungen oder Zugriffsbeschränkungen.
4. **Erstellung von Datensammlungen:** Dateningenieure und Datenwissenschaftler verwenden die AI Data Engine Console, um den Metadatenkatalog abzufragen, Ergebnisse zu filtern und kuratierte Datensammlungen für spezifische AI-Aufgaben zusammenzustellen.
5. **Vektorisierung:** Für Sammlungen, die eine semantische Suche erfordern, generiert AIDE Einbettungen mithilfe ausgewählter KI-Modelle. Vektoren werden in der Vektordatenbank für eine leistungsstarke Suche gespeichert.
6. **KI/ML-Nutzung:** Anwendungen greifen über mehrere Pfade auf Daten zu:
 - Direkter Datei-/Objektzugriff über NFS oder SMB
 - Semantische Suchanfragen in der Vektordatenbank
 - RAG-Endpunkte, die Datenabruf mit GenAI-Modellintegration kombinieren
 - REST API-Zugriff für programmatische Arbeitsabläufe

Dieser automatisierte, richtlinienbasierte Workflow reduziert die Zeit und den manuellen Aufwand, die für die Vorbereitung von Daten für KI erforderlich sind, sodass sich die Teams auf die Modellentwicklung und Erkenntnisse statt auf die Datenaufbereitung konzentrieren können.

AI Data Engine Architektur

AIDE basiert auf einer skalierbaren, fehlertoleranten Architektur, die Speicher und Rechenleistung trennt und so hohe Leistung und Flexibilität für KI-Workloads ermöglicht.

Physikalische Komponenten



AFX-Controller-Knoten

AFX-Controller-Knoten führen eine spezialisierte Variante der ONTAP Software aus, die entwickelt wurde, um die Anforderungen der AFX-Umgebung zu unterstützen. Clients greifen über mehrere Protokolle, einschließlich NFS und SMB, auf die Knoten zu. Jeder Knoten hat eine vollständige Sicht auf den Speicher, auf den er basierend auf den Client-Anfragen zugreifen kann. Die Knoten sind zustandsbehaftet mit nichtflüchtigem Speicher, um kritische Zustandsinformationen zu speichern, und beinhalten zusätzliche Verbesserungen, die speziell auf die Ziel-Workloads zugeschnitten sind.

Für AIDE-Bereitstellungen sind mindestens vier AFX-Controller-Knoten erforderlich, um hohe Verfügbarkeit und Leistung sicherzustellen.

Datenrechenknoten

Datenrechenknoten (DCNs) sind Linux-basierte Server mit hoher CPU-, RAM- und GPU-Leistung, die speziell für KI-Datenverarbeitungsaufgaben vorgesehen sind. Sie hosten KI-spezifische Dienste wie Metadatenkatalogisierung, Vektorsuche und Einbettungspipelines.

Für AIDE-Bereitstellungen werden genau *drei* DCNs benötigt.

Cluster-/Storage-Switches

Redundante Hochgeschwindigkeits-Switches (100GbE oder höher) verbinden ONTAP und DCNs für Datentransfer mit niedriger Latenz und hoher Verfügbarkeit.

Storage-Shelfs

NVMe-oF-Einschübe mit SSDs hoher Dichte bieten ultra-niedrige Latenz und Redundanz und unterstützen PB-skalierbaren Speicher.

Netzwerk

Alle DCNs und ONTAP-Speicherknoten sind über redundante, Hochgeschwindigkeits-Cluster-Switches (mindestens 100GbE) verbunden. Diese Architektur trennt Rechen- und Speicherressourcen, sodass jede unabhängig skaliert werden kann und sowohl die Leistung als auch die Ressourcenauslastung optimiert

werden.

Die Netzwerkverbindung zwischen DCNs und ONTAP-Knoten ist durch dedizierte VLANs und IPspaces auf den Cluster-Switches isoliert. Dadurch wird sichergestellt, dass alle Kommunikationsarten, wie Datenzugriff, Management-APIs und interner Dienstverkehr, sicher und effizient bleiben und andere Netzwerkoperationen nicht beeinträchtigen.

AI Data Engine Hauptmerkmale

Die AI Data Engine (AIDE) Hauptfunktionen arbeiten zusammen, um den AI Data Engine Datenlebenszyklus zu automatisieren, zu sichern und zu beschleunigen. Jede Funktion ist als Satz von Microservices implementiert, die auf DCNs laufen, in ONTAP Storage integriert sind und über REST-APIs und Management-Schnittstellen bereitgestellt werden.

Metadata Engine

Die Metadata Engine generiert automatisch eine strukturierte, aktuelle und interaktive Ansicht Ihres NetApp Datenbestands.

Lizenz und Zugang

Die Metadata Engine ist in der ONTAP One-Basislizenz enthalten und ist nach der Installation von AIDE verfügbar.

Sie können über ONTAP System Manager darauf zugreifen.

Fähigkeiten

- Katalogisiert Metadaten für alle Datenquellen, einschließlich Volumes, die lokal auf dem AFX-Cluster gespeichert sind, und solche, die von entfernten ONTAP Clustern synchronisiert werden.
- Extrahiert automatisch Metadaten und füllt den Katalog, während Daten aufgenommen oder geändert werden.
- Bietet REST-API-Zugriff zum Abfragen von Metadaten und ermöglicht Datenexperten und Speicheradministratoren, Daten zu entdecken, zu klassifizieren und zu verstehen.
- Lagert Metadatenabfragen vom Datenpfad aus und reduziert so die NFS-Verkehrslast auf Speichersystemen.
- Unterstützt große Metadatensätze mit Indizierung und Suchfunktionen.
- Integriert sich in Workspace- und Datenerfassungsabstraktionen, um Zugriffskontrolle und Governance durchzusetzen.

Datensynchronisierung

Data Sync ist ein automatisierter Hintergrunddienst, der sicherstellt, dass der Metadatenkatalog und die Datensammlungen aktuell und konsistent mit den zugrunde liegenden Datenquellen bleiben, selbst wenn sich die Quelldaten ändern.

Lizenz und Zugang

Die Datensynchronisierungsfunktion ist in der ONTAP One-Basislizenz enthalten und steht nach der Installation von AIDE zur Verfügung.

Fähigkeiten

- Synchronisiert Daten von entfernten oder lokalen ONTAP Clustern mithilfe richtliniengesteuerter SnapMirror-Replikation. Daten von entfernten Clustern werden für die AIDE-Verarbeitung in den lokalen AFX-Cluster kopiert.

- Aktualisiert inkrementell basierend auf erkannten Änderungen und propagiert nur die modifizierten Daten.
- Gewährleistet sichere, inkrementelle Datenmobilität und Synchronisierung über den gesamten Datenbestand hinweg.
- Plant und überwacht Synchronisierungsintervalle mit konfigurierbaren Aktualisierungsraten pro Arbeitsbereich.
- Integriert sich in Workflows zur Erstellung von Arbeitsbereichen, um Metadaten zu extrahieren und zu aktualisieren, wenn neue Datenquellen hinzugefügt werden.

Data Guardrails

Der Data Guardrails Dienst bietet kontinuierliche, automatisierte Steuerung und Schutz für sensible Daten während des gesamten KI-Lebenszyklus.

Lizenz und Zugang

Data Guardrails Funktionalität ist nicht in der ONTAP One Base-Lizenz enthalten und erfordert eine separate AIDE Lizenz.

Sie können auf die Guardrails-Funktionalität über die AI Data Engine Console zugreifen.

Fähigkeiten

- Scannt, klassifiziert und kategorisiert kontinuierlich Daten.
- Identifiziert sensible Daten und Risiken mithilfe integrierter und anpassbarer Klassifikatoren für Aufgaben wie PII-Erkennung.
- Automatisiert den Umgang mit sensiblen Daten durch richtlinienbasierte Schwärzung, Maskierung und Zugriffsbeschränkungen.
- Setzt Unternehmens- und Regulierungsstandards durch an Arbeitsbereiche angehängte Guardrail-Richtlinien durch.
- Beschränkt den Zugriff auf sensible Dateien oder Datenträger gemäß der Konfiguration, mit Audit-Protokollierung und Compliance-Berichterstattung.
- Integriert sich in die Arbeitsbereichs- und Datenerfassungsverwaltung, um Data Guardrails konsistent über AI Data Workflows hinweg anzuwenden.

Data Curator

Der Data Curator Service ermöglicht schnelle Datenentdeckung, Suche, Vektorisierung und Abruf für AI- und GenAI-Anwendungen.

Lizenz und Zugang

Data Curator Funktionalität ist nicht in der ONTAP One Basislizenz enthalten und erfordert eine separate AIDE Lizenz.

Sie können auf Data Curator über die AI Data Engine Console zugreifen.

Fähigkeiten

- Durchsucht den Speicher nach relevanten Daten mithilfe des zentralen Metadatenkatalogs.
- Bietet Tools für Datenwissenschaftler, um kuratierte Datensammlungen zu erstellen.
- Erzeugt automatisch Vektoreinbettungen auf der Speicherschicht.
- Bietet einen sicheren Abruf-Endpunkt für KI-Anwendungen, unterstützt vektorsemantische Suche und Neubewertung.

- Lässt sich in KI-Tools und -Technologien integrieren, einschließlich Retrieval-Augmented Generation (RAG)-Pipelines und agentic AI-Frameworks.
- Bietet REST-APIs für den programmatischen Zugriff auf Datensammlungen, Vektorsuche und Abrufendpunkte.

Sicherheit und Mandantenfähigkeit

Die Plattform setzt sowohl rollenbasierte Zugriffssteuerung (RBAC) als auch ressourcenbasierte Zugriffskontrolllisten (ACLs) durch. Alle API- und Benutzeraktionen werden protokolliert, und alle Daten werden im Ruhezustand und während der Übertragung verschlüsselt. Einzelne Mandanten sind für Daten und Metadaten isoliert.

Verwandte Informationen

- ["Installieren Sie eine AIDE-Lizenz"](#)
- ["Data-to-RAG-Schnellstart"](#)

AIDE-Komponenten und Verantwortlichkeiten nach Rolle

AI Data Engine Komponenten und rollenbasierte Interaktionen

AI Data Engine (AIDE) besteht aus zahlreichen Kernkomponenten, die zusammenarbeiten, um eine umfassende Datenmanagement- und -verarbeitungsplattform für KI-Workloads bereitzustellen. Zu diesen Komponenten gehören Arbeitsbereiche, Datensammlungen, Vektordatenbanken, Schutzmechanismen, Metadatenkataloge, Abruf-Endpunkte und Klassifikatoren. Jede Komponente spielt eine spezifische Rolle bei der effizienten Datenfindung, Kuration, Verwaltung und Integration mit KI/ML-Anwendungen.

Jeder AIDE-Benutzer interagiert je nach seiner Rolle unterschiedlich mit den AIDE-Komponenten.

Speicher- und datenfokussierte Benutzerrollen

AIDE führt neue Benutzerrollen ein und unterstützt gleichzeitig weiterhin die traditionellen ONTAP-Systemadministratorrollen:

Speicherbenutzer

- **Speicheradministrator:** Verwaltet AFX- und AIDE-Cluster-Einrichtung, Netzwerk, Speicherbereitstellung und Benutzerzugriff.

Datennutzer

- **Data engineer:** Erstellt und optimiert KI/ML-Pipelines, verwaltet Datensammlungen und integriert KI-Modelle.
- **Data scientist:** Entdeckt, kuratiert und analysiert Datensätze, erstellt Datensammlungen und nutzt Retrieval-Endpunkte für GenAI-Anwendungen.

Rolle (RBAC name)	Beschreibung
Speicheradministrator (admin)	Verwaltet AFX- und AIDE-Cluster-Einrichtung, Netzwerk, Speicherbereitstellung und Benutzerzugriff. Weist Benutzern rollenbasierte Zugriffssteuerungsrollen (RBAC) zu, die den Grad des Zugriffs auf AIDE-Schnittstellen und -Funktionen bestimmen. Diese Administratorrolle hat vollständigen Verwaltungszugriff über ONTAP System Manager und AI Data Engine Console.
Data Engineer (data-engineer)	Erstellt und optimiert KI/ML-Pipelines, verwaltet Datensammlungen und integriert KI-Modelle. Diese Rolle hat Zugriff auf die AI Data Engine Console für Data-Engineering-Workflows.
Datenwissenschaftler (data-scientist)	Identifiziert, kuratiert und analysiert Datensätze, erstellt Datensammlungen und nutzt Abruf-Endpunkte für GenAI-Anwendungen. Diese Rolle hat Zugriff auf die AI Data Engine Console für Data-Science-Workflows.

AIDE Systemkomponenten

Jeder AIDE-Benutzer (Speicheradministratoren, Dateningenieure und Datenwissenschaftler) interagiert entsprechend seiner Rolle mit den AIDE-Komponenten.

Arbeitsbereiche

Ein Workspace ist ein logisches Datensegment innerhalb des Clusters, das Volumes für ein bestimmtes Projekt, Team oder einen Workflow gruppiert. Workspaces definieren den Umfang der Datensichtbarkeit, des Zugriffs und der Governance in AIDE.

Metadatenkatalog

Eine zentrale, skalierbare Datenbank speichert Metadatensätze für alle Dateien und Objekte im lokalen Cluster, einschließlich Daten, die von entfernten ONTAP-Clustern mithilfe von ONTAP SnapMirror oder Cluster-Peering synchronisiert wurden. Sie ermöglicht eine umfassende, interaktive Suche und Filterung.

Klassifikatoren

Klassifikatoren sind Werkzeuge (integriert oder benutzerdefiniert), die Dateien nach bestimmten Arten von sensiblen Daten (zum Beispiel PII, Finanzdaten, Gesundheitsdaten) durchsuchen und kennzeichnen oder Dokumente nach Typ kategorisieren (zum Beispiel Rechts-, Personal-, Vertriebsdokumente).

Datensammlungen

Eine Datensammlung ist eine kuratierte Gruppe zusammengehöriger Dateien oder Objekte aus einem Arbeitsbereich, die durch eine benutzerdefinierte Abfrage für die Verwendung in GenAI-Workflows definiert wird. Der Inhalt der Dateien in der Datensammlung steht nach der Veröffentlichung über APIs für GenAI-Anwendungen zur semantischen Suche zur Verfügung.

Vektor-Datenbank

Die Vektor-Datenbank speichert Einbettungen, die aus Datensammlungen generiert werden, und ermöglicht so eine leistungsstarke semantische Suche und Abfrage für AI- und GenAI-Anwendungen.

Leitplanken

Leitplanken sind richtlinienbasierte Mechanismen, die die Datenverwaltung, Klassifizierung und den Schutz (wie Schwärzung oder Zugriffsbeschränkungen) während des gesamten KI-Daten-Lebenszyklus durchsetzen.

Abrufendpunkt (RAG endpoint)

Ein Retrieval-Endpunkt (manchmal auch Retrieval-Augmented Generation oder „RAG“-Endpunkt genannt) ist eine sichere API, die es KI- und GenAI-Anwendungen ermöglicht, auf relevante Daten, Kontext oder Einbettungen aus kuratierten Sammlungen und der Vektor-Datenbank zuzugreifen.

RAG-Endpunkte sind für die Unterstützung fortschrittlicher KI-Workflows konzipiert, beispielsweise für die semantische Suche und kontextbezogene Antworten in generativen KI-Modellen. Durch die Anbindung Ihrer KI-Anwendungen an einen Retrieval-Endpunkt können Sie die Genauigkeit und Relevanz der Modelle verbessern, indem Sie Echtzeit-Zugriff auf kuratierte, KI-bereite Datensätze ermöglichen, die von AIDE verwaltet werden.

Verwandte Informationen

- ["Wie AIDE-Speicheradministratoren mit AIDE-Komponenten arbeiten"](#)
- ["Wie AIDE-Dateningenieure mit AIDE-Komponenten arbeiten"](#)
- ["Wie AIDE-Datenwissenschaftler mit AIDE-Komponenten arbeiten"](#)

AI Data Engine Schnittstellen

AI Data Engine (AIDE) bietet drei primäre Schnittstellen für Benutzerinteraktion und Automatisierung. Jede Rolle, wie Speicheradministratoren, Dateningenieure und Datenwissenschaftler, nutzt diese Schnittstellen entsprechend ihren spezifischen Aufgaben und Verantwortlichkeiten.

ONTAP System Manager

ONTAP System Manager ist eine webbasierte Schnittstelle für Speicheradministratoren. Sie bietet Workflows für die Cluster-Einrichtung, die Workspace-Verwaltung, DCN-Monitoring und das Anhängen von Guardrail-Richtlinien.

AI Data Engine Console

Die AI Data Engine Console ist eine dedizierte Schnittstelle für Data Engineers und Data Scientists. Sie ermöglicht es Nutzern, Datenquellen zu erkunden, Datensammlungen zu erstellen und zu verwalten, Datenpipelines zu konfigurieren, Klassifikatoren anzuwenden und mit Schutzmechanismen und Vektorsuchfunktionen zu interagieren. Die Konsole bietet fortschrittliche Werkzeuge für die Datenermittlung, Kuration und Integration mit AI/ML-Workflows.

REST API

AIDE stellt die ONTAP REST API für Automatisierung, Integration und programmatischen Zugriff bereit. Die API unterstützt Cluster-Setup, Workspace- und Collection-Management, Metadatenabfragen, Vektorsuche und Retrieval-Endpunkte.

Erfahren Sie, wie AI Data Engine-Speicheradministratoren mit AIDE-Komponenten arbeiten

Als Speicheradministrator verwalten Sie die AIDE Infrastruktur über ONTAP und die AIDE

Console, stellen Arbeitsbereiche bereit, fügen Guardrail-Richtlinien hinzu und überwachen den Systemzustand. Ihre Aufgabe besteht darin, eine zuverlässige, sichere und konforme Datenspeicherung für AI-Workloads zu gewährleisten.

Zugriff auf die Komponente des Storage Administrators

Komponente	Zugriffsebene	Workflow des Speicheradministrators
ONTAP System Manager	Verwalten (erstellen, bearbeiten, löschen)	Sie verwenden ONTAP System Manager als Ihre primäre Schnittstelle für die Clusterverwaltung, die Bereitstellung von Arbeitsbereichen, die Verwaltung von Guardrail-Richtlinien und die Überwachung des Systemzustands.
AI Data Engine Console	Verwalten (erstellen, bearbeiten, löschen)	Sie verwenden die AI Data Engine Console, um Arbeitsbereiche zu überwachen, den Sammlungsstatus anzuzeigen und die Systemaktivitäten in der gesamten AIDE-Umgebung zu überwachen.
ONTAP REST API	Verwalten (erstellen, bearbeiten, löschen)	Sie verwenden die REST API, um Infrastrukturaufgaben zu automatisieren, Arbeitsbereiche und Guardrail-Richtlinien programmatisch zu verwalten und die AIDE Administration mit externen Tools und Workflows zu integrieren.
Arbeitsbereiche	Verwalten (erstellen, bearbeiten, löschen)	Sie erstellen und verwalten Arbeitsbereiche mit ONTAP System Manager. Sie wählen aus, welche Datenquellen einbezogen werden, weisen Dateningenieuren und Data Scientists Berechtigungen zu und fügen Guardrail-Richtlinien hinzu, um Governance und Compliance durchzusetzen. Sie überwachen außerdem die Gesundheit und den Zugriff der Arbeitsbereiche.
Datensammlungen	Ansicht (nur lesen)	Sie sehen den Status und die Integrität von Datensammlungen in jedem Arbeitsbereich mit System Manager. Sie stellen sicher, dass die zugrunde liegenden Datenquellen verfügbar und geschützt sind, erstellen oder ändern jedoch keine Sammlungen.
Leitplanken	Verwalten (erstellen, bearbeiten, löschen)	Sie definieren und ordnen Schutzrichtlinien mithilfe von System Manager Arbeitsbereichen zu. Sie überwachen den Status der Schutzrichtlinien und Compliance-Berichte. Sie stellen sicher, dass die Richtlinien durchgesetzt und bei Bedarf aktualisiert werden.

Komponente	Zugriffsebene	Workflow des Speicheradministrators
Metadatenkatalog	Überwachen (Gesundheit, Status, Aktivität anzeigen)	Sie stellen sicher, dass der Metadatenkatalog gefüllt und aktuell ist. Sie überwachen den Zustand des Katalogs und unterstützen die Zugriffskontrolle.
Vektor-Datenbank	Bereitstellen/Überwachen (bereitstellen, konfigurieren, Status anzeigen)	Sie stellen die Vektor-Datenbank-Infrastruktur bereit und überwachen sie, wobei Sie sicherstellen, dass Datenverarbeitungsknoten mit GPU-Ressourcen und ordnungsgemäße Lizenzierung vorhanden sind. Sie unterstützen die Umgebung, verwalten aber Embeddings oder Abfragen nicht direkt.
Klassifikatoren	Verwalten (erstellen, bearbeiten, löschen)	Sie erstellen, konfigurieren und verwalten Klassifikatoren und deren Kategorien. Sie wenden Klassifikatoren auf Arbeitsbereiche an und überwachen deren Effektivität.

