



Documentação do AI Data Engine

AI Data Engine

NetApp

March 13, 2026

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

Índice

Documentação do AI Data Engine	1
Notas de lançamento	2
O que há de novo no AI Data Engine	2
O que há de novo na versão inicial do AIDE 9.18.1	2
Limitações conhecidas do AI Data Engine	3
Limitações conhecidas para o lançamento inicial do AIDE 9.18.1	3
Comece agora	6
Saiba mais sobre o seu sistema AI Data Engine	6
Saiba mais sobre AI Data Engine	6
Arquitetura do AI Data Engine	7
Componentes e responsabilidades do AIDE por função	11
Guia rápido para AI Data Engine	16
Instale o AIDE	17
Requisitos para instalação do AI Data Engine	17
Instale o sistema de storage AFX para AI Data Engine	18
Instale seus nós de computação de dados	18
Configure seu sistema AIDE	28
Configurar AI Data Engine	28
Instale uma licença AIDE em seu sistema AFX	31
Configurar OpenID Connect para AIDE no ONTAP	32
Configurar espaços de trabalho	35
Preparar dados para AI Data Engine	35
Crie um workspace no AI Data Engine	36
Criar um espaço de trabalho	36
Revise os detalhes do espaço de trabalho	37
Atualizações e versionamento do espaço de trabalho	37
Atribuir acesso de usuário aos workspaces do AI Data Engine	38
Administrar e monitorar	40
Monitorar processos de cluster	40
Visualize o status do sistema AIDE e do cluster	40
Veja Insights para otimizar seu sistema AIDE	42
Visualizar eventos do AIDE, tarefas e log de auditoria	43
Gerenciar workspaces do AI Data Engine	45
Revisar o status do workspace	45
Editar propriedades do espaço de trabalho e agendamento de atualização	46
Adicionar contêineres de dados a um espaço de trabalho existente	46
Remova contêineres de dados de um espaço de trabalho	47
Gerenciar usuários do workspace	47
Excluir um workspace	47
Atualize e faça a manutenção do seu sistema AIDE	48
Atualizações e compatibilidade do sistema AI Data Engine	48
Atualizar o AI Data Engine software	49
Adicione nós de computação de dados ao seu cluster AIDE	50

Substitua um nó no seu AIDE cluster	52
Gerenciar vetorização e coleções de dados	55
Início rápido Data-to-RAG para AI Data Engine	55
Explore metadados de workspace no AI Data Engine Console	56
Sign in no AIDE Console como engenheiro de dados ou cientista de dados	56
Veja seus espaços de trabalho acessíveis	56
Qual é o próximo passo?	57
Crie coleções de dados no AI Data Engine Console	57
Criar uma coleta de dados a partir de metadados do workspace	57
Publicar uma coleta de dados	58
Atualizar ou excluir uma coleta de dados	58
Qual é o próximo passo?	59
Visualizar coleções de dados no AI Data Engine	59
Visualizar coleta de dados de todo o cluster	60
Monitorar trabalhos e eventos relacionados à coleta	60
Visualize as coleções de dados do AIDE Console	61
Qual é o próximo passo?	61
Implementar guardrails	62
Defina suas políticas de Data Guardrails no AI Data Engine para seu data estate	62
Entenda os tipos de políticas	62
Ativar classificadores	62
Gerenciar categorias de classificação	63
Criar e gerenciar políticas de Data Guardrails	64
Como as políticas interagem com os workspaces	65
Informações relacionadas	65
Perguntas frequentes sobre NetApp AI Data Engine	66
Noções básicas de AIDE	66
Usuários e funções	66
Requisitos e implantação	66
Gerenciamento e interfaces	67
Recursos e funcionalidades	67
Integração e interoperabilidade	68
Implantação e licenciamento	68
Avisos legais	70
Direitos autorais	70
Marcas registradas	70
Patentes	70
Política de privacidade	70
Código aberto	70
AI Data Engine	70

Documentação do AI Data Engine

Notas de lançamento

O que há de novo no AI Data Engine

AI Data Engine (AIDE) 9.18.1 é a versão inicial da plataforma de AI Data Engine da NetApp para gerenciamento de dados de IA. Esta versão introduz um mecanismo de metadados e fluxos de trabalho de gerenciamento que permitem às organizações catalogar e organizar dados não estruturados para cargas de trabalho de IA, fornecendo a base para recursos avançados de governança e vetorização. Governança avançada (guardrails) e vetorização estão disponíveis para clientes que possuem as licenças apropriadas do AI Data Engine.

O que há de novo na versão inicial do AIDE 9.18.1

AIDE 9.18.1 introduz as seguintes funcionalidades fundamentais:

"Metadata Engine para catalogação de dados de IA"

A versão inicial inclui um mecanismo de metadados que cataloga arquivos e objetos em clusters ONTAP.

Principais recursos incluem:

- Extração automatizada de metadados (atributos principais e estendidos, tags de objetos) de volumes ONTAP locais e remotos em clusters emparelhados.
- APIs REST centralizadas para consulta e filtragem para aplicações que exigem uma visão global dos dados corporativos.
- Armazenamento de metadados escalável.
- Extração automática de metadados acionada durante a criação do espaço de trabalho.

"Gerenciamento do workspace"

Os workspaces proporcionam um agrupamento lógico de fontes de dados (volumes) para projetos de IA.

A versão inicial oferece suporte a:

- Criação de espaços de trabalho que abrangem volumes ONTAP locais e remotos (usando o emparelhamento de clusters).
- Atribuição de controles de acesso a espaços de trabalho, com suporte a ambientes multiusuário e alocação a vários clientes.
- Extração automática de metadados e preenchimento do catálogo após a criação do espaço de trabalho.

"Sincronização de dados para atualização automática de dados"

O Data Sync mantém os catálogos de metadados e as coleções de dados atualizados conforme os dados de origem são alterados, sem intervenção manual.

Principais recursos incluem:

- Sincronização automatizada de dados de clusters ONTAP remotos ou locais usando replicação voltada a políticas SnapMirror.
- Atualizações incrementais que propagam apenas dados modificados, reduzindo a sobrecarga.

- Intervalos de atualização configuráveis por espaço de trabalho.
- Monitoramento do status e da atividade de sincronização em nível de workspace.

"Configuração e gerenciamento de clusters"

A versão inicial inclui os seguintes fluxos de trabalho:

- Descoberta e adição de nós de computação de dados (DCN) durante a configuração do cluster.
- Criação de máquinas virtuais de armazenamento de metadados dedicadas para o Metadata Engine.
- Configuração das interfaces de serviço do Data Engine para acesso a metadados em todo o cluster.
- Interligação com outros ONTAP clusters para estender a catalogação de metadados em todo o ambiente de dados.

"Autenticação OpenID Connect (OIDC)"

- Autenticação baseada em OIDC/OAuth para acesso seguro ao ONTAP System Manager e ao Data Engine Console com Microsoft Entra ID e Active Directory Federation Services (ADFS).
- Controles de acesso baseados em funções para gerenciamento de espaço de trabalho e metadados.

"Capacidades avançadas de curadoria e governança de dados"

As seguintes funcionalidades estão disponíveis para clientes que possuem as licenças apropriadas do AI Data Engine:

- **Vetorização e RAG:** Criação de coleções de dados, incorporações e pontos de extremidade de recuperação no AI Data Engine Console, usando metadados dos workspaces do AIDE.
- **Governança baseada em guardrails:** definição de políticas de guardrail no AI Data Engine Console e associação dessas políticas com workspaces no ONTAP System Manager.

Hardware e plataformas compatíveis

AI Data Engine 9.18.1 é executado em clusters da plataforma de dados ONTAP AI que combinam:

- Nós de armazenamento AFX 1K
- Nós de computação de dados NetApp

Informações relacionadas

- ["Limitações conhecidas do AI Data Engine"](#)
- ["Saiba mais sobre a arquitetura e os componentes do AI Data Engine"](#)

Limitações conhecidas do AI Data Engine

As limitações conhecidas identificam plataformas, dispositivos ou funções que não são compatíveis com esta versão do produto ou que não interoperam corretamente com ela. Analise essas limitações cuidadosamente.

Limitações conhecidas para o lançamento inicial do AIDE 9.18.1

Essas limitações se aplicam ao Metadata Engine, aos nós de computação de dados e aos fluxos de trabalho de gerenciamento em AIDE 9.18.1.

Requisitos e gerenciamento de nós de computação

- **Requisito mínimo de nó de computação de dados**

Os clusters AIDE requerem um mínimo e um máximo de 3 nós de computação de dados (DCNs) para a funcionalidade do Metadata Engine. Clusters com menos de 3 DCNs não podem ativar os recursos do Metadata Engine.