Erfahren Sie, wie Dateningenieure und Datenwissenschaftler mit den AIDE-Komponenten von AI Data Engine arbeiten.

Als Data Engineer oder Data Scientist nutzen Sie die AI Data Engine Console, um Arbeitsbereiche zu erkunden, zu denen Sie Zugriff erhalten haben, Datenkollektionen zu erstellen und zu verwalten, semantische Suchen durchzuführen und Retrieval-Endpunkte in AI/ML-Workflows zu integrieren.

Data Engineers konzentrieren sich darauf, Rohdaten in KI-fähige Datensätze umzuwandeln, indem sie Sammlungen erstellen, Einbettungspipelines konfigurieren und steuern, welche Benutzer auf veröffentlichte Sammlungen zugreifen können. Data Scientists konzentrieren sich darauf, kuratierte Datensätze für Analysen, Modelltraining und GenAI-Anwendungen zu nutzen, ohne Zugriffskontrolle oder Infrastruktur verwalten zu müssen.

Zugriff auf Datenbenutzerkomponenten

Komponente	Zugriffsebene	Arbeitsablauf des Data Engineers	Arbeitsablauf eines Data Scientist
AI Data Engine Console	Verwalten (erstellen, bearbeiten, löschen)	Die AI Data Engine Console ist Ihre primäre Schnittstelle für alltägliche Aufgaben, einschließlich Datenermittlung, Sammlungsverwaltung, Pipeline-Konfiguration und Veröffentlichung von RAG- oder Retrieval-Endpunkten für die Arbeitsbereiche, zu denen Sie berechtigt sind.	Die AI Data Engine Console ist Ihre primäre Schnittstelle für die Datenexploration, Verfeinerung und Versionierung von Sammlungen innerhalb von Arbeitsbereichen, auf die Sie zugreifen können, sowie für die Verbindung von kuratierten Datensätzen und Abrufendpunkten mit Analyse-, Modellierung- und GenAI-Workflows.

Komponente	Zugriffsebene	Arbeitsablauf des Data Engineers	Arbeitsablauf eines Data Scientist
ONTAP REST API	Verwalten (erstellen, bearbeiten, löschen)	Sie verwenden die REST API, um Vorgänge im Lebenszyklus von Sammlungen zu automatisieren, Einbettungspipelines auszulösen und zu überwachen sowie Daten-Workflows programmatisch mit externen Tools zu integrieren.	Sie verwenden die REST API, um programmatisch auf Datensammlungen zuzugreifen, Vektorsuchabfragen auszuführen und Abrufendpunkte in AI/ML-Anwendungen und agentenbasierte Frameworks zu integrieren.
Arbeitsbereiche	Ansehen/Verwenden (nur lesen)	Sie erkunden Ihre zugewiesenen Arbeitsbereiche, um verfügbare Datenquellen zu identifizieren und zu verstehen, bevor Sie Collections erstellen.	Sie durchsuchen Ihre zugewiesenen Arbeitsbereiche, um Dateien und Objekte zu finden, die für bestimmte Forschungs- oder Modellierungsaufgaben relevant sind.
Datensammlungen	Verwalten (erstellen, bearbeiten, löschen)	Sie erstellen Datensammlungen, indem Sie Quelldaten mithilfe von Tags, Klassifizierung und anderen Attributen auswählen und filtern, und Sie verwalten den vollständigen Lebenszyklus der Sammlung von der Erstellung und Versionierung bis zur Veröffentlichung als RAG-Endpunkte für die KI-Nutzung. Sie verwalten außerdem, welche Data Scientists und andere Nutzer auf jede Sammlung zugreifen können.	Sie erstellen, wählen aus, annotieren, versionieren und verfeinern Datensammlungen in den Arbeitsbereichen, auf die Sie Zugriff haben. Sie verwenden diese Sammlungen als Grundlage für semantische Such- und GenAI-Workflows.
Metadatenkatalog	Abfrage/Nutzung (Verbrauch für Workflows)	Sie verwenden den Metadatenkatalog, um Datenquellen für die Aufnahme zu bewerten und auszuwählen, Abfragen auszuführen, um relevante Dateien zu finden und zu bestätigen, dass sie die Anforderungen der Sammlungen erfüllen, die Sie in Ihren zugewiesenen Arbeitsbereichen erstellen.	Sie durchsuchen und filtern Metadaten in den Arbeitsbereichen, auf die Sie Zugriff haben, um Dateien und Objekte zu finden, die für die Analyse oder das Modelltraining benötigt werden, und verlassen sich dabei auf die Katalogstruktur, die von Dateningenieuren erstellt und gepflegt wurde.

Komponente	Zugriffsebene	Arbeitsablauf des Data Engineers	Arbeitsablauf eines Data Scientist
Vektor-Datenbank	• Einbettungen/Suche verwalten (Data Engineer) • Verwenden/Suchen (Data Scientist)	Sie lösen Einbettungspipelines aus, überwachen den Vektorisierungsstatus, konfigurieren Chunking- und Einbettungsparameter und stellen Abruf-Endpunkte bereit, die auf Vektorsuche basieren. Anwendungen und Agenten fragen diese Endpunkte dann über die API für semantische Suche und RAG-Workflows ab.	Sie führen semantische Suchanfragen gegen Einbettungen durch, die von Dateningenieuren verwalteten Pipelines generiert werden, und integrieren die Abrufergebnisse in GenAI- oder RAG-Workflows für kontextbezogene Modellantworten. Sie konfigurieren weder Chunking, Einbettungen noch Pipeline-Parameter.
Klassifikatoren	Verwendung (klassifizierte Daten verwenden)	Sie verwenden die Ergebnisse der Klassifizierung, um Quelldaten während der Vorbereitung der Datenerfassung zu annotieren und zu kennzeichnen und so sicherzustellen, dass die in Ihre Pipelines eingehenden Inhalte für nachgelagerte KI-Workflows ordnungsgemäß gekennzeichnet sind.	Sie verwenden vorklassifizierte Daten, um sicherzustellen, dass nur konforme und relevante Inhalte in Ihren Analysen und Ihrer Modellierung verwendet werden.

Schnellstart für AI Data Engine

Um Ihr AI Data Engine system in Betrieb zu nehmen, müssen Sie Ihre Hardwarekomponenten installieren, Ihren Cluster einrichten, den Datenzugriff von Ihren Hosts auf das Speichersystem einrichten und Ihren Speicher bereitstellen.

1

Wird AIDE mit einem neuen oder einem bestehenden AFX-Cluster installiert?

Sie müssen entscheiden, ob AIDE und AFX gleichzeitig installiert werden oder ob AIDE mit einem bestehenden AFX-Cluster integriert wird.

2

Installieren und richten Sie Ihre Hardware ein

"[Installieren und einrichten](#)" Ihre AIDE cluster compute nodes. Je nach Installationsumgebung stellen Sie außerdem sicher, dass Sie auch "[installieren](#)" die AFX hardware.

3

Richten Sie Ihren Cluster ein

Verwenden Sie ONTAP System Manager, um sich durch einen schnellen und einfachen Prozess führen zu lassen "[AIDE mit einem AFX-Cluster einrichten](#)".

4

Arbeitsbereiche und Datenzugriff einrichten

"Richten Sie einen Arbeitsbereich und Benutzer für diesen Arbeitsbereich ein, die auf AI Data Engine Daten zugreifen können."

Was kommt als Nächstes?

Sie können jetzt den ONTAP System Manager verwenden, um Ihre AI Data Engine zu verwalten und Ihren Dateningenieuren und Datenwissenschaftlern den Einstieg in ihre Arbeitsbereiche und Konfigurationen zu ermöglichen.

Installieren Sie AIDE

Voraussetzungen für die Installation von AI Data Engine

Prüfen Sie die Anforderungen für die Installation der AI Data Engine. AIDE benötigt ein AFX-Speichersystem, mindestens drei Data Compute Node, Netzwerk-Switches und Kabel.

Hardwareanforderungen

Die AI Data Engine benötigt ein AFX-Speichersystem und mindestens drei Data Compute Nodes. Das AFX-System stellt die Speicherinfrastruktur bereit, während die Data Compute Nodes die AIDE-Softwarekomponenten hosten, die Datenmanagement, Datenkuratierung und AI-Funktionen ermöglichen.

- **AFX-Speichersystem:** Besteht aus einem AFX-Controller, einem Festplattengehäuse und Netzwerk-Switch. Das AFX-Speichersystem ist für die AIDE-Bereitstellung erforderlich.
- **Datenrechenknoten:** Es werden mindestens drei Datenrechenknoten benötigt. Die Datenrechenknoten sind von NetApp bereitgestellte Hardwareknoten, die die AIDE-Software hosten, einschließlich der Metadata Engine, Data Sync, Data Curator und Data Guardrails.
 - Jeder Data Compute Node verfügt über die I/O-Steckplätze 4 und 5 für die Konnektivität. Steckplatz 3 ist für die GPU reserviert. Steckplätze 1 und 2 sind nicht belegt oder zugänglich.
 - Die Ports e4a und e5a werden für Clusterverbindungen verwendet.
 - Die Ports e4b und e5b werden für Host-Netzwerkverbindungen verwendet.

Anforderungen an Netzwerk-Switch

AI Data Engine benötigt Netzwerk-Switches, um den Netzwerkzugriff für Hosts und die Kommunikation zwischen den Data Compute Nodes zu ermöglichen.

- Client-Switches für die Host-Netzwerk-konnektivität (Cisco Nexus 9332D-GX2B oder Cisco Nexus 9364D-GX2A)
- Cluster-Switches für die Inter-Node-Kommunikation (Cisco Nexus 9332D-GX2B oder Cisco Nexus 9364D-GX2A)
- Optionale Management-Switches für Netzwerkmanagement

Verkabelungsanforderungen

Folgende Kabel werden benötigt, um die Data Compute Node mit den Netzwerk-Switches und dem Management-Netzwerk zu verbinden.

- 400-GbE-zu-100-GbE (4x100GbE) Breakout-Kabel zum Verbinden von Knoten mit Client- und Cluster-Switches

Unterstützung mehrerer Cluster

Nachdem AIDE mit AFX bereitgestellt wurde, kann es eine Verbindung zu und Verwaltung von Daten anderer ONTAP 9.18.1- und späterer Cluster mithilfe von SnapMirror und Cluster-Peering herstellen.

Installieren Sie das AFX-Speichersystem für AI Data Engine

Installieren Sie das AFX-Speichersystem als ersten Schritt bei der Bereitstellung der AI Data Engine. Das AFX-Speichersystem bildet die Grundlage der Speicherinfrastruktur und ist vor der Installation der Data Compute Node erforderlich.

Folgen Sie dem ["AFX 1K Installationsdokumentation"](#) zur Installation des AFX-Speichersystems.

Was kommt als Nächstes

Nach Abschluss der Installation des AFX-Speichersystems ["Installieren Sie die Data Compute Nodes"](#).

Installieren Sie Ihre Datenrechenknoten

Installations- und Einrichtungsworkflow für Datenrechenknoten für AI Data Engine

Zur Installation und Konfiguration Ihrer Data Compute Nodes (DCN) überprüfen Sie die Hardwareanforderungen, bereiten Ihren Standort vor, installieren und verkabeln die Hardwarekomponenten, schalten das System ein und richten Ihr ONTAP Cluster ein.

1

["Überprüfen Sie die Hardware-Installationsanforderungen"](#)

Stellen Sie sicher, dass Sie über ein vorhandenes AFX 1K-Speichersystem verfügen, und überprüfen Sie anschließend die Hardwareanforderungen für die Installation der Datenrechenknoten für AIDE. Informationen zur Installation des AFX 1K-Speichersystems finden Sie unter ["AFX 1K Speichersystem-Installationsdokumentation"](#).

2

["Bereiten Sie die Installation Ihrer Data Compute Nodes vor"](#)

Zur Vorbereitung der Installation Ihrer Datenrechenknoten müssen Sie den Standort vorbereiten, die Umgebungs- und Stromversorgungsanforderungen prüfen und sicherstellen, dass ausreichend Platz in den Serverschränken vorhanden ist. Entpacken Sie anschließend die Geräte, vergleichen Sie den Inhalt mit dem Lieferschein und registrieren Sie die Hardware, um Supportleistungen in Anspruch nehmen zu können. Mindestens drei Datenrechenknoten sind erforderlich, um die AI Data Engine bereitzustellen.

3

["Installieren Sie die Hardware für Ihre Data Compute Nodes"](#)

Installieren Sie die Schienensätze für Ihre Datenrechenknoten. Sichern Sie Ihre Datenrechenknoten im Schrank. Bringen Sie abschließend Kabelmanagement-Geräte an der Rückseite des Systems an, um eine organisierte Kabelführung zu gewährleisten.

4

"Verkabeln Sie Ihre Datenrechenknoten"

Um die Hardware zu verkabeln, verbinden Sie zuerst die Datenrechenknoten mit Ihrem Datennetzwerk, dann verbinden Sie die Datenrechenknoten mit den Cluster-Switches.

5

"Schalten Sie Ihre Datenrechenknoten ein"

Nachdem Sie die Rack-Hardware installiert und Ihre Datenrechenknoten verkabelt haben, sollten Sie Ihre DCNs und Ihre Controller-Knoten für das AFX-Speichersystem einschalten, falls diese noch nicht eingeschaltet sind.

Installationsanforderungen für Datenrechenknoten für AI Data Engine

Überprüfen Sie die benötigte Ausrüstung und die Hebevorsichtsmaßnahmen für Ihre Data Compute Nodes für AI Data Engine.

Voraussetzungen

Bevor Sie die Datenrechenknoten für AIDE installieren, stellen Sie sicher, dass Sie Folgendes haben:

- Ein AFX 1K Storage-System.



Informationen zur Installation des AFX 1K-Speichersystems finden Sie unter ["AFX 1K Speichersystem-Installationsdokumentation"](#).

Für die Installation benötigte Ausrüstung

Zur Installation der Data Compute Nodes für AIDE benötigen Sie die folgende Ausrüstung und Werkzeuge.

- Zugriff auf einen Webbrowser zur Konfiguration Ihrer Datenrechenknoten
- Elektrostatische Entladung (ESD) Armband
- Taschenlampe
- Laptop oder Konsole mit USB-/serieller Verbindung
- Kreuzschlitzschraubendreher Nr. 2

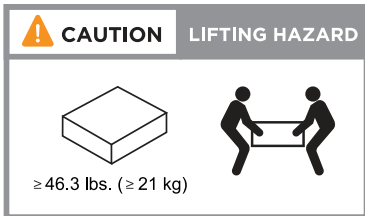
Vorsichtsmaßnahmen beim Heben

Die Datenverarbeitungsknoten sind schwer. Seien Sie beim Anheben und Bewegen dieser Gegenstände vorsichtig.

Daten-Compute-Knotengewichte

Treffen Sie die notwendigen Vorsichtsmaßnahmen beim Bewegen oder Anheben Ihres Data Compute Node.

Ein Datenrechenknoten kann bis zu 46,3 lbs (21 kg) wiegen. Zum Anheben des Datenrechenknotens sollten zwei Personen oder ein hydraulischer Lift verwendet werden.



Verwandte Informationen

- ["Sicherheitsinformationen und behördliche Hinweise"](#)

Was kommt als Nächstes?

Nachdem Sie die Hardwareanforderungen geprüft haben, ["Bereiten Sie die Installation Ihrer Data Compute Nodes vor"](#).

Bereiten Sie die Installation Ihrer Datenrechenknoten für AI Data Engine vor

Bereiten Sie die Installation Ihrer Datenrechenknoten für AI Data Engine vor, indem Sie den Standort vorbereiten, die Kartons entpacken und den Inhalt der Kartons mit dem Lieferschein vergleichen sowie das System registrieren, um Supportleistungen in Anspruch nehmen zu können.

Schritt 1: Bereiten Sie den Standort vor

Um Ihre Daten-Compute-Nodes zu installieren, stellen Sie sicher, dass der Standort und der Schrank oder das Rack, das Sie verwenden möchten, die Spezifikationen für Ihre Konfiguration erfüllen.

Schritte

1. Verwenden Sie ["NetApp Hardware Universe"](#), um zu bestätigen, dass Ihr Standort die Umwelt- und Elektroanforderungen für Ihre Datenrechenknoten erfüllt.
2. Stellen Sie sicher, dass innerhalb der bestehenden AFX 1K Storage-Systeminstallation ausreichend Schrankplatz für Ihre Data Compute Nodes vorhanden ist.
 - 1U für jeden Daten-Compute-Node
 - 2U für jeden AFX 1K-Controller-Knoten
 - 2U pro NX224 shelf
 - 1U oder 2U pro Switch, abhängig vom Switch-Modell.

Schritt 2: Entpacken Sie die Kartons

Nachdem Sie sichergestellt haben, dass der Standort und der Schrank oder das Rack, das Sie für Ihre Datenrechenknoten verwenden möchten, die erforderlichen Spezifikationen erfüllen, entpacken Sie alle Kartons und vergleichen Sie den Inhalt mit den Angaben auf dem Lieferschein.

Schritte

1. Öffnen Sie alle Kartons vorsichtig und legen Sie den Inhalt in einer organisierten Weise aus.
2. Vergleichen Sie die entpackten Artikel mit dem Lieferschein und notieren Sie etwaige Abweichungen.

Sie können Ihre Packliste abrufen, indem Sie den QR-Code an der Seite des Versandkartons scannen.

Die folgenden Elemente sind einige der Inhalte, die Sie in den Boxen sehen könnten.

Hardware	Kabel	
• Blende • Schienenbausätze mit Anleitung • Data-Compute-Knoten	• Netzkabel	

Schritt 3: Registrieren Sie Ihre Daten-Compute-Nodes

Nachdem Sie sichergestellt haben, dass Ihre Site die Anforderungen an die Spezifikationen Ihres Data Compute Node erfüllt und Sie überprüft haben, dass Sie alle bestellten Teile erhalten haben, sollten Sie Ihr System registrieren.

Schritte

1. Suchen Sie die Seriennummern Ihrer Daten-Compute-Nodes.

Die Seriennummern finden Sie an den folgenden Stellen:

- Auf dem Lieferschein
- In Ihrer Bestätigungs-E-Mail
- Auf jedem Datenrechenknoten oder auf einigen Systemen auf dem Systemmanagement-Modul jedes Datenrechenknotens.



2. Navigieren Sie zu ["NetApp Support Site"](#).
3. Prüfen Sie, ob Sie Ihr Speichersystem registrieren müssen:

Wenn Sie ein/e...	Folgen Sie diesen Schritten...
Bestehender NetApp-Kunde	a. Sign in mit Ihrem Benutzernamen und Passwort. b. Wählen Sie Systems > My Systems. c. Bestätigen Sie, dass die neue Seriennummer aufgeführt ist. d. Falls die Seriennummer nicht aufgeführt ist, befolgen Sie die Anweisungen für neue NetApp Kunden.
Neuer NetApp-Kunde	a. Klicken Sie auf Register Now, um ein Konto zu erstellen. b. Wählen Sie Systeme > Systeme registrieren. c. Geben Sie die Seriennummer des Storage-Systems und die angeforderten Details ein. Sobald NetApp Ihre Registrierung genehmigt, können Sie die benötigte Software herunterladen. Die Genehmigung kann bis zu 24 Stunden dauern.

Was kommt als Nächstes?

Nachdem Sie die Installation Ihrer Datenrechenknoten vorbereitet haben, "[Installieren Sie die Data Compute Nodes](#)".

Installieren Sie Ihre Daten-Compute-Nodes für AI Data Engine

Installieren und sichern Sie Ihre Data Compute Nodes im Schrank.

Bevor Sie beginnen

- Vergewissern Sie sich, dass die Anleitung dem Schienenbausatz beiliegt.
- Verstehen Sie die Sicherheitsbedenken im Zusammenhang mit dem Gewicht des Datenrechenknotens, des Speichersystems und des Speicherregals.
- Beachten Sie, dass der Luftstrom durch das Speichersystem von vorne eintritt, wo die Blende oder die Endkappen installiert sind, und hinten wieder austritt, wo sich die Ports befinden.



Im Allgemeinen sollten die Switches mittig im Schrank installiert werden. Die Ablagefächer sollten unterhalb des Switches und oberhalb eines zweiten installierten Switches installiert werden. Die Controller-Knoten können oberhalb oder unterhalb der Switches im Schrank installiert werden. Datenverarbeitungsknoten können oberhalb oder unterhalb der Controller-Knoten im Schrank installiert werden.

Schritte

1. Installieren Sie die Schienenbausätze für Ihre Data Compute Nodes nach Bedarf anhand der den Kits beiliegenden Anweisungen.
2. Installieren und sichern Sie Ihre Daten-Compute-Nodes im Schrank:
 - a. Positionieren Sie den Daten-Compute-Node auf den Schienen in der Mitte des Schanks und stützen Sie dann das Appliance von unten ab und schieben Sie es an seinen Platz.
 - b. Befestigen Sie den Datenrechenknoten mit den mitgelieferten Befestigungsschrauben am Schrank.
3. Bringen Sie die Blende an der Vorderseite des Datenrechenknotens an.

Was kommt als Nächstes?

Nachdem Sie die Datenrechenknoten installiert haben, "[Verkabeln Sie die Daten-Compute-Nodes für AIDE](#)".

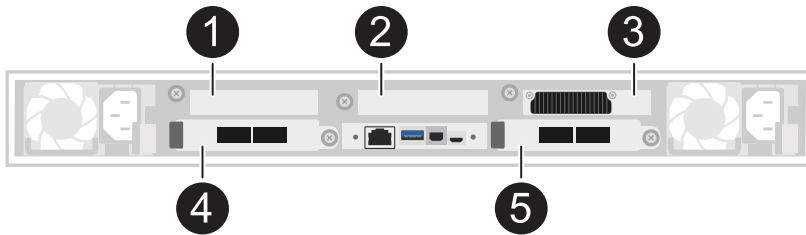
Anforderungen an die Verkabelung von Data Compute Node für AI Data Engine

Data Compute Nodes werden über Host-Netzwerk- und Cluster-Netzwerkverbindungen in Ihr AFX 1K Storage System integriert. Überprüfen Sie die I/O-Steckplatzkonfiguration, die Kabeltypen und die Verbindungsanforderungen für Ihre Bereitstellung.

Verkabelungskonfiguration

Data Compute Nodes werden an dieselben Cluster-Switches wie die AFX 1K Controller Nodes angeschlossen und erweitern so Ihr Speichersystem um Rechenressourcen, die für AI- und Machine-Learning-Workloads optimiert sind.

Die anfängliche AI Data Engine (AIDE) Konfiguration unterstützt mindestens drei Data Compute Nodes. Ausführliche Konfigurationsdetails und Slot-Prioritäten finden Sie unter "[NetApp Hardware Universe](#)".



1	Unbenutzter Steckplatz auf dem Data Compute Node.
2	Unbenutzter Steckplatz auf dem Data Compute Node.
3	GPU-Steckplatz auf dem Data Compute Node.
4	E/A-Steckplatz am Data Compute Node.
5	E/A-Steckplatz am Data Compute Node.

E/A-Steckplatzkonfiguration

Der Data Compute Node verwendet ein spezielles Steckplatznummerierungsschema, das sich von Standardserverkonfigurationen unterscheidet. Das Verständnis des Steckplatzlayouts ist für die korrekte Verkabelung unerlässlich.

- **Steckplatz 3:** Für GPU reserviert (nicht zugänglich für I/O-Kabel)
- **Steckplätze 4 und 5:** E/A-Steckplätze, die für Netzwerkverbindungen verwendet werden
 - Port a: Cluster-Netzwerkverbindungen
 - Port b: Host-Netzwerkverbindungen
- **Slots 1 und 2:** Nicht belegt und nicht zugänglich für die Nutzung

Netzwerkverbindungen

Data Compute Nodes benötigen zwei Arten von Netzwerkverbindungen, um mit dem AFX 1K-Speichersystem integriert zu werden.

• Host-Netzwerkverbindungen

Host-Netzwerkverbindungen bieten Zugriff auf Clientdaten und ermöglichen es den Data Compute Nodes, Workloads zu verarbeiten. Jeder Data Compute Node verwendet die Ports e4b und e5b für redundante Verbindungen zu separaten Netzwerk-Switches.

Hafenzuweisungen:

- e4b: Verbindet sich mit dem Netzwerk-Switch A
- e5b: Verbindet sich mit dem Host Netzwerk-Switch B

• Cluster-Netzwerkverbindungen

Cluster-Netzwerkverbindungen ermöglichen die Kommunikation zwischen Data Compute Nodes und AFX

1K Controller Nodes innerhalb des Storage Clusters. Jeder Data Compute Node verwendet die Ports e4a und e5a für redundante Verbindungen zu separaten Cluster-Netzwerk-Switches.

Hafenzuweisungen:

- e4a: Verbindet sich mit dem Netzwerk-Switch A des Clusters
- e5a: Verbindet sich mit dem Cluster Netzwerk-Switch B

Unterstützte Hardwarekomponenten

Die Data Compute Nodes erfordern spezielle Kabel und Netzwerk-Switches, um eine ordnungsgemäße Konnektivität und Leistung mit dem AFX 1K storage system sicherzustellen.

Data Compute Node	Unterstützte Switches	Unterstützte Kabel
Data Compute Nodes (mindestens drei erforderlich)	• Cisco Nexus 9332D-GX2B (400GbE) • Cisco Nexus 9364D-GX2A (400GbE)	• 400GbE QSFP-DD Breakout auf 4x100GbE QSFP56-Kabel für Verbindungen zu Data Compute Nodes: ◦ 100GbE zu den Data Compute Node-Cluster-Netzwerkports (e4a, e5a) ◦ 100GbE zu den Host-Netzwerkports des Data Compute Node (e4b, e5b) • RJ-45-Kabel für Managementverbindungen

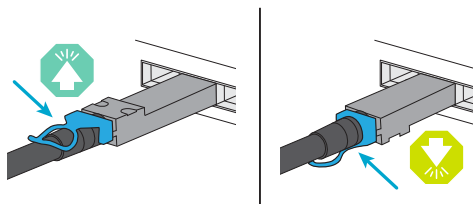


Breakout-Kabel stellen vier 100-GbE-Verbindungen von jedem 400-GbE-Switch-Port bereit. Verbinden Sie das 400-GbE-Ende mit den Switches und das 100-GbE-Ende mit den Data Compute Node I/O-Ports.

Kabelausrichtung

Beim Anschließen von Kabeln an Data Compute Nodes sorgt die richtige Kabelausrichtung für zuverlässige Verbindungen.

Die Verkabelungsgrafiken in den Installationsanweisungen zeigen Pfeilsymbole, die die korrekte Ausrichtung (oben oder unten) der Zuglasche des Kabelsteckers beim Einführen eines Steckers in einen Port anzeigen. Beim Einführen des Steckers sollten Sie spüren, wie er einrastet. Wenn Sie kein Einrasten spüren, entfernen Sie ihn, drehen Sie ihn um und versuchen Sie es erneut.



Behandeln Sie die empfindlichen Steckverbinderkomponenten vorsichtig, wenn Sie sie einrasten lassen.

Was kommt als Nächstes?

Nach Überprüfung der Verkabelungskonfiguration "[Verkabeln Sie die Hardware](#)" für Ihre Data Compute Nodes.

Verkabeln Sie Ihre Data Compute Node für AI Data Engine

Verbinden Sie Ihre Data Compute Nodes mit den Netzwerk-Switches des Host-Netzwerks und des Clusters, um die Verarbeitung von KI-Workloads und die Integration mit Ihrem AFX 1K-Speichersystem zu ermöglichen. Dieses Verfahren verwendet 100GbE-Verbindungen sowohl für den Netzwerkzugriff als auch für die Clusterkommunikation, sodass die Nodes die bestehende Clusterinfrastruktur nutzen können, ohne das AFX-System herunterzufahren.

Über diese Aufgabe

Diese Verfahren zeigen gängige Konfigurationen. Die spezifische Verkabelung hängt von den für Ihr Speichersystem bestellten Komponenten ab. Ausführliche Konfigurationsdetails und Steckplatzprioritäten finden Sie unter "[NetApp Hardware Universe](#)".



Sie müssen das AFX 1K storage system beim Verkabeln der Data Compute Node nicht ausschalten. Sie können die Data Compute Node zu einem bestehenden AFX 1K storage system hinzufügen, das bereits eingeschaltet und konfiguriert ist.

Bevor Sie beginnen

- Sie haben bereits ein AFX 1K-Speichersystem installiert. Informationen zur Installation des AFX 1K-Speichersystems finden Sie unter "[AFX 1K Speichersystem-Installationsdokumentation](#)".
- Die erforderlichen Netzwerk-Switches sind installiert und konfiguriert. Wenden Sie sich an Ihren Netzwerkadministrator, um Informationen zum Anschluss des Systems an Ihre Netzwerk-Switches zu erhalten.
- Sie haben die "[Verkabelungsanforderungen für die Data Compute Node](#)" überprüft.



Mindestens drei Data Compute Nodes sind erforderlich, um die AI Data Engine bereitzustellen.

Schritt 1: Verbinden Sie die Data Compute Nodes mit dem Hostnetzwerk

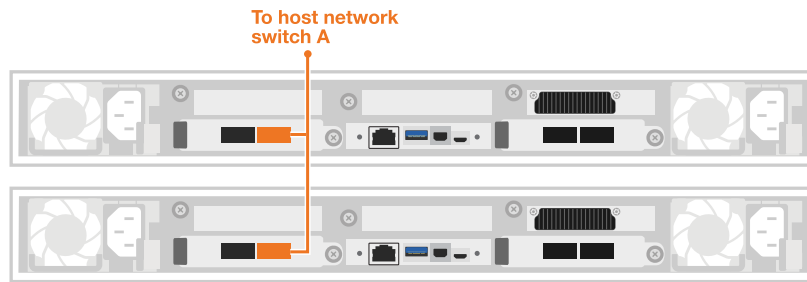
Sie können die Data Compute Node-Ports mit Ihrem Host-Netzwerk verbinden.

Schritte

1. Verbinden Sie Port e4b der folgenden Data Compute Nodes mit dem Ethernet-Datennetzwerk Netzwerk-Switch A:
 - Data Compute Node 1, Port e4b
 - Data Compute Node 2, Port e4b

100GbE-Kabel

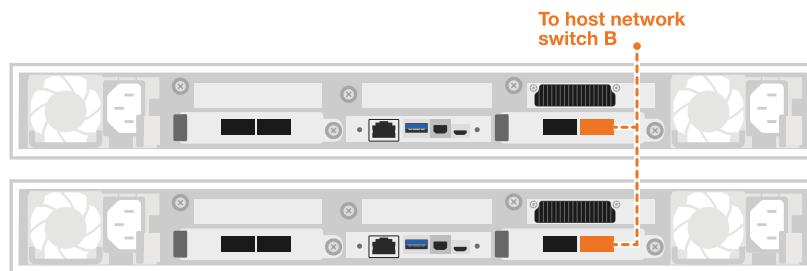




2. Verbinden Sie Port e5b der folgenden Data Compute Nodes mit dem Ethernet-Datennetzwerk-Switch B:

- Data Compute Node 1, Port e5b
- Data Compute Node 2, Port e5b

100GbE-Kabel



Schritt 2: Verkabeln Sie die Verbindungen des Data Compute Node-Clusters

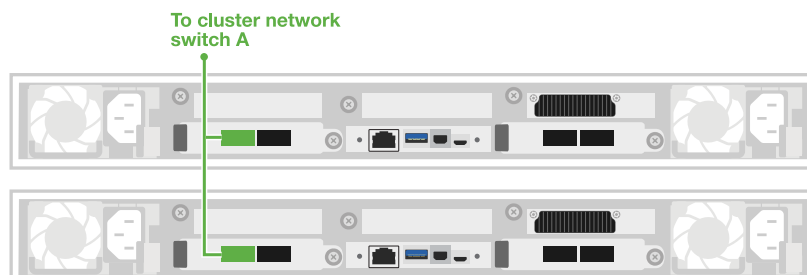
Für Datenrechenknoten verwenden Sie 4x100GbE-Breakout-Kabel, um die e4a/e5a-Ports für die Clusterverbindungen anzuschließen.

Schritte

1. Verbinden Sie Port e4a der folgenden Data Compute Node mit einem Nicht-ISL-Port am Cluster Netzwerk-Switch A:

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

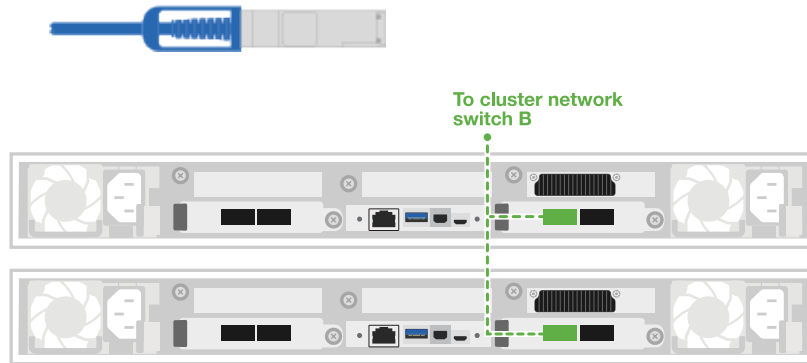
4x100GbE Breakout-Kabel



2. Verbinden Sie Port e5a der folgenden Data Compute Node mit einem Nicht-ISL-Port am Cluster Netzwerk-Switch B:

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

4x100GbE Breakout-Kabel



Was kommt als Nächstes?

Nachdem Sie die Hardware verkabelt haben, ["Schalten Sie Ihre Data Compute Node ein"](#).

Schalten Sie Ihre Datenrechenknoten für AI Data Engine ein

Nachdem Sie die Rack-Hardware installiert und Ihre Datenrechenknoten verkabelt haben, sollten Sie Ihre DCNs und Ihre Controller-Knoten für das AFX-Speichersystem einschalten, falls diese noch nicht eingeschaltet sind.

Bevor Sie beginnen

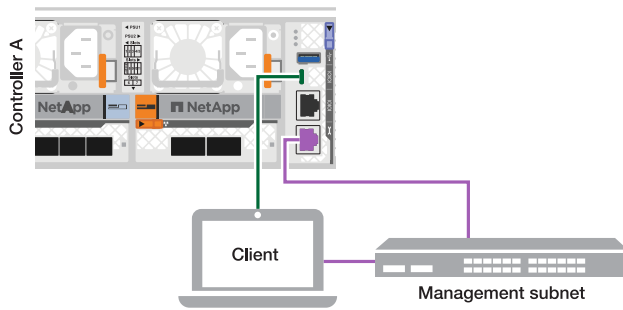
- Stellen Sie sicher, dass Ihre Regale eingeschaltet sind und jedem eine eindeutige Regal-ID zugewiesen wurde. Informationen zur Zuweisung von Regal-IDs für das AFX-Lagersystem finden Sie in der ["Dokumentation zur Vergabe eindeutiger Regal-IDs"](#).

Schritte

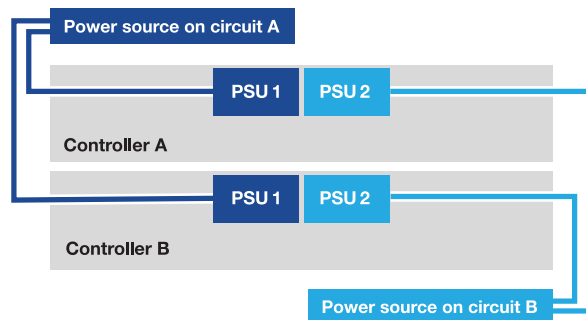
Nachdem Sie Ihre Storage Shelves eingeschaltet und die eindeutigen IDs zugewiesen haben, schalten Sie Ihre DCNs und die Storage Controller Nodes ein, falls diese noch nicht eingeschaltet sind.

1. Schließen Sie Ihren Laptop an den seriellen Konsolenanschluss an. Dadurch können Sie die Bootsequenz überwachen, wenn die Controller eingeschaltet werden.
 - a. Stellen Sie den seriellen Konsolenanschluss am Laptop auf 115.200 Baud mit N-8-1 ein.

Siehe die Online-Hilfe Ihres Laptops für Anweisungen zur Konfiguration des seriellen Konsolenanschlusses.
 - b. Verbinden Sie das Konsolenkabel mit dem Laptop, und verbinden Sie den seriellen Konsolenanschluss am Controller mit dem Konsolenkabel, das mit Ihrem Storage-System geliefert wurde.
 - c. Verbinden Sie den Laptop mit dem Switch im Management-Subnetz.



2. Weisen Sie dem Laptop eine TCP/IP-Adresse zu, wobei Sie eine verwenden, die sich im Management-Subnetz befindet.
3. Stecken Sie die Netzkabel in die Netzteile der Controller und verbinden Sie sie dann mit Stromquellen an verschiedenen Stromkreisen.



- Das System beginnt zu starten. Der erste Startvorgang kann bis zu acht Minuten dauern.
 - Die LEDs blinken auf und die Lüfter starten, was darauf hinweist, dass die Controller eingeschaltet werden.
 - Die Lüfter können beim Start laut sein, was normal ist.
4. Stecken Sie die Netzkabel in die Netzteile der Daten-Compute-Nodes und verbinden Sie sie dann mit Stromquellen an verschiedenen Stromkreisen.
 5. Sichern Sie die Netzkabel mit der Sicherungsvorrichtung an jedem Netzteil.
 6. Schalten Sie die Datenrechenknoten ein.

Um an den Netzschalter zu gelangen, müssen Sie möglicherweise die Blende entfernen; denken Sie in diesem Fall daran, sie anschließend wieder anzubringen.

Was kommt als Nächstes?

Nachdem Sie Ihre Datenrechenknoten eingeschaltet haben, ["ONTAP AIDE Cluster einrichten"](#).

Richten Sie Ihr AIDE System ein

AI Data Engine einrichten

Als ONTAP-Speicheradministrator können Sie AI Data Engine (AIDE) in Ihre AFX-Systembereitstellung integrieren. Sie können entweder AIDE im Rahmen einer erstmaligen AFX-Speichersystembereitstellung einrichten oder AIDE zu einem AFX-Cluster hinzufügen, der bereits bereitgestellt wurde und Daten bereitstellt. In beiden Fällen müssen Sie nach Abschluss der Basiseinrichtung auch die AIDE-Konfiguration

abschließen.

1. Richten Sie AIDE ein und integrieren Sie es in ein AFX-Speichersystem

Es gibt zwei Möglichkeiten, AIDE einzurichten und in ein AFX-Speichersystem zu integrieren. Wählen Sie die Option, die für Ihre Umgebung geeignet ist.



Bevor Sie AIDE einrichten, stellen Sie sicher, dass die Data Compute Nodes installiert und mit den Cluster-Switches verbunden sind. Siehe ["Installieren Sie AIDE"](#) für weitere Informationen.

AIDE mit einem neuen AFX-Cluster einrichten

Wenn Sie AIDE im Rahmen einer erstmaligen AFX-Systembereitstellung einrichten, ist der AIDE-Einrichtungsprozess Teil des Standard-Workflows für die AFX-Cluster-Einrichtung. Während der ONTAP-Cluster-Einrichtung erkennt System Manager automatisch die verbundenen DCNs und integriert sie in die Clusterkonfiguration. Danach sind Sie bereit, die AIDE-Konfiguration einschließlich der erforderlichen Lizenzkonfiguration abzuschließen.

AIDE mit einem vorhandenen AFX-Cluster einrichten

Wenn Sie einem bestehenden AFX-Speichersystem einen Cluster von Data Compute Nodes hinzufügen, müssen Sie die DCNs konfigurieren und die Einrichtung Ihres AIDE Clusters abschließen.

Fügen Sie die Data Compute Nodes zum Cluster hinzu

ONTAP erkennt dynamisch die neuen Data Compute Nodes, die mit Ihrem Cluster-Netzwerk verbunden sind, und zeigt sie im System Manager an. Sie können ["Fügen Sie Ihrem AIDE Cluster Datenrechenknoten hinzu"](#). Beachten Sie außerdem Folgendes:

- Die Data Compute Nodes müssen physisch mit den Cluster-Switches verbunden sein. Genau drei DCNs sind erforderlich, wenn ein AIDE Cluster erstellt wird.
- Sie müssen über ausreichend IP-Adressen für das DCN-Netzwerk und andere erforderliche Netzwerkkonfigurationsinformationen verfügen.

Schließen Sie die Einrichtung Ihres AIDE Clusters ab

Nach der Integration der Data Compute Nodes und dem Abschluss der ersten Netzwerkkonfiguration können Sie die Einrichtung Ihres neuen AIDE Clusters abschließen.

Über diese Aufgabe

Ein Willkommensfenster mit Anleitung wird auf der rechten Seite der Seite mit den notwendigen Konfigurationsschritten angezeigt. Ein Häkchen wird links neben jedem Schritt oder jeder Aktion angezeigt, die bereits abgeschlossen wurde. Ein Pfeil (→) wird für Elemente angezeigt, die noch nicht abgeschlossen wurden. Für jede der unvollständigen Aktionen:

- Wählen Sie den aktivierten Titellink aus und schließen Sie die Einrichtungsaktion für Ihre Umgebung ab.
- Wenn der Link deaktiviert ist, bewegen Sie den Mauszeiger über den Titel, um die erforderliche Aktion zum Aktivieren anzuzeigen und mit der Einrichtung fortzufahren.

Jeder Schritt enthält, sofern möglich, einen Link zu weiterführender Dokumentation. Einige Konfigurationseinstellungen sind möglicherweise bereits in der AFX-Cluster-Einrichtung enthalten. Siehe die ["AFX-Dokumentation"](#) für weitere Informationen.