- **Não há suporte para adicionar nós DCN usando NetApp Console**

As atualizações e adições de nós DCN devem ser realizadas usando ONTAP System Manager, não através do NetApp Console.

Fontes de dados suportadas

- **Não há suporte para buckets ONTAP S3 ou StorageGRID como fontes de dados**

Somente volumes ONTAP (locais ou remotos) são suportados como fontes de dados para espaços de trabalho e catalogação de metadados. Buckets ONTAP S3 e objetos StorageGRID não podem ser adicionados a espaços de trabalho e não são indexados pelo Metadata Engine nesta versão.

- **Não há suporte para a criação de espaços de trabalho com volumes FlexCache**

FlexCache volumes não podem ser adicionados como fontes de dados aos workspaces.

Limitações de atualização e reversão de software

- **Atualizações manuais de software apenas para nós DCN**

As atualizações automáticas de software para DCN clusters não são suportadas no AIDE 9.18.1. O software do nó DCN só pode ser atualizado fazendo upload da imagem a partir de um cliente local. O download de imagens de servidores externos (HTTP/FTP) não é suportado.

- **Não é possível reverter o software do cluster DCN**

O software do cluster DCN não pode ser revertido para uma versão anterior. Somente upgrades para versões posteriores são permitidos.

- **Sem reversão ONTAP de clusters de armazenamento AFX**

Os clusters de armazenamento AFX não podem ser revertidos para versões anteriores do ONTAP. Somente atualizações para versões mais recentes são permitidas.

Ciclo de vida do workspace e configuração de acesso

- **Sem exclusão lógica ou restauração para espaços de trabalho**

A exclusão de um espaço de trabalho é permanente. Não há opção para restaurar espaços de trabalho excluídos.

- **Não há suporte para configuração OIDC durante a configuração inicial do cluster**

A configuração OIDC/OAuth deve ser realizada após a criação do cluster usando ONTAP System Manager.

Informações relacionadas

- ["O que há de novo no AI Data Engine"](#)

Comece agora

Saiba mais sobre o seu sistema AI Data Engine

Saiba mais sobre AI Data Engine

O NetApp AI Data Engine (AIDE) é uma plataforma de nível empresarial projetada para acelerar e simplificar o processamento de dados, o gerenciamento e a governança de dados orientados por IA. AIDE pode ajudar a transformar grandes quantidades de dados não estruturados em conjuntos de dados estruturados e prontos para IA. Ele foi desenvolvido para atender às demandas das cargas de trabalho modernas de aprendizado de máquina (ML) e IA generativa (GenAI), oferecendo suporte tanto às operações tradicionais de TI quanto às novas funções centradas em IA.

AIDE aborda os desafios da IA

AIDE foi projetado para ajudar as organizações a gerenciar dados para cargas de trabalho de IA e oferece as seguintes funcionalidades principais:

- **Gestão centralizada de metadados:** AIDE coleta e cataloga metadados de volumes ONTAP, possibilitando a busca, classificação e aplicação de políticas de governança a conjuntos de dados.
- **Processamento de dados automatizado:** AIDE oferece suporte à criação de fluxos de dados para cargas de trabalho de IA e ML, incluindo a capacidade de gerar incorporações vetoriais para pesquisa semântica (com o licenciamento apropriado).
- **Isolamento de dados e controle de acesso:** AIDE aplica controles de acesso e isolamento básico de dados para várias equipes ou projetos.
- **Integração com NetApp tools:** AIDE funciona com ONTAP System Manager para administração de storage e fornece uma interface (AI Data Engine Console) para que engenheiros e cientistas de dados gerenciem coleções de dados e fluxos de trabalho.

Características de projeto de alto nível

As seguintes características de design definem como AI Data Engine foi construído para atender às necessidades das cargas de trabalho de AI:

- **Serviços baseados em microsserviços:** Usa o Kubernetes para orquestrar serviços modulares e resilientes para catalogação de metadados, busca vetorial e gerenciamento de infraestrutura.
- **Segurança de nível empresarial:** Implementa criptografia, controle de acesso baseado em funções (RBAC) e auditoria em todos os dados e metadados.
- **Acesso a dados multiprotocolo:** Suporta NFS e SMB para ingestão e recuperação flexíveis de dados.
- **Fluxos de dados automatizados:** Rastreia alterações de dados, cria embeddings e gerencia bancos de dados vetoriais para aplicações de IA.