Schritte

1. Wählen Sie **Link-Aggregation-Gruppen und VLANs konfigurieren**.

Falls für Ihre Umgebung erforderlich, können Sie LAGs und VLANs auf Clusterebene einrichten. Durch Auswahl dieser Option wird die Ethernet-Ports-Seite geöffnet.

2. Wählen Sie **Netzwerkprotokolle konfigurieren** aus.

Sie können die Storage VM (SVM) Netzwerkprotokolle konfigurieren.

3. Wählen Sie **Update data compute software** aus.

Wenn ein Data Compute Software-Image zur Aktualisierung verfügbar ist, wird diese Option aktiviert. Wählen Sie diese Option, um die Data Compute Software-Update-Seite zu öffnen. Weitere Informationen finden Sie unter ["Aktualisieren Sie die AI Data Engine software"](#).

4. Wählen Sie **Configure data engine networking**.

Diese Option zeigt die Seite zur Konfiguration der AI Data Engine-Netzwerkadresse an.

5. Wählen Sie **Intercluster-Netzwerkschnittstellen erstellen**

Diese Option öffnet die Seite zur Konfiguration der Intercluster-Netzwerkschnittstellen als Teil eines zweistufigen Verfahrens, wenn Sie beabsichtigen, Remote-Daten zu katalogisieren.

6. Wählen Sie **Peer mit anderen ONTAP storage systems**.

Nach der Erstellung der Intercluster-Netzwerkschnittstellen können Sie Peering-Beziehungen mit anderen ONTAP Storage-Systemen herstellen.

7. Wählen Sie **Add a data container**.

Sie können ein Volume zur Verwendung durch AIDE hinzufügen und es einem bestimmten SVM zuordnen.

8. Wählen Sie **Add Workspaces**.

Dadurch wird die Data Engine-Arbeitsbereiche-Seite geöffnet, auf der Sie ["einen Arbeitsbereich erstellen"](#) können.

9. Wählen Sie **Configure OpenID Connect**.

Sie müssen ["OpenID-Authentifizierung konfigurieren"](#) aktivieren, um den Zugriff auf die AIDE Console zu ermöglichen.

Nachdem Sie fertig sind

Nachdem alle Aufgaben erledigt sind, wird das Hilfefenster ausgeblendet. Sie können es bei Bedarf manuell schließen und wieder öffnen.

2. AIDE-Konfiguration abschließen

Nachdem Ihr AIDE Cluster eingerichtet und in das AFX-Speichersystem integriert wurde, müssen Sie die AIDE-Konfiguration abschließen.

Schritte

1. "Fügen Sie die erforderlichen AIDE-Lizenzen hinzu".

Es gibt eine AI Data Engine-Lizenz, die Sie für den vollen Funktionsumfang von AIDE, einschließlich Vektorisierungs- und Schutzplankenfunktionen, installieren müssen.

2. "OIDC-Authentifizierung konfigurieren".

Stellen Sie sicher, dass OIDC für alle Bereitstellungen konfiguriert ist. Sie müssen OIDC konfigurieren, um auf die AI Data Engine Console zuzugreifen.

Installieren Sie eine AIDE-Lizenz in Ihrem AFX-System

Sie müssen eine NetApp-Lizenz installieren, um auf die vollständigen Funktionen der AI Data Engine (AIDE) zugreifen zu können. Als ONTAP-Clusteradministrator können Sie die Lizenzverwaltung mit ONTAP System Manager durchführen.

Bereiten Sie die Lizenzierung der AI Data Engine vor

Eine Lizenz ist ein Nachweis für eine oder mehrere Softwareberechtigungen. Alle AFX-Lizenzen werden als NetApp Lizenzdatei (NLF) bereitgestellt, bei der es sich um eine einzelne Datei handelt, die mehrere Funktionen aktiviert. Es gibt mehrere Dinge, die Sie beachten sollten, bevor Sie eine Lizenz auf Ihrem AFX-Speichersystem zur Unterstützung von AIDE installieren.

Lizenzarten

Es gibt zwei Haupttypen von Lizenzen, die benötigt werden, um AIDE über Ihr AFX-Speichersystem zu verwenden.

ONTAP One-Lizenz

Die Metadata Engine Basic-Lizenz ist typischerweise werkseitig als Teil der ONTAP One-Lizenz installiert. Sie ermöglicht den Zugriff auf die Metadata Engine-Funktionalität, die die wesentliche Grundlage für AIDE-Operationen bietet. Alle Kernfunktionen, die zur Verwaltung Ihres ONTAP-Systems benötigt werden, sind ebenfalls enthalten.

AI Data Engine

Sie müssen die AI Data Engine-Lizenz erwerben und installieren, um auf die Premium-Dienste zuzugreifen, die zur Aktivierung der vollständigen Funktionen von AIDE erforderlich sind. Die Lizenz schaltet Ihre Datenrechenknoten frei und ermöglicht AI-Funktionen wie Vektorisierung, Governance Guardrails, Inferenz und eine integrierte UI-Erfahrung. Die Lizenz beinhaltet eine GPU-Anzahl sowie ein Ablaufdatum.

Lizenzinstallationsanforderungen

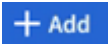
Sie müssen die AIDE-Lizenz erwerben und die zugehörige NLF-Datei auf Ihr lokales System herunterladen. Anschließend können Sie die Datei über den System Manager in Ihr AFX-Speichersystem hochladen. Stellen Sie außerdem sicher, dass Sie Folgendes haben:

- Administratoranmeldeinformationen für die Anmeldung bei ONTAP System Manager
- Ein AFX-Cluster mit ONTAP 9.18.1 und höher

Installieren Sie eine Lizenz auf Ihrem AFX-System

Sie können eine AIDE-Lizenz installieren, um die zusätzlichen AIDE-Funktionen zu aktivieren, die für Ihr AFX Storage System benötigt werden.

Schritte

1. Im System Manager wählen Sie **Cluster** und dann **Einstellungen**.
2. Neben **Lizenzen** wählen Sie ➔.
3. Wählen Sie die Registerkarte **Features**, um die verfügbaren ONTAP Features anzuzeigen.
4. Um eine Lizenz zu installieren, wählen Sie die Registerkarte **Installed licenses**.
5. Wählen Sie  **Add**.
6. Wählen Sie eine lokale Lizenzdatei aus und wählen Sie **Hinzufügen**.

Verwandte Informationen

- ["ONTAP-Lizenzierungsübersicht"](#)
- ["So laden Sie NLF-Lizenzen von der NetApp Support Site herunter"](#)
- ["ONTAP CLI: system license add"](#)

OpenID Connect für AIDE in ONTAP konfigurieren

Als ONTAP-Clusteradministrator können Sie mit ONTAP System Manager die OpenID Connect (OIDC) Authentifizierung für einen AI Data Engine (AIDE) Cluster konfigurieren. Dies bietet eine sichere und zentrale Anmeldung über einen externen Identitätsanbieter (IdP).



Sie müssen OIDC konfigurieren, um auf die AI Data Engine Console zuzugreifen. Wenn OIDC konfiguriert ist, erfolgt die gesamte Authentifizierung über OIDC. Wenn OIDC nicht konfiguriert ist, steht die Konsole Administratoren sowie Data Engineers und Data Scientists nicht zur Verfügung. In diesem Fall wird die Anmeldung am System Manager auf die lokale Authentifizierung zurückgesetzt.

Beachten Sie außerdem Folgendes zur OIDC-Konfiguration für AIDE-Zugriff:

- Sie können eine bestehende OIDC-Konfiguration nicht ändern. Wenn Sie eine Änderung vornehmen müssen, löschen Sie zuerst die Konfiguration und erstellen Sie eine neue mit den gewünschten Einstellungen.
- Wenn Sie OIDC deaktivieren oder entfernen, wird System Manager auf die lokale ONTAP Benutzerauthentifizierung zurückgesetzt.

OIDC-Übersicht

OpenID Connect (OIDC) ist ein Authentifizierungsprotokoll, das auf dem OAuth 2.0-Framework basiert. Es erweitert OAuth 2.0, das primär für die Autorisierung verwendet wird, um eine Identitätsschicht. OIDC führt das Konzept eines ID-Tokens ein, eines JSON Web Tokens (JWT), das Angaben über das Authentifizierungsereignis und die Identität des Benutzers enthält.

Sie müssen einen externen Identitätsanbieter (IdP) auswählen und konfigurieren, der von AFX mit AIDE unterstützt wird. Der IdP authentifiziert Benutzer und stellt Token aus, die AFX über System Manager verwenden kann, um Zugriff auf die AIDE Console zu gewähren.

Konfigurieren Sie Identitätsanbieter von Drittanbietern

Um sich mit OIDC zu authentifizieren, müssen Sie zunächst einen externen IdP konfigurieren. Die ONTAP-Implementierung von OIDC verwendet Rollenansprüche in den Token, um RBAC durchzusetzen. Stellen Sie bei der Einrichtung eines IdP sicher, dass dieser so konfiguriert ist, dass er Rollenansprüche im id token und access token zurückgibt. ONTAP unterstützt zwei IdPs für OIDC-Authentifizierung: Entra ID und Active Directory Federation Services (AD FS).

Entra ID

Sie können Entra ID mit den folgenden Schritten auf hoher Ebene konfigurieren:

1. Erstellen Sie eine neue App Registration auf der Entra ID-Konfigurationsseite.
2. Setzen Sie den Wert für Redirect URI (Web) auf `https://$CLUSTER_MGMT_IP/oidc/callback`, wobei Sie die entsprechende Cluster-Management-IP-Adresse oder den FQDN einsetzen.
3. Erstellen Sie die erforderlichen Rollen unter App Roles und weisen Sie diese Ihren Benutzern zu.
4. Aktualisieren Sie die Token-Ansprüche unter Token Configuration, um Rollen in id-token und access-token zurückzugeben.

Siehe ["Richten Sie einen OpenID Connect-Anbieter mit Entra ID ein"](#) für weitere Informationen.

Active Directory Federation Services

Sie können AD FS mit den folgenden Schritten auf hoher Ebene konfigurieren:

1. Erstellen Sie eine neue Anwendungsgruppe und wählen Sie **Server application accessing a web API** aus.
2. Setzen Sie den Wert für Redirect URI (Web) auf `https://$CLUSTER_MGMT_IP/oidc/callback`, wobei Sie die entsprechende Cluster-Management-IP-Adresse oder den FQDN einsetzen.
3. Konfigurieren Sie Claims so, dass Rollen in Tokens zurückgegeben werden.

Siehe ["AD FS als OpenID Connect-Identitätsanbieter hinzufügen"](#) für weitere Informationen.

OIDC in System Manager konfigurieren

Nach der Konfiguration Ihres IdP können Sie die OIDC-Authentifizierung im System Manager einrichten, um einen sicheren Zugriff auf die AIDE Console zu ermöglichen.

Bevor Sie beginnen

- Sie benötigen Administratorzugriff auf System Manager.
- Ihr OIDC-Identitätsanbieter muss konfiguriert und erreichbar sein.

Schritte

1. Im System Manager wählen Sie **Cluster** und dann **Einstellungen** aus; suchen Sie die OpenID Connect card.
2. Wenn OIDC bereits konfiguriert ist, können Sie die Konfiguration bearbeiten oder deaktivieren. Wenn OIDC nicht konfiguriert ist, wählen Sie  aus, um den Einrichtungsprozess zu starten.
3. Geben Sie unter „OpenID Connect konfigurieren“ Werte für die folgenden Felder an:
 - Anbieter
 - Emittent

- JSON Web Key Set URI
 - Autorisierungsendpunkt
 - Token-Endpunkt
 - Endpunkt zum Beenden der Sitzung
 - Access Token Issuer (optional)
4. Geben Sie unter Client configuration Werte für die folgenden Felder an:
- Client-ID
 - Remote-Benutzeranspruch
 - Aktualisierungsintervall
5. Geben Sie unter Connection details Werte für die folgenden Felder an:
- Cluster-IP-Adresse oder FQDN
 - Ausgehender Proxy (optional)
6. Unter „Externe Rollenzuordnung“ wählen Sie eine vorhandene Rollenzuordnung aus oder definieren eine neue Rolle für den ONTAP `admin` Benutzer.
7. Wählen Sie **Jetzt aktivieren** und dann **Speichern**. System Manager wird aktualisiert, um die neuen Authentifizierungseinstellungen anzuwenden.
8. Melden Sie sich mit Ihren IdP-Zugangsdaten an; nach erfolgreicher Authentifizierung werden Sie zum System Manager zurückgeleitet.

Verwandte Informationen

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

Arbeitsbereiche einrichten

Daten für AI Data Engine vorbereiten

Nachdem Sie einen Cluster erstellt haben, richten Sie einen Datencontainer ein, der die Daten enthält, die Sie mit AI Data Engine (AIDE) verwenden möchten. Dieser Datencontainer muss ein ONTAP Volume sein, entweder ein lokales Volume oder ein Volume aus einem verbundenen ONTAP Cluster, das ONTAP 9 ausführt.

Anstatt ONTAP-Clusterdaten manuell in AIDE hochzuladen, müssen Sie die Zielcluster und SVMs mit dem AIDE Cluster peeren und dann entscheiden, welche NFS-Volumes Sie mit dem AIDE-Metadatenkatalog verwenden möchten. Nachdem Sie den Datencontainer erstellt haben, können Sie einen Arbeitsbereich erstellen und den Datencontainer mit diesem Arbeitsbereich verknüpfen. Benutzer dieses Arbeitsbereichs können dann auf die mit dem Arbeitsbereich verbundenen Datensammlungen und Ressourcen für ihre AI-Workloads zugreifen und mit ihnen interagieren.

Über diese Aufgabe

Sie müssen jede SVM, die die Daten enthält, die Sie mit AI Data Engine verwenden möchten, peeren. Das Peeren nur des Clusters ist nicht ausreichend. Dadurch wird sichergestellt, dass die AI Data Engine wie vorgesehen auf Ihre Daten zugreifen, sie integrieren und indizieren kann.

Sie sollten die SVM im AIDE Cluster, die als SnapMirror Ziel fungieren wird, per Peering verbinden.

Sie müssen keine SnapMirror-Beziehungen zwischen den ONTAP Clustern und dem AIDE Cluster herstellen. Diese Beziehungen werden während der Erstellung des Arbeitsbereichs automatisch erstellt.

Bevor Sie beginnen

- Sie benötigen *Storage-Administrator*-Rechte, um Cluster und SVMs zu verbinden und Datencontainer auszuwählen.
- Sie haben die ONTAP Cluster und SVMs identifiziert, die die Daten enthalten, die Sie mit AI Data Engine verwenden möchten.
- Sie haben bestätigt, dass die Datenquellen-Volumes die folgenden Anforderungen erfüllen:
 - Volumes sind online und zugänglich.
 - Das NFS-Protokoll ist aktiviert. Nur NFS-aktivierte Volumes werden als Datencontainer für AI Data Engine unterstützt. SMB- und CIFS-Volumes werden nicht unterstützt.
 - Volumes sind keine FlexCache Volumes.
 - Datenquellen sind Lese-/Schreibvolumes. Datenschutzvolumes werden nicht unterstützt.

Schritte

1. "[Verbinden Sie jeden ONTAP Cluster und jede SVM als Peers](#)" die Daten enthält, die Sie mit AI Data Engine verwenden möchten.
2. "[Wählen Sie die Volumes aus, die Sie mit AI Data Engine verwenden möchten](#)".

Notieren Sie für jeden Volume die folgenden Informationen:

- Volumenname
- UUID

- SVM-Name und UUID
- Cluster name und UUID

Was kommt als Nächstes?

"[Einen Arbeitsbereich erstellen](#)" und ordnen Sie die von Ihnen erstellten Datencontainer einem Arbeitsbereich zu.

Verwandte Informationen

- "[Optionen für die Datenmigration](#)"

Erstellen Sie einen Arbeitsbereich in AI Data Engine

Nachdem Sie einen Cluster eingerichtet haben, können Sie einen Arbeitsbereich erstellen. Arbeitsbereiche ermöglichen es Ihnen, Daten im Cluster zu segmentieren, den Datenzugriff für einzelne Benutzer zu steuern und Daten auszuschließen, auf die AI Data Engine (AIDE) nicht zugreifen soll.

Wenn Sie Speicher verwalten, verwenden Sie ONTAP System Manager, um Arbeitsbereiche zu erstellen und zu verwalten.

Organisationen erstellen Arbeitsbereiche basierend auf Teams, Projekten, Datensensibilitätsstufen oder anderen relevanten Kriterien. Beispielsweise könnten Sie im Gesundheitswesen klinische Daten in einem Arbeitsbereich segmentieren, aber Daten aus den Bereichen IT, Recht oder anderen Abteilungen ausklammern.

Über diese Aufgabe

Systemverarbeitungslimits beeinflussen die Erstellung von Arbeitsbereichen (typischerweise bis zu 15 GB pro Tag und Cluster). Wenn Sie mehrere Arbeitsbereiche parallel oder in schneller Folge erstellen, kann die Verarbeitung jedes Arbeitsbereichs länger dauern, und Sie könnten erhebliche Verzögerungen erleben.

Überwachen Sie den Status der Arbeitsbereichserstellung auf der Arbeitsbereichsinventarseite. Für beste Ergebnisse vermeiden Sie es, viele Arbeitsbereiche gleichzeitig zu erstellen, wenn Sie sofortigen Zugriff auf diese Funktionen benötigen.

Bevor Sie beginnen

- Sie benötigen *Storage Administrator*-Berechtigungen, um Arbeitsbereiche zu erstellen und Datensammlungen zuzuordnen.
- Sie haben die entfernten (per Peering verbundenen) und lokalen Datenquellen bestimmt, die Sie mit dem Workspace und mit AI Data Engine verwenden möchten.
- Sie haben "[mindestens einen Datencontainer erstellt](#)", die der Arbeitsbereich nutzen kann, wie z. B. ein lokales Volume oder ein Volume aus einem verbundenen Cluster.



Fügen Sie einem Arbeitsbereich ein Volume hinzu, das Sie während der erwarteten Lebensdauer dieses Arbeitsbereichs nicht löschen werden. Wenn Sie ein Volume nach dem Hinzufügen zu einem Arbeitsbereich löschen, wird der Arbeitsbereich in einen Fehlerzustand versetzt. Bestätigen Sie die längerfristige Nutzbarkeit des Volumes, bevor Sie einen Arbeitsbereich einrichten.

- Stellen Sie sicher, dass NFS auf dem Volume aktiviert ist, aber dass CIFS nicht aktiviert ist. Workspaces unterstützen nur Volumes mit NFS. Volumes mit CIFS (SMB) werden nicht unterstützt.

Einen Arbeitsbereich erstellen

Erstellen Sie einen Arbeitsbereich und verknüpfen Sie Datencontainer, die die Daten enthalten, die Sie mit AI Data Engine verwenden möchten.

Schritte

1. Navigieren Sie im ONTAP System Manager zu **Data Engine > Workspaces**.
2. Wählen Sie **Add**.
3. Im Dialogfeld **Arbeitsbereich hinzufügen** wählen Sie mindestens einen verfügbaren Datencontainer aus, um ihn mit dem Arbeitsbereich zu verknüpfen.
4. Konfigurieren Sie "**verbundene Cluster**" so, dass auf die Daten dieser Cluster innerhalb des Arbeitsbereichs zugegriffen werden kann.
5. Wenn Sie den Benutzerzugriff auf den Arbeitsbereich konfigurieren möchten, können Sie dies jetzt tun oder "**Warten Sie, bis der Arbeitsbereich erstellt ist**".
6. Konfigurieren Sie ein Aktualisierungsintervall, wie oft der Arbeitsbereich mit den zugehörigen Datencontainern synchronisiert wird, um neue oder aktualisierte Daten zu erfassen (zum Beispiel sechs Stunden).



Wählen Sie ein Intervall, das ein ausgewogenes Verhältnis zwischen Datenaktualität und Systemleistung gewährleistet. Wenn Sie einen Datencontainer mehreren Arbeitsbereichen hinzufügen, verwendet das System automatisch das aggressivste (kürzeste) Intervall. Weitere Informationen finden Sie in der Dokumentation zu [Aktualisierung und Versionierung von Arbeitsbereichen](#).

7. Wählen Sie **Continue**.
8. Geben Sie im Dialogfeld **Finalize workspace** einen Arbeitsbereichsnamen und eine Beschreibung ein.
9. Wählen Sie **Add**, um den Arbeitsbereich zu erstellen.

Ergebnis

Der Arbeitsbereichserstellungsprozess dauert je nach zugehörigem Datensatz und dessen Dateianzahl, Dateigröße und anderen Faktoren mehrere Minuten bis Stunden.

Das System extrahiert automatisch Metadaten für alle Datenquellen und speichert sie in einem Metadatenkatalog, mit dem Benutzer die für ihre Projekte benötigten Dateien finden können. Nachdem Sie Benutzer dem Arbeitsbereich zugewiesen haben, können Dateningenieure die zum Arbeitsbereich gehörenden Komponenten über die AI Data Engine Console einrichten und mit ihnen interagieren.

Der neue Arbeitsbereich wird auf der Seite „Arbeitsbereiche“ im `Creating` Status angezeigt, bis der Prozess abgeschlossen ist und sich der Status in `ready` ändert.

Arbeitsbereichsdetails überprüfen

Nach der Erstellung des Arbeitsbereichs überprüfen Sie die Details des Arbeitsbereichs.

Schritte

1. Überprüfen Sie die Details des Arbeitsbereichs, einschließlich der Gesamtgröße, des Prozentsatzes der genutzten Clusterkapazität und des Datums der letzten Aktualisierung des Arbeitsbereichs.
2. Wählen Sie den Arbeitsbereichsnamen aus, um die Detailseite zu öffnen.
3. Auf der Registerkarte „Übersicht“ können Sie Details zum Arbeitsbereich einsehen, darunter zugehörige

Aktualisierungen und Versionierung von Arbeitsbereichen

Bei jeder Aktualisierung eines Arbeitsbereichs wird eine unveränderliche Version erstellt, die den aktuellen Zustand aller Dateien und Objekte im Arbeitsbereich erfasst. Versionen enthalten vollständige Metadaten, Verweise auf während der Extraktion verwendete Snapshots und eine Job-ID zur Nachverfolgbarkeit. Dies unterstützt Datenherkunft, Reproduzierbarkeit und Auditierung.

Aktualisierungen erfolgen entweder nach dem von Ihnen konfigurierten Zeitplan (z. B. alle sechs Stunden) oder wenn Sie sie manuell auslösen. Das minimale unterstützte Aktualisierungsintervall beträgt eine Stunde; das maximale beträgt ein Jahr. Wenn ein Datencontainer in mehreren Arbeitsbereichen enthalten ist, verwendet das System für die Planung der Metadatenextraktion das häufigste, kürzeste Aktualisierungsintervall.

Standardmäßig speichert das System vorherige, aktuelle und nächste (in Bearbeitung befindliche) Versionen. Das System speichert ältere Versionen gemäß den Richtlinien Ihrer Organisation und kann sie bei Bedarf entfernen.

Sie können alle Versionen eines Arbeitsbereichs auflisten und die Unterschiede zwischen den Versionen anzeigen, um festzustellen, welche Dateien oder Objekte hinzugefügt, geändert oder gelöscht wurden. So können Sie Änderungen im Zeitverlauf nachverfolgen und die Entwicklung Ihrer Arbeitsbereichsdaten nachvollziehen.

Benutzerzugriff auf AI Data Engine Arbeitsbereiche zuweisen

Als Speicheradministrator weisen Sie Benutzern je nach Rolle – Dateningenieure, Datenwissenschaftler oder andere Rollen, abhängig von der Struktur und den Bedürfnissen Ihrer Organisation – einen Arbeitsbereich zu. Benutzer melden sich mit ihren Zugangsdaten an der AI Data Engine (AIDE) Console an und greifen auf die Datencontainerressourcen innerhalb des zugewiesenen Arbeitsbereichs zu.

ONTAP System Manager ermöglicht es Ihnen, zu verwalten, welche Benutzer Zugriff auf AI Data Engine Arbeitsbereiche haben. Sie können Benutzer hinzufügen oder entfernen, um zu steuern, wer die Arbeitsbereichsdaten und -aktivitäten anzeigen, bearbeiten und damit interagieren kann.

Bevor Sie beginnen

- Sie benötigen *storage administrator*-Berechtigungen, um den Benutzerzugriff auf Arbeitsbereiche zu verwalten.
- Bestätigen Sie, dass Sie den Arbeitsbereich erstellt haben und dass er im Arbeitsbereichsinventar aktiv ist.
- Vergewissern Sie sich, dass alle relevanten Datencontainer dem Arbeitsbereich hinzugefügt wurden und zugänglich sind.
- ["Bestätigen Sie, dass OIDC für den Cluster aktiviert und konfiguriert ist"](#). Die Rollenzuordnung von IdP zu ONTAP-Rollen muss für jeden relevanten Data Engineer und Data Scientist IdP-Benutzer bzw. jede Gruppe abgeschlossen werden.

Schritte

- Einen Benutzer zu einem Arbeitsbereich hinzufügen:
 - a. Navigieren Sie im ONTAP System Manager zu **Data Engine > Workspaces**.

- b. Wählen Sie den Arbeitsbereichsnamen aus, um die Detailseite zu öffnen.
- c. Wechseln Sie zur Registerkarte „Benutzer“.
- d. Wählen Sie die Schaltfläche **Hinzufügen** aus, um den Dialog zum Hinzufügen von Benutzern zu öffnen.
- e. Geben Sie die Details eines oder mehrerer Benutzer ein. Geben Sie die Details als durch Kommas getrennte Liste von OIDC-Benutzern ein.
- f. Wählen Sie **Add**, um dem Benutzer Zugriff auf den Arbeitsbereich zu gewähren.
- Einen Benutzer aus einem Arbeitsbereich entfernen:
 - a. Suchen Sie auf der Registerkarte „Benutzer“ der Workspace-Detailseite den Benutzer, den Sie entfernen möchten, und wählen Sie ihn aus.
 - b. Wählen Sie die Schaltfläche **Remove**.
 - c. Bestätigen Sie die Entfernung im Dialog.
 - d. Das System entfernt den Benutzer umgehend und entzieht ihm den Zugriff auf den Arbeitsbereich.

Ergebnis

Nur Benutzer, die auf der Registerkarte „Benutzer“ des Arbeitsbereichs aufgeführt sind, können auf die Daten und Aktivitäten des Arbeitsbereichs zugreifen und mit ihnen interagieren.

Verwalten und überwachen

Clusterprozesse überwachen

AIDE-System- und Clusterstatus anzeigen

Als Speicheradministrator können Sie mit ONTAP System Manager auf das Dashboard zugreifen und den Clusterstatus anzeigen. Dies ist ein guter erster Schritt, bevor Sie mit Ihren AIDE-Verwaltungsaufgaben beginnen oder wenn Sie ein Betriebsproblem vermuten.

Bevor Sie beginnen

- Sie benötigen *Storage-Administrator*-Berechtigungen, um administrative Aufgaben im Zusammenhang mit AIDE ONTAP durchzuführen.

Überwachen Sie den Zustand und die Kapazität von AIDE über das Dashboard.

1. Stellen Sie eine Verbindung zum ONTAP System Manager über die Cluster Management-Adresse her:

`https://$FQDN_OR_IP/`

2. Sign in mit einem Administratorkonto.
3. Wählen Sie **Dashboard** im linken Navigationsbereich.
4. Überprüfen Sie die Kachel **Health**:
 - Bestätigen Sie die allgemeine Cluster-Gesundheit.
 - Überprüfen Sie die Anzahl und den Status der **Data compute nodes**.
 - Auf Warnmeldungen prüfen:
 - DCN-Knotenprobleme oder Verbindungsprobleme
 - Arbeitsbereiche oder Datensammlungen mit Fehlern (zum Beispiel Fehler bei der Veröffentlichung von Sammlungen)
5. Überprüfen Sie die **Capacity**-Kachel:
 - Beachten Sie die Gesamtkapazität des Clusters und die genutzte Kapazität.
 - Für AIDE-Cluster überprüfen Sie Folgendes:
 - Von AIDE-Metadaten und Anwendungs-Volumes belegte Kapazität (metadata Storage VM)
 - Von Arbeitsbereichen und Datensammlungen genutzte Kapazität (falls verfügbar)
6. Optional können Sie die **Netzwerk**- und **Leistung**-Kacheln überprüfen, um das clusterweite Verhalten zu verstehen, das sich auf AIDE workloads auswirken könnte (zum Beispiel Netzwerküberlastung oder Schutzverzögerung).

DCN-Gesundheits- und Nutzungsdaten anzeigen

1. Wählen Sie im Navigationsbereich **Cluster** und anschließend **Übersicht** aus.
2. Wählen Sie die Registerkarte **Data compute** aus.

Diese Registerkarte zeigt alle DCN-Knoten im Cluster mit:

- Knotenname, Modell, Seriennummer und Softwareversion
 - Gesamtknotenzustand
 - CPU- und Speicherauslastung
 - GPU-Auslastung (falls GPUs vorhanden sind)
 - Alle Fehlerindikatoren auf Knotenebene
3. Erweitern Sie einen DCN node, um die Detailansicht zu öffnen und zu überprüfen:
 - System-CPU- und Speicherauslastung
 - GPU-Speichernutzung
 - Gemeldete Hardware- oder Serviceprobleme
 4. Wählen Sie auf der Seite **Cluster > Übersicht Verkabelung** aus, um zu überprüfen, ob DCN-Knoten korrekt mit den Cluster-Switches verkabelt sind und um etwaige Port- oder Verbindungsprobleme zu identifizieren.

Arbeitsbereiche und Metadaten-Fußabdruck überwachen

1. Wählen Sie im Navigationsbereich **Data engine** und anschließend **Workspaces** aus.
2. Überprüfen Sie die Zusammenfassung des Arbeitsbereichs oben auf der Seite:
 - Anzahl der Arbeitsbereiche und deren Zustände (zum Beispiel `Processing`, `Healthy`, `Error`).
 - Gesamtgröße des Arbeitsbereichs.
 - Prozentsatz der von allen Arbeitsbereichen verbrauchten Clusterkapazität.
3. Überprüfen Sie das Arbeitsbereichsraster:
 - Bestätigen Sie, dass die kritischen Arbeitsbereiche den Status **Healthy** anzeigen.
 - Überprüfen Sie die Arbeitsbereichsgrößen und den Kapazitätsverbrauch.
 - Suchen Sie nach Arbeitsbereichen in `Error` oder lang andauernden `Processing` Zuständen.
4. Um die Details für einen bestimmten Arbeitsbereich anzuzeigen, wählen Sie dessen Namen aus:
 - Bestätigen Sie auf der Registerkarte **Overview** Folgendes:
 - Arbeitsbereichsstatus und -größe
 - Eingeschlossene Datencontainer (Volumes) und deren Elementanzahl
 - Letzte Aktualisierungszeit für jede Datenquelle
 - Bestätigen Sie auf der Registerkarte **Data collections** Folgendes:
 - Welche Datensammlungen existieren für diesen Arbeitsbereich (Datensammlungen sind im System Manager schreibgeschützt)
 - Ihr Status, ihre Größe und der Zeitpunkt der letzten Aktualisierung
 - Überprüfen Sie auf der Registerkarte **Benutzer**, welche Benutzer der AI Data Engine Console Zugriff haben.

Metadaten Storage-VM und von AIDE verwalteten Schutz überwachen

1. Wählen Sie im Navigationsbereich **Cluster** und anschließend **Storage VMs** aus.
2. Suchen Sie die Storage VM mit Subtyp `data-engine` (die Metadata SVM):
 - Bestätigen Sie, dass die Metadaten-SVM online ist.

- Optional können Sie die Details öffnen, um die Zähler für Folgendes anzuzeigen:
 - Volumes
 - LIFs mit Typ `Data compute network` (verwendet für DCN-ONTAP-Kommunikation)
- 3. Wählen Sie **Schutz** und dann **Beziehungen**, um den Schutz für Remote-Datenquellen anzuzeigen, die in Arbeitsbereichen verwendet werden:
 - Identifizieren Sie von AIDE erstellte SnapMirror-Beziehungen anhand des Namensmusters:
 - Zielvolumen: `<source_volume_name>_dest_<source_volume_UUID>`
 - Richtlinie: `<source_volume_name>_dest_aide_policy_<source_volume_UUID>`
 - Nutzen Sie diese Ansicht, um zu überprüfen, ob die Beziehungen intakt sind und ob die Verzögerungszeit den Erwartungen an die Aktualisierung des Arbeitsbereichs entspricht.



Ändern Sie die Metadaten-Storage-VM, die von AIDE erstellten SnapMirror-Beziehungen oder die von AIDE verwalteten Snapshots (oder deren Zeitpläne) nicht direkt in ONTAP. Änderungen können den AIDE-Versionverlauf beeinträchtigen. "[Arbeitsbereichsaktualisierungseinstellungen anpassen](#)" wenn Sie das Aktualisierungsverhalten anpassen müssen.

AIDE-bezogene Warnungen und Benachrichtigungen prüfen

1. Im Navigationsbereich wählen Sie **Events & Jobs** und anschließend **System alerts**.
2. Überprüfen Sie alle aktiven Warnmeldungen im Zusammenhang mit:
 - DCN node health oder Konnektivität
 - Netzwerkprobleme der Data Engine
 - Fehler im Arbeitsbereich oder bei der Datenerfassung
 - Softwareversionskonflikte zwischen ONTAP und DCN cluster
3. Konfigurieren Sie bei Bedarf Benachrichtigungsziele (zum Beispiel E-Mail, Syslog) in **Cluster > Settings > Notification management**, um sicherzustellen, dass AIDE-bezogene Warnmeldungen an Ihre Operations-Tools weitergeleitet werden.

Verwandte Informationen

["Bereiten Sie sich auf die Administration Ihres AFX-Speichersystems vor"](#)

Insights anzeigen, um Ihr AIDE System zu optimieren

Als Speicheradministrator können Sie die Funktion „Insights“ des ONTAP System Managers nutzen, um empfohlene Konfigurationsaktualisierungen anzuzeigen, die den NetApp Best Practices entsprechen. Diese Änderungen können die Sicherheit und Leistung Ihres AIDE Clusters optimieren.

Über diese Aufgabe

Die einzelnen Erkenntnisse werden als separate Kachel oder Karte auf der Seite dargestellt, die Sie umsetzen oder verwerfen können. Sie können außerdem den zugehörigen Dokumentationslink auswählen, um mehr über eine bestimmte Technologie zu erfahren.

Schritte

1. Im System Manager wählen Sie **Analysis** und dann **Insights**.

2. Prüfen Sie die verfügbaren Empfehlungen.

Was kommt als Nächstes

Führen Sie eine der empfohlenen Maßnahmen durch, um bewährte Konfigurationspraktiken umzusetzen.

AIDE-Systemereignisse, Jobs und Revisionsprotokoll

Als Speicheradministrator können Sie die von AIDE generierten Ereignisse, Jobs und Revisionsprotokollnachrichten überprüfen, um interne Prozesse zu verfolgen und potenzielle Probleme zu diagnostizieren. Das AIDE System kann so konfiguriert werden, dass diese Informationen zusammen mit anderen zugehörigen Daten zur weiteren Verarbeitung und Archivierung weitergeleitet werden.

Bevor Sie beginnen

- Sie benötigen *Storage-Administrator*-Berechtigungen, um administrative Aufgaben im Zusammenhang mit AIDE ONTAP durchzuführen.

Überwachen Sie AIDE-Aktivitäten, -Ereignisse und -Jobs

Sie können die zentrale **Activity**-Ansicht verwenden, um AIDE-spezifische Ereignisse und Aufgaben in allen Arbeitsbereichen zu überwachen oder Aktivitäten zu überprüfen, die auf einzelne Arbeitsbereiche beschränkt sind.

Clusterweite AIDE Aktivität anzeigen

Überwachen Sie Workspace-Operationen, beheben Sie Probleme bei der Metadatenextraktion und verfolgen Sie die Veröffentlichung der Datenerfassung in Ihrer gesamten AIDE Deployment.

1. Wählen Sie im ONTAP System Manager im Navigationsbereich **Data engine** und anschließend **Activity** aus.
2. Wählen Sie die Registerkarte **Events** aus:
 - Überprüfen Sie aktuelle AIDE-spezifische Ereignisse wie die folgenden:
 - Erstellung, Aktualisierung oder Löschung eines Arbeitsbereichs
 - Operationen zum Hinzufügen/Entfernen von Datencontainern
 - Veröffentlichung der Datenerfassung (falls vorhanden)
 - Verwenden Sie Filter (nach Schweregrad, Objekttyp, Workspace oder Zeitbereich), um sich auf aktive oder kritische Ereignisse zu konzentrieren.
3. Wählen Sie ein einzelnes Ereignis aus, um es zu öffnen und zu überprüfen:
 - Beschreibung und Zeitstempel
 - Betroffener Arbeitsbereich, Datenerfassung oder Datenquelle
 - Empfohlene Aktion, falls angegeben
4. Wählen Sie die Registerkarte **Jobs** aus:
 - Überwachen Sie langlaufende Jobs wie:
 - Erste Metadatenextraktion für einen Arbeitsbereich
 - Arbeitsbereichsaktualisierung / Katalogaktualisierung Jobs

- Datenerfassungs-, Veröffentlichungs- oder Aktualisierungsjobs
 - Überprüfen Sie den Jobstatus und den Fortschritt.
5. Wählen Sie einen Auftrag aus, um die Peek-Ansicht zu öffnen und ihn zu überprüfen:
- Start- und Endzeit
 - Fortschrittsprozentsatz und Phase (zum Beispiel, **Scannen**, **Veröffentlichen**)
 - Betroffener Arbeitsbereich, Datenerfassung oder Datenquelle
 - Fehlermeldungen für fehlgeschlagene Jobs

Arbeitsbereichsspezifische Aktivitäten anzeigen

Um Probleme in einem bestimmten Workspace zu beheben, öffnen Sie die Workspace-Details (**Data engine > Workspaces**, wählen Sie einen Workspace aus) und verwenden Sie die Registerkarte **Aktivität**:

- Überprüfen Sie Ereignisse und Jobs, die nur auf diesen Arbeitsbereich beschränkt sind.
- Nutzen Sie diese Ansicht, um Probleme wie einen einzelnen Arbeitsbereich zu isolieren, der im `Processing` Zustand feststeckt.

Clusterweite Ereignisse anzeigen

Prüfen Sie die Ereignismeldungen, um einen wertvollen Überblick über die Systemaktivität zu erhalten. Jedes Ereignis enthält eine Beschreibung und eine eindeutige Kennung sowie eine empfohlene Vorgehensweise.

1. Im ONTAP System Manager wählen Sie **Events & jobs** und dann **Events**.
2. Prüfen Sie die empfohlenen Maßnahmen oben auf der Seite und reagieren Sie darauf, wie z. B. das Aktivieren der automatischen Aktualisierung.
3. Wählen Sie die Registerkarte **Events log**, um eine Liste der Meldungen anzuzeigen.
4. Wählen Sie eine Ereignismeldung aus, um sie genauer zu untersuchen, einschließlich der Sequenznummer, Beschreibung, Ereignis und empfohlenen Maßnahmen.
5. Optional können Sie die Registerkarte **Active IQ suggestions** auswählen und sich bei Active IQ registrieren, um detaillierte Risikoinformationen für den Cluster zu erhalten.

Clusterweite Jobs anzeigen

Alle auf dem AIDE-Cluster laufenden Jobs anzeigen, einschließlich AIDE-spezifischer Jobs und allgemeiner ONTAP-Jobs.

1. Im ONTAP System Manager wählen Sie **Events & Jobs** und dann **Jobs**.
2. Passen Sie die Anzeige individuell an sowie suchen und laden Sie die Jobinformationen nach Bedarf herunter.

Revisionsprotokoll anzeigen

Nutzen Sie das Revisionsprotokoll, um eine Aufzeichnung der Systemaktivitäten basierend auf der Verwendung von Zugriffsprotokollen wie HTTP zu überprüfen.

1. Im ONTAP System Manager wählen Sie **Ereignisse & Jobs** und dann **Revisionsprotokolle**.
2. Wählen Sie **Settings**, um die Vorgänge zu aktivieren oder zu deaktivieren, die verfolgt werden.

Benachrichtigungen verwalten

Konfigurieren Sie Benachrichtigungsziele, um AIDE-Ereignisse und Revisionsprotokolle automatisch weiterzuleiten.

Schritte

1. Im ONTAP System Manager wählen Sie **Cluster** und dann **Settings**.
2. Navigieren Sie zu **Benachrichtigungsverwaltung** und wählen Sie .
3. Wählen Sie die entsprechende Aktion aus, um die von AIDE verwendeten Ziele anzuzeigen oder zu konfigurieren:
 - a. Ereignisziele: Wählen Sie **View event destinations**
 - b. Revisionsprotokoll-Ziele: Wählen Sie **Revisionsziele anzeigen**
4. Wählen Sie **Hinzufügen** aus, wenn zutreffend, und geben Sie die Zielinformationen an.
5. Wählen Sie **Save**.

Verwandte Informationen

- ["ONTAP Ereignis-, Leistungs- und Gesundheitsüberwachung"](#)

AI Data Engine Arbeitsbereiche verwalten

Ein Arbeitsbereich ist eine Gruppe von Datenquellen (Volumes), die die AI Data Engine (AIDE) verwendet, um einen Metadatenkatalog für ein bestimmtes Projekt oder einen bestimmten Anwendungsfall zu erstellen und zu aktualisieren. Als Speicheradministrator können Sie ONTAP System Manager verwenden, um den Zustand des Arbeitsbereichs zu überwachen, die Konfiguration anzupassen, Datenquellen zu steuern, Benutzer zu verwalten und Arbeitsbereiche zu löschen, wenn sie nicht mehr benötigt werden.

Bevor Sie beginnen

- Sie benötigen *storage administrator*-Berechtigungen, um Arbeitsbereiche zu verwalten.