Como os dados fluem pelo AIDE

Compreender como os dados fluem pelo AIDE ajuda a ilustrar o valor da plataforma para equipes de AI/ML:

1. **Ingestão de dados:** Os arquivos são armazenados em volumes ONTAP usando protocolos padrão (NFS

e SMB). Os dados podem residir no armazenamento local do AIDE (o cluster AFX dentro da sua implantação do AIDE) ou em clusters ONTAP remotos. Os dados dos clusters remotos são sincronizados com o cluster AFX local usando ONTAP SnapMirror, de modo que todos os dados processados pelo AIDE sejam, em última análise, armazenados e acessados localmente.



Buckets S3 não são suportados como fontes de dados para workspaces ou coletas de dados.

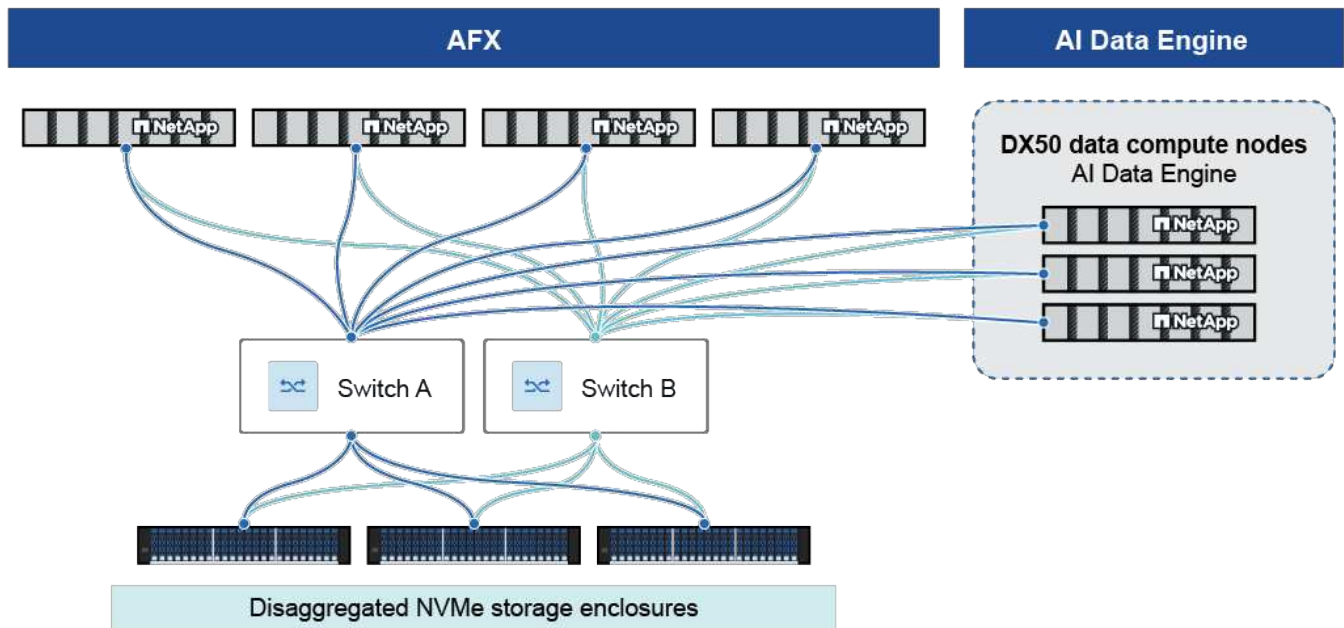
1. **Criação de workspaces:** Os administradores de armazenamento definem workspaces no ONTAP System Manager, agrupando volumes ONTAP relacionados para projetos, equipes ou fluxos de trabalho específicos. As permissões de acesso e as políticas de governança são atribuídas no nível do workspace.
2. **Extração de metadados:** AIDE examina automaticamente arquivos e objetos em espaços de trabalho, extraíndo metadados (tipo de arquivo, tamanho, carimbos de data/hora, atributos personalizados) e armazenando-os em um catálogo centralizado. Isso ocorre continuamente à medida que os dados são alterados.
3. **Classificação e governança:** Classificadores examinam os dados em busca de informações sensíveis (PII, dados financeiros) ou tipos de documento (jurídicos, RH). Políticas de Data Guardrails aplicam redação ou restrições de acesso automaticamente.
4. **Criação de coleta de dados:** engenheiros e cientistas de dados usam o AI Data Engine Console para consultar o catálogo de metadados, filtrar resultados e reunir coletas de dados selecionadas para tarefas específicas de IA.
5. **Vetorização:** para coleções que exigem busca semântica, AIDE gera embeddings usando modelos de IA selecionados. Vetores são armazenados no banco de dados de vetores para recuperação de alto desempenho.
6. **Consumo de IA/ML:** aplicativos acessam dados por meio de múltiplos caminhos:
 - Acesso direto a arquivo/objeto usando NFS ou SMB
 - Consultas de pesquisa semântica no banco de dados de vetores
 - Pontos de extremidade RAG que combinam recuperação de dados com integração de modelo GenAI
 - Acesso à API REST para fluxos de trabalho programáticos