Arbeitsbereichsstatus überprüfen

Überprüfen Sie die Arbeitsbereich-Gesundheit, die Kapazitätsauslastung und den Metadatenstatus, um sicherzustellen, dass die Metadata Engine wie erwartet funktioniert und keine unerwarteten Ressourcen verbraucht.

Schritte

1. Wählen Sie im ONTAP System Manager im Navigationsbereich **Data engine > Workspaces** aus.
2. Überprüfen Sie die Zusammenfassung oben auf der Seite für die Gesamtzahl der Arbeitsbereiche, den allgemeinen Zustand der Arbeitsbereiche und die Kapazitätsauslastung.
3. Für arbeitsbereichsspezifische Informationen wählen Sie einen Arbeitsbereichsnamen aus. Bestätigen Sie auf der Registerkarte **Übersicht**:
 - Arbeitsbereichsstatus und -größe.
 - Datencontainer (Volumes), die im Arbeitsbereich enthalten sind.
 - Anzahl der Elemente und Zeitpunkt der letzten Aktualisierung pro Datenquelle.

- Alle Warnungen auf Arbeitsbereichsebene.



Wenn ein Arbeitsbereich oder eine Datensammlung einen Fehlerstatus anzeigt, vergewissern Sie sich, dass alle Quellvolumes online und zugänglich sind.

4. Wählen Sie die Registerkarte **Data collections** aus, um Folgendes anzuzeigen:

- Alle mit diesem Arbeitsbereich verknüpften Datensammlungen.
- Status (wie `Published` oder `Error`), Größe und Zeitpunkt der letzten Aktualisierung.



System Manager ist für Datensammlungen schreibgeschützt. Dateningenieure können Datensammlungen in der AI Data Engine Console erstellen und verwalten.

5. Wählen Sie die Registerkarte **Users** aus, um Folgendes anzuzeigen:

- Die Liste der Benutzer mit Zugriff auf diesen Arbeitsbereich.

6. Wählen Sie die Registerkarte **Aktivität** aus, um Ereignisse und Aufgaben anzuzeigen, die sich nur auf diesen Arbeitsbereich beziehen.

Arbeitsbereichseigenschaften bearbeiten und Aktualisierungsintervall bearbeiten

Sie können den Namen, die Beschreibung, das Aktualisierungsintervall und (sofern lizenziert) die Guardrail-Richtlinie eines Arbeitsbereichs anpassen.

Schritte

1. Wählen Sie unter **Data engine > Workspaces** neben einem Workspace und wählen Sie **Edit**.

2. Arbeitsbereichseigenschaften bearbeiten:

- Aktualisieren Sie **Name** und **Description** nach Bedarf.
- Passen Sie das **Aktualisierungsintervall** (Metadatenaktualisierungshäufigkeit) innerhalb des zulässigen Bereichs (Stunden und Tage) an.
- Wenn eine AIDE-Lizenz installiert ist, können Sie **Guardrail policy** auswählen.

3. Wählen Sie **Save**.



Änderungen am Aktualisierungsintervall oder an der Metadatenverarbeitung können sich darauf auswirken, wie häufig Remote-SnapMirror Beziehungen für diesen Arbeitsbereich aktualisiert werden.

Fügen Sie einem bestehenden Arbeitsbereich Datencontainer hinzu

Sie können zusätzliche eingebundene Volumes (lokal oder von einem verbundenen Remote-Cluster) hinzufügen, sodass deren Metadaten im Workspace-Katalog enthalten sind.

1. Führen Sie unter **Data engine > Workspaces** einen der folgenden Schritte aus:

- Wählen Sie neben dem Arbeitsbereich und wählen Sie **Datencontainer hinzufügen**.
- Öffnen Sie den Arbeitsbereich, wählen Sie die Registerkarte **Übersicht** und wählen Sie dann im Abschnitt Datencontainer **Hinzufügen** aus.

2. Im Dialogfeld **Datencontainer zum Arbeitsbereich hinzufügen**:

- Lokalisieren Sie lokale Datenträger auf dem AIDE Cluster.

- Erweitern Sie die Peering-Cluster, um Remote-Volumes auszuwählen (Remote-Volumes erfordern Cluster- und SVM-Peering).



Es können nur geeignete, Online-Volumes ausgewählt werden, die nicht global ausgeschlossen sind und noch nicht Teil des Arbeitsbereichs sind.

3. Wenn Sie zur Remote-Volume-Zuordnung aufgefordert werden:
 - Wählen Sie die Ziel-Storage-VM im AIDE Cluster aus, um SnapMirror-Ziele für die ausgewählten Remote-Volumes zu empfangen.
4. Wählen Sie **Add**.
5. Verwenden Sie die Registerkarte **Aktivität** im Arbeitsbereich oder **Data engine > Aktivität**, um die Metadatenextraktion und alle SnapMirror Initialisierungen für neue Datenquellen zu verfolgen.

Datencontainer aus einem Arbeitsbereich entfernen

Sie können einen Datencontainer entfernen, wenn er für den Zweck des Arbeitsbereichs nicht mehr relevant ist oder wenn Sie den Umfang der Metadatenverwaltung für diesen Arbeitsbereich reduzieren möchten. Durch das Entfernen eines Datencontainers wird die Metadatenaktualisierung für dieses Volume gestoppt und seine Metadaten werden aus dem Metadatenkatalog entfernt.



Löschen Sie kein Quellvolume aus ONTAP, das einem Arbeitsbereich hinzugefügt wurde. Wenn Sie das Volume löschen, gerät der Arbeitsbereich in einen Fehlerzustand. Entfernen Sie immer zuerst den Datencontainer aus dem Arbeitsbereich, bevor Sie zugrunde liegende ONTAP Volumes löschen.

Schritte

1. Navigieren Sie zu **Data engine > Workspaces** und wählen Sie den Arbeitsbereich aus, der den Datencontainer enthält.
2. Suchen Sie auf der Registerkarte **Overview** den zu entfernenden Datencontainer.
3. Wählen Sie **Entfernen** neben dem Datencontainer.
4. Überprüfen Sie den Bestätigungsdialog und wählen Sie **Remove**.



Das Entfernen eines Datencontainers aus einem Arbeitsbereich löscht weder das zugrunde liegende ONTAP Volume noch dessen SnapMirror Beziehung. Es betrifft lediglich die Metadatenutzung innerhalb der AI Data Engine.

Workspace-Benutzer verwalten

Sie können Dateningenieuren und Datenwissenschaftlern Zugriff auf einen Arbeitsbereich gewähren oder entziehen. Diese Benutzer sind in Ihrem Identitätsanbieter (OIDC) definiert und ONTAP Rollen zugeordnet. Siehe "[Benutzer zu Arbeitsbereichen zuweisen](#)" Dokumentation, um zu erfahren, wie Sie den Benutzerzugriff verwalten.

Einen Arbeitsbereich löschen

Sie können einen Arbeitsbereich löschen, um die Arbeitsbereichsdefinition und die zugehörigen AIDE Metadaten zu entfernen. Alle mit dem Arbeitsbereich verknüpften Datensammlungen und Vektoreinbettungen werden ebenfalls entfernt.



Zugrunde liegende ONTAP Daten (Volumes, SnapMirror relationships) werden nicht gelöscht.

Schritte

1. Führen Sie unter **Data engine > Workspaces** einen der folgenden Schritte aus:
 - Löschen Sie einen einzelnen Arbeitsbereich, wählen Sie  und wählen Sie **Löschen**.
 - Mehrere Arbeitsbereiche löschen: Aktivieren Sie die Kontrollkästchen für die Arbeitsbereiche und wählen Sie dann **Löschen**.
2. Überprüfen Sie im Bestätigungsdialog die Auswirkungen der Aktion, bevor Sie fortfahren:
 - Die Metadaten des Workspace werden dauerhaft gelöscht.
 - Die mit dem Arbeitsbereich verknüpften Datensammlungen und Embeddings werden dauerhaft gelöscht.



Es gibt keine Option zum Soft-Delete oder Wiederherstellen.

3. Setzen Sie das Häkchen, um Ihr Verständnis zu bestätigen, und wählen Sie **Löschen**.

Verwandte Informationen

- ["Benutzer zu Arbeitsbereichen zuweisen"](#)

Aktualisieren und warten Sie Ihr AIDE-System

AI Data Engine systemaktualisierungen und Kompatibilität

Halten Sie die AI Data Engine (AIDE) Systemkomponenten auf dem neuesten Stand, um optimale Leistung und Zugriff auf neue Funktionen zu gewährleisten. Aktualisieren Sie die Komponenten nach der Bereitstellung, wenn neue Software oder Firmware verfügbar ist, wenn Sie Knoten hinzufügen oder ersetzen, oder regelmäßig für Funktionsupdates.

AIDE Systemkomponenten

Es gibt zwei Hauptkomponenten, die in einem AIDE System aktualisiert werden müssen: ONTAP software und AIDE software, zu der auch DCN firmware gehört.

ONTAP Software

ONTAP ist das Betriebssystem, das auf NetApp Speichersystemen, einschließlich der in AIDE Deployments verwendeten, läuft. Halten Sie ONTAP aktuell, um die Systemstabilität, Sicherheit und Kompatibilität mit AIDE Komponenten zu erhalten. AIDE Komponenten werden separat aktualisiert.

AIDE Software-Updates

AIDE Console software und DCN Firmware-Updates werden zusammen als ein einziges Paket (.tgz) verteilt und sind nicht in ONTAP Images eingebettet. Updates gewährleisten die ordnungsgemäße Funktion der Hardwarekomponenten innerhalb des AIDE Systems und bieten neue Funktionen, Leistungsverbesserungen und Fehlerbehebungen.

Den Aktualisierungsprozess verstehen

AIDE software updates können über ONTAP System Manager verwaltet werden.

AIDE unterstützt die automatische Software-Update-Funktionalität von ONTAP nicht. Sie können sich für Benachrichtigungen von ["NetApp Support Site Downloads"](#) anmelden, aber alle Updates der AIDE software werden manuell vom Administrator durchgeführt.

Veröffentlichungsarten und -umfang:

- ONTAP Hauptversionen (9.x.x) und AIDE Hauptversionen (9.x.x U0) führen neue Funktionen, APIs oder Änderungen ein, die die ONTAP Integration betreffen.
- ONTAP patch releases (9.x.x Px) und AIDE update releases (9.x.x Ux) enthalten Korrekturen und Aktualisierungen, die die ONTAP Integration nicht beeinträchtigen.

Kompatibilitätsmatrix

Stellen Sie die Kompatibilität zwischen ONTAP und AIDE software sicher, wenn Sie Updates planen.

AIDE software wird als "U"-Releases veröffentlicht. Ein AIDE Major Release ist ein "U0"-Release und nachfolgende Minor Releases werden "U1" und später sein.

ONTAP und AIDE Kompatibilität

AIDE Releases	Unterstützte ONTAP Versionen
9.18.1 U0	9.18.1 GA und alle 9.18.1 Px
9.18.1 U1 und später	9.18.1 GA und alle 9.18.1 Px



„Px“ bezeichnet alle ONTAP Patch-Releases in der Hauptversion (zum Beispiel 9.18.1 P1, 9.18.2 und andere).

AIDE Upgrade-Pfade

Die folgenden Beispiele verwenden hypothetische zukünftige Versionen, um die zulässigen Upgrade- und Update-Pfade ab AIDE 9.18.1 U0 und 9.18.1 U1 zu zeigen.

Wenn Ihre aktuelle AIDE Release ist...	Und Ihre angestrebte AIDE release ist...	Ihr Upgrade- oder Updatepfad ist...
9.18.1 U0	9.18.1 U1	Direkt
9.18.1 U0	9.18.1 U3	Direkt (Sie können von jeder Ux auf jede spätere Ux in 9.18.1 aktualisieren)
9.18.1 U1	9.18.1 U3	Direkt (Sie können von jeder Ux auf jede spätere Ux in 9.18.1 aktualisieren)

Einschränkungen rückgängig machen

AIDE-Systeme unterstützen keine Rücksetzvorgänge für DCN-Firmware, AIDE-Software-Updates oder ONTAP auf AFX-Speichersystemen. Nachdem Sie ein Update oder Upgrade installiert haben, können Sie nicht zu einer vorherigen Version zurückkehren. Überprüfen Sie [Kompatibilitätsanforderungen](#) vor dem Update oder Upgrade.

Verwandte Informationen

- ["ONTAP-Software für AFX-Systeme aktualisieren"](#)
- ["AIDE-Software aktualisieren"](#)

Aktualisieren Sie die AI Data Engine software

Als Speicheradministrator können Sie AI Data Engine (AIDE) Software, Data Compute Node (DCN) Firmware und andere Systemdateien auf Ihrem AIDE-System mit ONTAP System Manager aktualisieren.

AIDE unterstützt die automatische Software-Update-Funktionalität von ONTAP nicht. Sie können sich für Benachrichtigungen von ["NetApp Support Site Downloads"](#) anmelden, aber alle Updates der AIDE software werden manuell vom Administrator durchgeführt.

Über diese Aufgabe

Das kombinierte AI Data Engine software package ist deutlich größer als ein typisches ONTAP update package (AIDE package ist ungefähr 40GB). Planen Sie längere Upload- und Installationszeiten ein, wenn Sie die AIDE software aktualisieren.

Bevor Sie beginnen

- Sie benötigen *Storage Administrator*-Berechtigungen, um die DCN-Firmware und die AI Data Engine software zu aktualisieren.
- Sie benötigen Ihre NetApp Support Site-Zugangsdaten für Ihr aktives Konto.
- ["Stellen Sie die Kompatibilität zwischen ONTAP, DCN-Firmware und AI Data Engine software sicher, wenn Sie Updates planen."](#)



Die Rückgängigmachungsfunktion wird für DCN Firmware- oder AI Data Engine software-Updates nicht unterstützt. Nach der Installation eines Updates können Sie nicht zu einer vorherigen Version zurückkehren.

Schritte

1. ["Laden Sie die kombinierte DCN Firmware- und AIDE software-Update-Datei auf einen lokalen Client herunter"](#).
2. Im System Manager wählen Sie **Cluster > Settings > Software updates**.
3. Neben **Software updates** wählen Sie ➔.
4. Unter AI Data Engine-Updates wählen Sie **Add AI Data Engine software files** aus und wählen Sie das Update-Paket aus.
5. Nach Abschluss des Paket-Uploads wählen Sie **Update**, um mit der Installation der Updates auf den DCN-Knoten zu beginnen.



AIDE Console ist nicht verfügbar oder zugänglich, wenn die DCN-Knotenaktualisierung läuft.

Ergebnis

Die DCNs werden mit AI Data Engine software aktualisiert und die aktualisierte Version wird für jeden Knoten angezeigt.

Verwandte Informationen

- ["ONTAP-Software für AFX-Systeme aktualisieren"](#)

Fügen Sie Ihrem AIDE-Cluster Datenrechenknoten hinzu

Sie können Datenrechenknoten (DCNs) im Rahmen der Erstellung eines neuen AI Data Engine (AIDE)-Clusters oder der Erweiterung eines bestehenden Clusters hinzufügen. Der Workflow umfasst das Erkennen und Konfigurieren der Knoten mithilfe von ONTAP System Manager.

Bereiten Sie das Hinzufügen von Knoten vor

Beim Hinzufügen von DCN-Knoten gibt es mehrere Überlegungen.

Beim Erstellen eines neuen AIDE-Clusters

Für einen neuen AIDE Cluster werden genau drei DCN-Knoten benötigt.

Hardwareinstallation und Adressierbarkeit

Stellen Sie sicher, dass die folgenden Voraussetzungen erfüllt sind:

- Die neue DCN-Hardware ist im Rack montiert, eingeschaltet und mit den Cluster-Switches verkabelt.
- Sie haben einen IP-Adressraumbereich für das DCN-zu-ONTAP-Backend-Subnetz verfügbar.
- Der ONTAP Cluster wurde initialisiert und ist vom Cluster Management LIF aus erreichbar.

System Manager-Anmeldeinformationen

Sie benötigen *storage administrator*-Berechtigungen, um Aufgaben zur Erstellung oder Erweiterung eines AIDE-Clusters durchzuführen.

Softwarekompatibilität

Prüfen Sie die folgende Dokumentation, um zu bestätigen, dass Ihre DCN-Hardware- und Softwareversionen mit Ihrem ONTAP Cluster kompatibel sind:

- ["DCN-Software ist mit der ONTAP Version kompatibel"](#).

Während des Hinzufügens von Knoten überprüft der System Manager, ob auf den neuen Knoten eine Softwareversion ausgeführt wird, die kompatibel ist mit:

- Die effektive Version (ECV) des ONTAP-Clusters, falls dies der erste DCN-Beitritt ist.
- Die bestehende DCN-Cluster-Version, wenn DCNs bereits vorhanden sind.

Wenn ein Knoten inkompatibel ist:

- Neben dem betroffenen DCN wird im Dialogfeld **Hinzufügen** ein Fehler angezeigt.
- Sie müssen zuerst die DCN-Software (oder ONTAP, wie zutreffend) auf eine kompatible Version aktualisieren.

Fügen Sie die Datenberechnungsknoten hinzu

Sie fügen DCN-Knoten im Rahmen der Erstellung eines neuen AIDE-Clusters oder der Erweiterung eines bestehenden Clusters hinzu.

Schritte

1. Wählen Sie im System Manager den **Dashboard** im Navigationsbereich und dann die Karte **Health** aus.
2. Bestätigen Sie, dass Knoten zum Hinzufügen vorhanden sind, und wählen Sie **Details anzeigen**, um die Liste anzuzeigen.

Die Liste enthält gefundene Knoten, die noch nicht Teil des AIDE Clusters sind

3. Alternativ können Sie **Cluster** und **Overview** sowie die Registerkarte **Data compute** auswählen, um die Liste anzuzeigen.
4. Wählen Sie am unteren Rand der Seite „Datenberechnung“ oberhalb der Liste der Knoten **Hinzufügen** aus.
5. Im Dialogfeld **Datenrechenknoten hinzufügen** wählen Sie die DCN-Knoten aus, die Sie hinzufügen möchten.

Sie können einzelne Knoten optional vor dem Hinzufügen umbenennen.

6. Wenn Sie zum ersten Mal Knoten hinzufügen und kein Backend-Subnetz existiert, wählen Sie **Subnetz hinzufügen** und geben Sie Folgendes an:

- Subnetzname (für interne Verwendung)
- Subnetzadresse und Maske
- IP-Adressbereich für DCN und ONTAP-Knoten auf diesem Backend-Netzwerk

System Manager überprüft, ob der Bereich genügend freie IP-Adressen für alle hinzuzufügenden DCNs, alle ONTAP-Knoten im Cluster und zusätzliche Floating IPs auf Clusterebene enthält, die für die Kommunikation zwischen DCN und ONTAP verwendet werden.

7. Ob Sie das Backend-Subnetz hinzugefügt haben oder es bereits existiert:

a. Überprüfen Sie die verfügbaren IP-Adressen.

b. Wählen Sie bei Bedarf **Subnetz bearbeiten** aus und erweitern Sie den IP-Bereich.

- Der Bereich kann nur erweitert werden. Das Verkleinern oder Ändern des Subnetzes wird nicht unterstützt.
- Eine Änderung des Subnetzes oder des IP-Bereichs kann die Neuerstellung des zugrunde liegenden Kubernetes-Clusters auf DCNs erforderlich machen und mehrere Minuten dauern.

8. Optional können Sie die Data Engine-Service-Schnittstelle konfigurieren, indem Sie Folgendes angeben:

- Dienst-IP-Adresse
- Netzwerkmaske
- Gateway (falls für Ihre Umgebung erforderlich)

Die einzelne IP wird über die DCNs verteilt und als Frontend-Adresse für die AI Data Engine Console und zugehörige APIs verwendet.

9. Überprüfen Sie die ausgewählten Knoten, das Backend-Subnetz und die Einstellungen der Data Engine-Service-Schnittstelle.
10. Wählen Sie **Hinzufügen** und warten Sie, bis der Vorgang abgeschlossen ist. System Manager führt die folgenden Aktionen aus:
 - Fügen Sie die ausgewählten Knoten dem DCN Cluster hinzu
 - Stellen Sie das Backend-Netzwerk bereit und fügen Sie Knoten dem Kubernetes-basierten DCN Cluster hinzu

- Aktualisiert interne Metadaten für DCN discovery

11. Nach Abschluss wählen Sie **Cluster** und **Overview** aus und bestätigen Sie:

- Unter **Datenverarbeitung** erscheinen die neuen DCNs als Teil des Clusters
- Alle Knoten sind **Healthy**
- Überprüfen Sie, ob die **Health**-Karte im Dashboard die aktualisierte Knotenanzahl anzeigt.

Ersetzen Sie einen Knoten in Ihrem AIDE Cluster

Sie müssen einen Datenrechenknoten (DCN) in Ihrem AI Data Engine (AIDE) Cluster ersetzen, wenn er nicht mehr funktioniert oder aufgrund eines Hardwareausfalls, Upgrades oder Wartungsarbeiten ausgetauscht werden muss. Dies stellt sicher, dass der AIDE Cluster gesund und funktionsfähig bleibt. Das Verfahren kann durchgeführt werden, ohne laufende Dienste zu unterbrechen.

Bereiten Sie den Austausch eines Knotens vor

Bevor Sie einen Knoten in Ihrem AIDE Cluster austauschen, gibt es mehrere Dinge zu beachten.

System Manager-Anmeldeinformationen

Sie benötigen *storage administrator*-Berechtigungen, um AIDE-Clusterknotenaustauschaufgaben durchzuführen.

Einschränkungen

Sie sollten sich der folgenden Einschränkungen bewusst sein, wenn Sie einen Knoten in Ihrem AIDE Cluster ersetzen:

- Der Austausch von Knoten wird nur über die CLI und optional die REST API unterstützt.
- Sie können den Knotenaustausch nicht mit dem System Manager durchführen.
- Der neue Knoten sollte mit der Softwareversion des Clusters übereinstimmen; ONTAP wird ihn bei Bedarf aktualisieren.
- Der ausgefallene Knoten darf nicht eingeschaltet werden, während er mit dem Cluster-Netzwerk verbunden ist, um IP-Adresskonflikte zu vermeiden.

Anforderungen

Sie benötigen Folgendes:

- Seriennummer für den neuen Ersatzknoten

Ersetzen Sie einen DCN-Knoten in Ihrem AIDE-Cluster

Sie können einen DCN-Knoten in Ihrem AIDE-Cluster mit folgendem Verfahren ersetzen.

Schritte

1. Entfernen Sie den defekten Knoten physisch

Schalten Sie den Knoten aus und trennen Sie ihn vom Clusternetzwerk. Stellen Sie sicher, dass der Knoten während des Austauschvorgangs nicht im Netzwerk hochgefahren ist.

2. Löschen Sie den ausgefallenen Knoten aus dem Cluster mit dem folgenden Befehl:

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

Geben Sie den tatsächlichen Namen für den <node_name> Wert an.

3. Schließen Sie den neuen Knoten physisch an den Cluster an

Stellen Sie sicher, dass der Knoten verkabelt, eingeschaltet und auffindbar ist.

4. Zeigen Sie die auffindbaren und nicht konfigurierten Knoten an, um zu überprüfen, dass der neue Knoten online ist:

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

5. Fügen Sie den Knoten mit dem folgenden Befehl zum Cluster hinzu:

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

ONTAP weist dem neuen Knoten eine IP-Adresse zu. Wenn die Softwareversion des Knotens nicht mit der des Clusters übereinstimmt, aktualisiert ONTAP den Knoten automatisch.

6. Überprüfen Sie den Clusterzustand und die Knotenintegration mit einem der folgenden Befehle:

```
dcn cluster node show
```

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

Verwandte Informationen

- ["Erweitern Sie Ihren Compute-Cluster"](#)

Vektorisierung und Datensammlungen verwalten

Data-to-RAG-Schnellstart für AI Data Engine

Gehen Sie mit diesem Workflow von einem neu bereitgestellten AI Data Engine (AIDE) System zu einem funktionsfähigen Retrieval-Augmented Generation (RAG) Endpunkt. Verstehen Sie, wie Speicheradministratoren, Dateningenieure und Data Scientists mithilfe von ONTAP System Manager und AIDE Console zusammenarbeiten.

Bevor Sie beginnen

- Sie haben Data compute nodes (DCNs) zum ONTAP Cluster installiert und hinzugefügt.
- Sie haben die AI Data Engine software für Vektorisierung und Leitplanken installiert und lizenziert.
- Sie haben ["OpenID Connect \(OIDC\)"](#) konfiguriert und Rollen für Administratoren, Dateningenieure und Datenwissenschaftler zugeordnet.

1

Datenumfang und Governance definieren

Als Speicheradministrator oder Sicherheitsadministrator möchten Sie die Umgebung in AIDE Console und ONTAP System Manager vorbereiten:

- ["Erstellen Sie einen oder mehrere Arbeitsbereiche"](#) from lokalen und entfernten Datenquellen.
- ["Klassifikatoren und Guardrail-Richtlinien konfigurieren"](#) in der AIDE Console.
- ["Weisen Sie Dateningenieuren und Datenwissenschaftlern Zugriff auf die Workspaces zu"](#).

2

Arbeitsbereichsmetadaten erkunden

Als Data Engineer oder Data Scientist möchten Sie die Metadaten des Arbeitsbereichs mithilfe der AIDE Console untersuchen:

- ["Arbeitsbereichsmetadaten erkunden"](#) um die verfügbaren Inhalte zu verstehen.
- Definieren Sie eine oder mehrere logische Teilmengen von Daten, die RAG speisen sollen (zum Beispiel Supportartikel, Produkthandbücher oder anonymisierte klinische Notizen).

3

Erstellen und Veröffentlichen einer Datensammlung

Als Data Engineer oder Data Scientist möchten Sie die ausgewählte Teilmenge in eine RAG-ready Sammlung umwandeln:

- ["Erstellen einer Datensammlung"](#) aus dem Arbeitsbereich unter Verwendung ausgewählter Filter.
- ["Veröffentlichen Sie die Datensammlung"](#) und überwachen Sie die Indizierung, bis sie den Ready Zustand erreicht.
- Kopieren Sie die Retrieval-Endpunkt-URI für die ausgewählte Collection und stellen Sie sie den Data Scientists oder Anwendungsentwicklern zur Verfügung.
- ["Status der Datenerfassung und Vektor-Fußabdruck anzeigen"](#) nach Bedarf.

Was kommt als Nächstes?

- ["Definieren Sie Ihren Datenbestand und Ihre Schutzrichtlinien in AI Data Engine"](#)
- ["Erkunden Sie Arbeitsbereichs-Metadaten in der AI Data Engine Console"](#)
- ["Erstellen Sie Datensammlungen in AI Data Engine Console"](#)

Erkunden Sie Arbeitsbereichs-Metadaten in der AI Data Engine Console

Als Data Engineer oder Data Scientist besteht Ihre erste Aufgabe in AI Data Engine (AIDE) darin, zu verstehen, welche Daten in Ihren Arbeitsbereichen verfügbar sind. Sie verwenden die AI Data Engine Console, um den Metadatenkatalog abzufragen, nach relevanten Dateien zu suchen und Datenteilmengen zu identifizieren, die Sie in Datensammlungen umwandeln.

Bevor Sie beginnen

- Sie benötigen *data engineer* oder *data scientist* Berechtigungen in AI Data Engine Console und Zugriff auf mindestens einen Workspace.
- Ein Speicheradministrator hat:
 - Ein oder mehrere Arbeitsbereiche im ONTAP System Manager wurden erstellt.
 - Ihrem Benutzer oder Ihrer Gruppe wurde Zugriff auf die relevanten Arbeitsbereiche zugewiesen.
- Die Metadatenextraktion für den Arbeitsbereich ist abgeschlossen und der Arbeitsbereich befindet sich im Ready Zustand.
- Klassifikatoren sind aktiviert, sodass Metadaten Klassifizierungs-Tags (zum Beispiel PII-Indikatoren) enthalten.

Sign in to AIDE Console als Dateningenieur oder Datenwissenschaftler

Schritte

1. Öffnen Sie im Browser die AIDE Console URL:

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

2. Authentifizieren Sie sich über den OIDC-Anbieter Ihrer Organisation.
3. Prüfen Sie, ob Ihre Rolle als Data Engineer oder Data Scientist erkannt wird (z. B. anhand des verfügbaren Arbeitsbereichs und der Datenerfassungsaktionen). Weitere Informationen finden Sie unter ["AIDE Rollendokumentation"](#), um zu erfahren, wie Data Engineers und Data Scientists mit AIDE Komponenten arbeiten.

Ergebnis

Sie sind in der AI Data Engine Console angemeldet und sehen nur die Arbeitsbereiche, für die Ihnen Zugriff gewährt wurde.

Ihre zugänglichen Arbeitsbereiche anzeigen

Schritte

1. Navigieren Sie in der AIDE Console zu **Data Curator > Workspaces**.
2. Überprüfen Sie die Liste der Arbeitsbereiche, auf die Sie zugreifen können.

3. Wählen Sie einen Arbeitsbereich aus, um dessen Details anzuzeigen.

Ergebnis

Sie haben nun eine auf den Arbeitsbereich beschränkte Ansicht des Datenbestands, den die Speicheradministratoren für Ihre Projekte zur Verfügung gestellt haben.

Was kommt als Nächstes?

- ["Erstellen Sie Datensammlungen für RAG aus einem Workspace"](#)

Erstellen Sie Datensammlungen in AI Data Engine Console

Datensammlungen sind die zentralen RAG-Bausteine der AI Data Engine (AIDE). Als Dateningenieur oder Data Scientist definieren Sie, welche Dateien zu einer Sammlung gehören, konfigurieren Einbettungs- und Indizierungsoptionen und veröffentlichen die Sammlung, sodass Anwendungen sie über einen Retrieval-Endpunkt abfragen können.

Sie führen alle Datenerfassungsaufgaben in der AI Data Engine Console aus.

Bevor Sie beginnen

- Sie benötigen Berechtigungen als *data engineer* oder *data scientist* in der AI Data Engine Console (https://<cluster_management_ip>/console).
- Sie haben Zugriff auf mindestens einen Arbeitsbereich, in dem Metadaten extrahiert und im Ready Status vorliegen.
- Sie haben die Metadaten des Arbeitsbereichs untersucht und Abfragen oder Filter identifiziert, die sinnvolle Datenteilmengen definieren.
- Die AI Data Engine software-Lizenz ist installiert und die Inferenzfunktionen sind aktiviert.

Erstellen einer Datensammlung aus Workspace-Metadaten

Schritte

1. Navigieren Sie zu **Data Curator > Workspaces** und wählen Sie den Workspace aus, der Ihre Zieldaten enthält.
2. Wählen Sie **Add data collection**.
3. Gehen Sie auf der Seite **Create new data collection** wie folgt vor:
 - a. Geben Sie einen Namen und eine Beschreibung für die Sammlung ein (zum Beispiel, `Support_KB_RAG_EN`).
 - b. Wählen Sie aus, ob die Sammlung wie folgt sein soll:
 - **Dynamisch:** Neue Dateien werden automatisch erkannt und der Datensammlung anhand der von Ihnen definierten Filterkriterien hinzugefügt. Dies geschieht während Aktualisierungen des Arbeitsbereichs.
 - **Statisch:** Sie wählen aus, welche Dateien in die Sammlung aufgenommen werden. Sie können die Dateien bearbeiten, wenn sich die Datensammlung im `draft` Status befindet. Nachdem die Datensammlung in den `Published` Status übergegangen ist, kann sie nicht mehr bearbeitet werden.
4. Geben Sie die Quellteilmenge an:

- a. Verwenden Sie Schlüsselwörter und Filter (Dateityp, Zeitstempel und andere Attribute), um die relevanten Dateien zu finden, die einbezogen werden sollen.



Sie können einen Dateinamen auswählen, um ein Vorschauenfenster des Inhalts zu öffnen.

5. Fügen Sie diese Dateien der Datensammlung hinzu.
6. Wählen Sie **Speichern**, um die Sammlung abzuschließen.

Ergebnis

Sie haben den Umfang der Datenerfassung definiert und die erforderlichen Dateien hinzugefügt. AIDE generiert Einbettungen und erstellt den Vektorindex, wenn Sie die Sammlung veröffentlichen.



Erstellen Sie kleine, fokussierte Sammlungen (z. B. pro Anwendungsfall oder Domäne) anstatt einer einzigen „Alles“-Sammlung. Dies verbessert die Relevanz der Suchergebnisse und die Verwaltbarkeit.

Veröffentlichen Sie eine Datensammlung

Veröffentlichen Sie die Datensammlung, um sie über einen RAG-Abfrageendpunkt für KI-Anwendungen abfragbar zu machen. Die Veröffentlichung generiert Vektoreinbettungen aus Ihren ausgewählten Dateien und indiziert sie für die semantische Suche. Nachdem die Sammlung den Ready Status erreicht hat, steht ihr Endpunkt Datenwissenschaftlern zur Integration in Notebooks, Pipelines und KI-Anwendungen für Retrieval-Augmented Generation (RAG) und Suche zur Verfügung.



Für große Sammlungen sollten Sie die anfängliche Veröffentlichung und größere Neuveröffentlichungen während Nebenzeiten einplanen, um Ressourcenkonflikte zu minimieren.

Schritte

1. Navigieren Sie zu **Data Curator > Data collections** und wählen Sie das Optionsmenü (⋮) für Ihre Datensammlung aus.
2. Wählen Sie **Publish**.
3. Wählen Sie eine Standard- oder benutzerdefinierte Optimierungskonfiguration aus.
4. Wählen Sie **Publish**, um die Datentransformation zu starten.
5. Öffnen Sie in der AIDE Console die Sammlungsdetailansicht (**Data Curator > Data collections**) für Statusaktualisierungen.

Ergebnis

Die Sammlung erreicht den Ready Status und steht nachgelagerten Anwendungen und Datenwissenschaftlern zur Verfügung.

Unter **Data Curator > Data collections** können Sie **Copy URI** auswählen, um die Informationen zu erhalten, die für den Zugriff auf die Datensammlung über eine API erforderlich sind.

Aktualisieren oder Löschen einer Datensammlung

Im Laufe der Zeit kann es erforderlich sein, Datensammlungen zu verfeinern oder außer Betrieb zu nehmen. Das Verfeinern einer Sammlung kann das Anpassen von Filtern zum Hinzufügen oder Entfernen von Dateien, das Ändern von Einbettungseinstellungen oder das Aktualisieren der Sammlungsbeschreibung umfassen. Das

Löschen einer Sammlung entfernt sie dauerhaft und macht ihren Abruf-Endpunkt nicht mehr verfügbar.

Eine Datensammlung aktualisieren

Sie können eine Datensammlung aktualisieren, wenn sie sich im `draft`-Status befindet.

Schritte

1. Navigieren Sie zu **Data Curator > Data collections**.
2. Wählen Sie die Sammlung aus, die Sie ändern möchten.
3. Wählen Sie **Edit**.
4. Passen Sie eines der folgenden Elemente an:
 - Name und Beschreibung
 - Filter (Pfade, Dateitypen, Klassifizierungs-Tags).
 - Einbettungs- und Chunking-Einstellungen.
5. Speichern Sie Ihre Änderungen.
6. Veröffentlichen Sie die Sammlung erneut, damit die neue Definition und die Einbettungen wirksam werden.

Ergebnis

Ein neuer Indizierungsauftrag wird mit der aktualisierten Konfiguration ausgeführt, und die Sammlung kehrt in einen `Ready` Zustand zurück, wenn sie abgeschlossen ist.

Eine Sammlung löschen

Das Löschen einer Sammlung ist endgültig. Stellen Sie sicher, dass keine Produktionsanwendung mehr vom Abrufendpunkt der Sammlung abhängt, bevor Sie sie löschen.

Schritte

1. Navigieren Sie zu **Data Curator > Data collections** und wählen Sie das Optionsmenü (`...`) für die Sammlung aus.
2. Wählen Sie **Delete**.
3. Bestätigen Sie die Löschung.

Ergebnis

Die Sammlungsdefinition und ihre Einbettungen werden aus AI Data Engine entfernt. Anwendungen, die versuchen, den ehemaligen Abrufendpunkt abzufragen, schlagen nach dem Entfernen der Sammlung fehl.

Was kommt als Nächstes?

- ["Datensammlungen anzeigen"](#)

Datensammlungen in AI Data Engine anzeigen

Nachdem Dateningenieure oder Datenwissenschaftler Datensammlungen aus Arbeitsbereichen erstellt und veröffentlicht haben, benötigen Sie Einblick in deren Status, Größe und Auswirkungen auf den AI Data Engine cluster.

Wenn Sie Speicheradministrator, Dateningenieur oder Datenwissenschaftler sind, können Sie

Datensammlungen in ONTAP System Manager und AIDE Console anzeigen.

Bevor Sie beginnen

- Sie benötigen entweder *storage administrator* Berechtigungen im ONTAP System Manager oder *data engineer* oder *data scientist* Berechtigungen in der AI Data Engine Console (https://<cluster_management_ip>/console), um Datensammlungen anzuzeigen.
- Mindestens ein Arbeitsbereich existiert mit erfolgreich extrahierten Metadaten.
- Dateningenieure oder Datenwissenschaftler haben mindestens eine Datensammlung aus der AI Data Engine Console erstellt und veröffentlicht.
- Die AI Data Engine software-Lizenz ist installiert und die Inferenzfunktionen sind aktiviert, sodass Vektorisierungs- und Abrufendpunkte aktiv sind.

Clusterweite Datensammlungen anzeigen

Für Speicheradministratoren bietet ONTAP System Manager eine clusterweite Übersicht über Datensammlungen und deren Speicherbedarf, erlaubt es Administratoren jedoch nicht, diese zu erstellen oder zu ändern.

Schritte

1. Navigieren Sie im System Manager zu **Data Engine > Data collections**.
2. Überprüfen Sie die Bestandsübersicht oben auf der Seite:
 - Gesamtzahl der Datenerhebungen nach Status
 - Gesamtspeicherplatz, der von der Vektor-Datenbank über alle Sammlungen hinweg belegt wird
 - Vektorraum als Prozentsatz der Gesamtkapazität des Clusters
3. Wählen Sie eine einzelne Datenerfassung und Überprüfung aus:
 - Sammlungsname und Beschreibung
 - UUID
 - Zugehöriger Arbeitsbereich
 - Status
 - Sammlungsgröße
 - Schöpfer
 - Letzte Aktualisierungszeit

Ergebnis

Sie haben nun einen Überblick über alle Datensammlungen im Cluster und deren Speicherbedarf. Nutzen Sie diese Ansicht, um Sammlungen zu identifizieren, die groß, veraltet oder in einem nicht bereiten Zustand sind.

Sie können außerdem sehen, ob eine einzelne Datensammlung aktiv aktualisiert wird und ob Ausfälle die Nutzung von RAG blockieren.

Überwachen Sie sammlungsbezogene Jobs und Ereignisse

Als Speicheradministrator können Sie Jobs, die Sammlungen erstellen und aktualisieren, von der clusterweiten **Aktivitäts**-Seite und von den Arbeitsbereichsdetails aus überwachen.

Schritte

1. Navigieren Sie im System Manager zu **Data Engine > Activity**.
2. Auf der Registerkarte **Events**:
 - a. Filtern nach Typ (zum Beispiel Arbeitsbereich, Datenerfassung) oder Schweregrad.
 - b. Erweitern Sie jedes Ereignis im Zusammenhang mit Datensammlungen (zum Beispiel „Data collection publish failed“), um weitere Details anzuzeigen.
3. Auf der Registerkarte **Jobs**:
 - a. Filter, um sich auf Datenerfassungs-, Indizierungs- und Veröffentlichungsaufgaben zu konzentrieren.
 - b. Öffnen Sie für jeden Job die Peek-Ansicht, um Folgendes zu sehen:
 - Prozentsatz des Fortschritts.
 - Start- und Endzeiten.
 - Alle gemeldeten Fehlermeldungen oder Warnungen.
4. Optional können Sie zum betroffenen Arbeitsbereich zurückkehren (**Data Engine > Workspaces**) und dessen Registerkarte **Aktivität** öffnen, um Ereignisse und Aufträge anzuzeigen, die nur für diesen Arbeitsbereich gelten.

Ergebnis

Sie können den Lebenszyklus von Datensammlungen verfolgen, ins Stocken geratene oder fehlgeschlagene Jobs identifizieren und Kontextinformationen sammeln, um sie an Dateningenieure, Datenwissenschaftler oder den Support weiterzugeben.



Wenn eine Datenerfassung sich über einen längeren Zeitraum im `Publishing` Status befindet, prüfen Sie auf der Seite „Aktivität“, ob ein entsprechender langlaufender Auftrag vorliegt, bevor Sie von einem Fehler ausgehen.

Datensammlungen aus AIDE Console anzeigen

Dateningenieure und Datenwissenschaftler überwachen Datensammlungen typischerweise direkt von der AIDE Console, wo sie erstellt und veröffentlicht werden.

Schritte

1. Melden Sie sich bei der AI Data Engine Console als Data Engineer oder Data Scientist an.
2. Navigieren Sie zu **Data Collections** und wählen Sie die gewünschte Datensammlung aus.
3. Für jede Kollektion:
 - a. Überprüfen Sie den Status (`Draft`, `Publishing`, `Ready`, oder `Failed`).
 - b. Wählen Sie den Namen der Datensammlung aus, um die Definitionsdetails (Filter, eingeschlossene Dateitypen, Klassifikatoroptionen, Einbettungseinstellungen) anzuzeigen.
 - c. Prüfen Sie die Zeitstempel für die letzte Veröffentlichung oder Aktualisierung.
4. Öffnen Sie bei Bedarf die Jobdetails oder Protokolle (sofern verfügbar), um Fehler oder unvollständige Ausführungen zu verstehen.

Ergebnis

Dateningenieure und Datenwissenschaftler können Sammlungsdefinitionen iterativ bearbeiten und erneut veröffentlichen, während sie Status und Zustand überwachen, ohne Speicheradministratoren einbeziehen zu müssen.

Was kommt als Nächstes?

- ["Erstellen Sie Datensammlungen für RAG in AIDE Console"](#)

Leitplanken implementieren

Definieren Sie Ihre Data Guardrails-Richtlinien in AI Data Engine für Ihren Datenbestand

Als Daten- oder Plattformbesitzer verwenden Sie AI Data Engine (AIDE) Console, um festzulegen, welche Daten für KI in Frage kommen, welche Daten immer tabu sind und welche Sicherheitsregeln gelten, wenn diese Daten für Klassifizierung und Retrieval-Augmented Generation (RAG) verwendet werden.

Verwenden Sie diese Verfahren, um diese Richtlinien in AIDE Console zu definieren, damit ONTAP System Manager sie auf alle Daten in Arbeitsbereichen anwenden kann.

Bevor Sie beginnen

- Sie benötigen *Storage Administrator*-Berechtigungen in AI Data Engine Console (https://<cluster_management_ip>/console, um globale Richtlinien zu erstellen und zu verwalten.
- Sie verfügen über einen AIDE Cluster mit bereitgestellten und fehlerfreien Data Compute Nodes.
- "[OpenID Connect \(OIDC\)](#)" ist konfiguriert, und Ihre IdP-Rolle ist einer AIDE admin Rolle zugeordnet, die die Verwaltung von Datenrichtlinien ermöglicht.
- Die AI Data Engine software-Lizenz ist installiert, sodass Data Guardrails und Inferenzfunktionen aktiviert sind.
- Es existiert mindestens ein Arbeitsbereich, oder Sie haben sich mit dem Administrator abgestimmt, um zu verstehen, welche Datenquellen (Volumes) in Arbeitsbereichen verwendet werden.

Richtlinientypen verstehen

AIDE Console stellt diese Richtlinientypen dar, die Ihre Datenlandschaft prägen:

- **Klassifikatoren:** Aktivieren Sie Klassifikatoren, um PII, Sicherheitsprobleme oder andere Muster in allen Arbeitsbereichen zu erkennen.
- **Classifier-Kategorien:** Gruppieren Sie Classifier in Compliance-Kategorien für Organisation und Management.
- **Guardrail policies:** Sicherheits- und Schwärzungsregeln, die zum Zeitpunkt des Abrufs oder der Schlussfolgerung angewendet werden.

Sie können ONTAP System Manager nicht verwenden, um diese Data Guardrails zu erstellen oder zu verwalten. Er liest sie nur aus und erzwingt sie, wenn ein Speicheradministrator sie auf Arbeitsbereiche anwendet. Die gesamte Richtliniendefinition und -wartung erfolgt in AIDE Console.

Klassifikatoren aktivieren

Klassifikatoren analysieren sowohl Metadaten als auch Inhalte, um Dateien und Objekte zu annotieren (zum Beispiel, um PII oder sensible Kategorien zu erkennen). Bevor Klassifikatoren auf Workspace-Daten ausgeführt werden können, müssen Sie sie in AIDE Console aktivieren.

Über diese Aufgabe

Das Verhalten von Klassifikatoren wird global in der AI Data Engine Console gesteuert. Alle aktivierten

Klassifikatoren werden in jedem Workspace ausgeführt. Da sie global angewendet werden, können sie nicht für einen einzelnen Workspace aktiviert oder deaktiviert werden. Sie können nur global aktiviert oder deaktiviert werden.

Schritte

1. Navigieren Sie in der AIDE Console zu **Data Guardrails > Classifiers**.
2. Wählen Sie eine Klassifikatorkategorie aus, um die darin enthaltenen Klassifikatoren anzuzeigen.
3. Wählen Sie die Kontrollkästchen für die Klassifikatoren aus, die Sie aktivieren möchten, oder wählen Sie alle Zeilen aus, um Klassifikatoren in großen Mengen zu aktivieren.
4. Wählen Sie **Enable**.



Verwenden Sie die Option „Massenauswahl“, um mehrere Klassifikatoren gleichzeitig zu aktivieren. Jedes Mal, wenn Sie einen Klassifikator aktivieren, wird eine Aktualisierung in allen Arbeitsbereichen ausgelöst. Um unnötige Aktualisierungen zu vermeiden, aktivieren Sie mehrere Klassifikatoren gleichzeitig, anstatt sie einzeln zu aktivieren.

Ergebnis

Alle neu erstellten und bestehenden Arbeitsbereiche führen die aktivierten Klassifikatoren während der Metadatenverarbeitung aus.

Klassifizierungs-Tags werden in den Metadatenkatalog geschrieben und stehen Dateningenieuren beim Erstellen von Datensammlungen zum Filtern zur Verfügung.

Klassifiziererkategorien verwalten

Klassifikatoren werden in Kategorien (wie „PII“ oder „Finanzdaten“) organisiert. Kategorien helfen Ihnen, zusammengehörige Klassifikatoren für eine einfachere Verwaltung und Compliance-Transparenz zu gruppieren. Sie können die von AIDE bereitgestellten Standardkategorien verwenden oder benutzerdefinierte Kategorien erstellen, um Ihre Compliance-Anforderungen zu erfüllen.

Schritte

1. Navigieren Sie in der AIDE Console zu **Data Guardrails > Classifiers**.
2. Sehen Sie sich die vorhandenen Klassifizierungskategorien an. Es gibt zwei Hauptkategorien der Klassifizierung:
 - Inhalt oder Daten: Erkennt bestimmte Arten von Daten in Dateien.
 - Dokument: Klassifiziert den Dokumenttyp basierend auf dem Inhalt.
3. Prüfen Sie, ob die standardmäßigen Unterkategorien des Klassifikators ausreichen oder ob Sie eine eigene Unterkategorie erstellen möchten.
 - Wenn Sie eine Standardklassifizierungsunterkategorie verwenden (zum Beispiel **General Privacy**):
 - i. Wählen Sie den Kategorienamen in Klassifikatorkategorien aus, um die zugehörigen Klassifikatoren anzuzeigen.
 - ii. Untersuchen Sie die Liste der Klassifikatoren.
 - iii. Wählen Sie **Add**, um nicht aufgeführte Klassifikatoren aus der vollständigen Liste der verfügbaren Klassifikatoren zu finden und hinzuzufügen.
 - Wenn Sie eine benutzerdefinierte Kategorie erstellen möchten, wählen Sie **+ Add**.
 - i. Fügen Sie einen eindeutigen Namen, eine Beschreibung hinzu und ordnen Sie der Kategorie verfügbare Klassifikatoren zu.

ii. **Hinzufügen** auswählen

- Um einen Klassifikator innerhalb einer Kategorie zu deaktivieren, wählen Sie **...** für den Klassifikator und wählen Sie **Deaktivieren**. Sie können auch alle Zeilen auswählen, um Statusänderungen in einem Schritt vorzunehmen.

Ergebnis

Kategorien strukturieren Klassifikatoren für Compliance-Transparenz. Dateningenieure können Klassifizierungs-Tags beim Filtern und Erstellen von Datensammlungen verwenden.

Guardrail-Richtlinien erstellen und verwalten

Guardrail-Richtlinien legen fest, wie AIDE reagiert, wenn Klassifikatoren sensible Inhalte erkennen oder wenn Eingabeaufforderungen und Suchergebnisse gegen Inhaltsregeln verstoßen.

Typische Data Guardrails-Verhaltensweisen sind:

- Maskierung oder Schwärzung von PII aus abgerufenen Ausschnitten.
- Blockieren von Antworten, die gegen die Compliance-Regeln verstoßen.
- Protokollierung oder Kennzeichnung von Verstößen für Audit-Zwecke.

Über diese Aufgabe

Sie erstellen und verwalten Data Guardrails-Richtlinien ausschließlich in der AI Data Engine Console.

Sie können Arbeitsbereiche im ONTAP System Manager jeweils nur mit einer einzigen Guardrail-Richtlinie verknüpfen.

Schritte

- Navigieren Sie in der AIDE Console zu **Data Guardrails > Guardrail policies**.
- Wählen Sie **Add**.
- Geben Sie einen Namen und eine Beschreibung ein, die den Geltungsbereich klar beschreiben (zum Beispiel `Customer PII redaction for support KB`).
- Konfigurieren Sie die datenklassifikatorgesteuerten Bedingungen, die für die Aktivierung der Data Guardrails erforderlich sind:
 - Definieren Sie die Bedingungen für die Aktivierung der Data Guardrails:
 - Wählen Sie die Klassifizierungskategorie oder den Klassifizierungstyp für jede Bedingung aus.
 - Fügen Sie nach Bedarf zusätzliche Bedingungen hinzu und definieren Sie diese.
 - Definieren Sie spezifische Suchkriterien in **Suchen** und wählen Sie dann **Akzeptieren**.
 - Definieren Sie Aktionen für die Guardrail-Richtlinie, wie das Anonymisieren von Inhalten oder das Blockieren und Entfernen einer Datei aus einer Datensammlung.
- Wählen Sie den Arbeitsbereich aus, auf den der Guardrail angewendet wird.
- Legen Sie den Richtlinienstatus fest:
 - **Enabled**: Aktiviert die Richtlinie sofort.
 - **Testmodus**: Ermöglicht es Ihnen, die Auswirkungen der Richtlinie zu validieren, bevor Sie sie aktivieren.
 - **Deaktiviert**: Speichert die Guardrail, ohne sie zu erzwingen.

7. Wählen Sie **Hinzufügen**, um die Richtlinie zu speichern und auf den Arbeitsbereich anzuwenden.



Verwenden Sie den **Testmodus** mit einem Pilotarbeitsbereich und einer nicht-produktiven Datenerfassung, um zu verstehen, wie viele Antworten betroffen wären, bevor Sie die strikte Durchsetzung aktivieren.

Ergebnis

Die neue Leitplankenrichtlinie ist aktiv und auf den ausgewählten Arbeitsbereich beschränkt.

Wie Richtlinien mit Workspaces interagieren

Nachdem die Richtlinien festgelegt wurden:

- Der Speicheradministrator verwendet ONTAP System Manager, um Arbeitsbereiche zu erstellen, Datencontainer auszuwählen und eine Guardrail-Policy zuzuordnen.
- Klassifikatoren werden automatisch auf Workspace-Inhalten ausgeführt, basierend auf dem, was Sie aktiviert haben.
- Die an den Arbeitsbereich angehängten Data Guardrails beeinflussen das Verhalten der Retrieval-Endpunkte.

Für Dateningenieure und Datenwissenschaftler:

- Der sichtbare Datenbestand (Arbeitsbereiche und Datensammlungen) ist bereits nach Rollenzuweisung gefiltert.
- Die von Ihnen abgefragten Metadaten (z. B. PII-Tags) werden durch die aktivierten Klassifikatoren gesteuert.
- Die Antworten, die Ihre RAG-Pipelines erhalten, werden durch die auf Workspace-Ebene konfigurierten Data Guardrails eingeschränkt.

Verwandte Informationen

- ["Datensammlungen in AI Data Engine anzeigen"](#)
- ["Data-to-RAG-Schnellstart für AI Data Engine"](#)

FAQ für NetApp AI Data Engine

Diese FAQ beantwortet häufig gestellte Fragen zur NetApp AI Data Engine (AIDE), einschließlich ihrer Architektur, Bereitstellung, Benutzertypen, technischen Merkmale, Integration und Lizenzierung.

AIDE-Grundlagen

Was ist NetApp AI Data Engine (AIDE)?

NetApp AI Data Engine (AIDE) ist ein speicherintegrierter AI Data Engine-Datendienst, der den gesamten KI-Lebenszyklus abdeckt – von der Ermittlung und Aufbereitung von Rohdaten bis hin zur Bereitstellung von Retrieval-Endpunkten, um generative KI (GenAI), Retrieval-Augmented Generation (RAG), agentic AI und AI factories zu unterstützen. AIDE automatisiert Sync und Änderungserkennung und bietet eine einheitliche, aktuelle Ansicht ausgewählter Daten für die Datenentdeckung und Kuration.

Wie funktioniert AIDE?

AIDE integriert sich direkt mit NetApp ONTAP-Speichersystemen, um eine globale, strukturierte Sicht auf den gesamten NetApp Datenbestand mit automatisierter Änderungserkennung und Synchronisierung zu erstellen. AIDE bietet Echtzeit-Vektorisierung mit Komprimierung und Deduplizierung, richtliniengesteuerte Data Guardrails und Integration mit KI-Tools.

Benutzer und Rollen

Wer verwendet die AI Data Engine?

Zu den Hauptnutzern von AIDE gehören:

1. **ONTAP Speicheradministratoren:** Verwaltung der Infrastruktur, KI-spezifischer Speicheranforderungen, Sicherheit und Compliance.
2. **Dateningenieure:** Verwalten die Datenbewegung, -aufbereitung und -integration über verschiedene Umgebungen hinweg.
3. **Data Scientists:** Bereiten Sie die relevanten Daten für die Nutzung durch KI auf und transformieren Sie sie.

Anforderungen und Bereitstellung

Welche Hardware wird benötigt?

AIDE benötigt für die Bereitstellung AFX-Systeme (einschließlich AFX-Controller, Disk Shelf und Netzwerk-Switch), kann aber Clusterdaten von Clustern mit ONTAP 9 mithilfe von SnapMirror und Cluster-Peering nutzen. Für AIDE-Bereitstellungen sind mindestens vier AFX-Controller-Knoten erforderlich, um hohe Verfügbarkeit und Leistung sicherzustellen.

AIDE läuft auf einem NetApp data compute node (DCN). Drei DCNs sind erforderlich. Der DCN hostet die AIDE Software, die die Metadata Engine, Data Sync, Data Curator und Data Guardrails umfasst.

Kann ich mein eigenes DCN verwenden?

Nein. Das DCN ist ein von NetApp bereitgestellter Datenverarbeitungshardwareknoten und der einzige Bereitstellungsmechanismus für die AI Data Engine.

Was ist die minimale Anzahl an DCNs, die erforderlich ist?

Genau drei DCNs sind erforderlich.

Welches Betriebssystem läuft auf den DCNs?

Die DCNs führen einen von NetApp bereitgestellten Software-Stack mit AIDE aus.

Kann AIDE ohne AFX bereitgestellt werden?

Nein. AIDE benötigt AFX für die Bereitstellung. AIDE verwendet Trident, um die AFX-Volumes für den internen Speicher (persistente Volumes) zu nutzen. Der AFX-Cluster, der den Speicher für AIDE bereitstellt, kann mit einem ONTAP 9-System oder -Cluster verbunden werden. Es wird Cluster-Peering und SnapMirror verwendet, um Daten vom entfernten ONTAP-Cluster mit dem AFX-System zu synchronisieren.

Verwaltung und Schnittstellen

Ist die AIDE Console Teil der NetApp Console oder eine separate Managementoberfläche?

Die AIDE Console ist eine separate Managementoberfläche, die auf DCNs ausgeführt wird. Sie verwenden die AIDE Console, um AIDE Services wie Data Guardrails und Data Curator zu verwalten. Sie können auch ONTAP System Manager verwenden, um den AIDE Cluster zu überwachen.

Funktionen und Fähigkeiten

Was sind die wichtigsten Funktionen von AIDE?

Es gibt vier Hauptfunktionen von AIDE:

Metadata Engine

- Erzeugt automatisch eine strukturierte, aktuelle und interaktive Ansicht Ihrer Daten.
- Funktioniert mit Daten, die auf ONTAP gespeichert sind.
- Ermöglicht Datenexperten die Zusammenarbeit mit Speicheradministratoren, um Daten zu finden und zu verstehen.
- APIs fragen Metadaten ab, um Funktionen bereitzustellen und gleichzeitig die NFS-Traffic-Last auf Speichersystemen zu reduzieren.
- Die Funktion zur Metadatenextraktion und Katalogisierung ist speziell für AIDE entwickelt, arbeitet kontinuierlich und nutzt ONTAP-Funktionen wie Snapshots.

Datensynchronisierung

- Hält die Daten automatisch auf dem neuesten Stand, wenn sich die Quelldaten ändern, ohne manuelle Eingriffe.
- Administratoren legen das Aktualisierungsintervall in Tagen oder Stunden fest.
- Ermöglicht inkrementelle Datenmobilität und Synchronisierung der Daten, um redundante Kopien von AI-Daten zu eliminieren.

Data Guardrails

- Identifiziert und schützt sensible Daten automatisch während des gesamten KI-Lebenszyklus. Es ist über die AI Data Engine Console zugänglich.
- Scannt, klassifiziert und kategorisiert kontinuierlich Daten.
- Identifiziert sensible Daten (wie z. B. personenbezogene Daten) und Risiken.
- Ermöglicht die Erstellung von Richtlinien für die automatische Verarbeitung von sensiblen Daten im Einklang mit den Unternehmens- und Regulierungsstandards.

- Bietet automatische Schwärzung sensibler Informationen zum Schutz der Daten.
- Beschränkt den Zugriff auf sensible Dateien nach Bedarf.

Data Curator

- Ermöglicht Datenwissenschaftlern, den gesamten Speicher nach relevanten Daten zu durchsuchen.
- Erstellt kuratierte Datensammlungen mit Daten, die auf AFX Volumes vorhanden sind.
- Erzeugt Vektoreinbettungen auf der Speicherschicht, um Daten bloat zu reduzieren und die Leistung zu steigern.
- Bietet einen Abruf-Endpunkt für KI-Anwendungen mit vektorsemantischer Suche und Re-Ranking.

Integration und Interoperabilität

Unterstützt AIDE föderierte Metadaten über mehrere ONTAP Cluster hinweg?

AIDE kann über SnapMirror und Cluster-Peering eine Verbindung zu mehreren ONTAP-Clustern herstellen, wodurch eine zentrale Metadaten-Sichtbarkeit ermöglicht wird.

Wo werden die Metadaten gespeichert?

AIDE speichert Metadaten auf dem verbundenen AFX-Cluster mithilfe eines von AFX bereitgestellten persistenten Volumes. Die DCNs verwenden lokalen Speicher für interne Operationen.

Klassifiziert die AIDE Metadata Engine Daten?

Nein. Die Metadata Engine katalogisiert Dateisystem-Metadaten und stellt APIs zur Abfrage dieser katalogisierten Metadaten bereit.

Welche Datenquellen werden unterstützt?

AIDE unterstützt ONTAP Volumes (lokal oder remote) als Datenquellen. Remote ONTAP Cluster müssen ONTAP 9 ausführen und über Cluster-Peering und SnapMirror verbunden sein.

ONTAP S3 buckets und StorageGRID objects werden in AIDE 9.18.1 nicht als Datenquellen unterstützt.

Welche Dateitypen kann AIDE für Klassifizierung, Vektorisierung und semantische Suche verarbeiten?

AIDE unterstützt eine breite Palette von Dateitypen, einschließlich PDF, DOCX, PPTX, TXT und Bilddateien mit OCR-Funktionen.

Unterstützt AIDE die Klassifizierung nicht-englischsprachiger Daten?

AIDE unterstützt ausschließlich englischsprachige Daten.

Welche Integrationen unterstützt AIDE?

AIDE bietet einen RAG-API-Endpunkt, der über direkte API-Aufrufe oder über einen Model Context Protocol (MCP)-Server zugänglich ist. Dies unterstützt die Integration mit agentic AI Frameworks und Tools.

Bereitstellung und Lizenzierung

Welche Bereitstellungsoptionen gibt es?

AIDE wird lokal auf der AFX-Infrastruktur mit DCNs bereitgestellt. Es integriert sich direkt mit NetApp ONTAP AFX-Installationen.

Wie wird AIDE lizenziert?

AIDE erfordert eine Softwarelizenz, um Data Guardrails und Data Curator auszuführen.

Wenn Sie nur die Metadata Engine benötigen, berechtigt Sie die ONTAP One-Lizenz, die in allen AFX-Systemen enthalten ist, ausschließlich zur Nutzung der Metadata Engine-Funktionen.

Verwandte Informationen

- ["Installieren Sie AIDE-Lizenzen im ONTAP System Manager"](#)
- ["Erfahren Sie mehr über die AIDE-Architektur und ihre Komponenten"](#)

Rechtliche Hinweise

Rechtliche Hinweise bieten Zugriff auf Urheberrechtsvermerke, Marken, Patente und mehr.

Copyright

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

Marken

NETAPP, das NETAPP-Logo und die auf der NetApp Trademarks-Seite aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen- und Produktnamen können Marken ihrer jeweiligen Eigentümer sein.

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

Patente

Eine aktuelle Liste der NetApp owned patents finden Sie unter:

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

Datenschutzrichtlinie

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

Open Source

Hinweisdateien enthalten Informationen über Urheberrechte und Lizenzen Dritter, die in NetApp Software verwendet werden.

AI Data Engine

["Hinweis für AIDE 9.18.1"](#)

Copyright-Informationen

Copyright © 2026 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGEND EINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.