Esse fluxo de trabalho automatizado e voltado a políticas reduz o tempo e o esforço manual necessários para preparar os dados para IA, permitindo que as equipes se concentrem no desenvolvimento de modelos e insights, em vez de na manipulação de dados.

Arquitetura do AI Data Engine

AIDE é construído sobre uma arquitetura escalável e tolerante a falhas que separa storage e processamento, permitindo alto desempenho e flexibilidade para cargas de trabalho de IA.

Componentes físicos



Nós controladores AFX

Os nós controladores AFX executam uma personalidade especializada do ONTAP projetada para atender aos requisitos do ambiente AFX. Os clientes acessam os nós por meio de múltiplos protocolos, incluindo NFS e SMB. Cada nó possui uma visão completa do storage, ao qual pode acessar com base nas solicitações do cliente. Os nós são stateful, com memória não volátil para persistir informações críticas de estado e incluem melhorias adicionais específicas para as cargas de trabalho de destino.

São necessários pelo menos quatro nós controladores AFX para implantações do AIDE, a fim de garantir alta disponibilidade e desempenho.

Nós de computação de dados

Os nós de computação de dados (DCNs, na sigla em inglês) são servidores baseados em Linux com recursos avançados de CPU, RAM e GPU, dedicados a tarefas de processamento de dados de IA. Eles hospedam serviços específicos de IA, como catalogação de metadados, busca vetorial e pipelines de incorporação.

São necessários exatamente *três* DCNs para implantações do AIDE.

Switches de cluster/storage

Switches redundantes de alta velocidade (100GbE ou superior) conectam ONTAP e DCNs para transferência de dados com baixa latência e alta disponibilidade.

Prateleiras de storage

Os chassis NVMe-oF com SSDs de alta densidade oferecem latência ultrabaixa e redundância, suportando armazenamento em escala de PB.

Rede

Todos os DCNs e nós de armazenamento ONTAP são conectados por meio de switches de cluster redundantes e de alta velocidade (mínimo 100GbE). Essa arquitetura separa recursos de computação e de storage, permitindo que cada um seja dimensionado independentemente e otimizando tanto o desempenho quanto a utilização de recursos.

A comunicação entre DCNs e os nós do ONTAP é isolada por meio de VLANs e IPspaces dedicados nos switches do cluster. Isso garante que todas as comunicações, como acesso a dados, APIs de gerenciamento e tráfego de serviços internos, permaneçam seguras, eficientes e não interfiram em outras operações de rede.

Principais funcionalidades do AI Data Engine

Os principais recursos do AI Data Engine (AIDE) trabalham juntos para automatizar, proteger e acelerar o ciclo de vida dos dados de IA. Cada recurso é implementado como um conjunto de microserviços em execução em DCNs, integrados ao storage ONTAP e expostos por meio de APIs REST e interfaces de gerenciamento.

Metadata Engine

O Metadata Engine gera automaticamente uma visão estruturada, atualizada e interativa do seu conjunto de dados NetApp.

Licença e acesso

O Metadata Engine está incluído na licença básica do ONTAP One e está disponível após a instalação do AIDE.

Você pode acessá-lo por meio do ONTAP System Manager.

Capacidades

- Cataloga metadados de todas as fontes de dados, incluindo volumes armazenados localmente no cluster AFX e aqueles sincronizados de clusters ONTAP remotos.
- Extrai metadados automaticamente e preenche o catálogo à medida que os dados são ingeridos ou alterados.
- Fornece acesso à API REST para consulta de metadados, permitindo que profissionais de dados e administradores de storage descubram, classifiquem e compreendam os dados.
- Descarrega as consultas de metadados do caminho de dados, reduzindo a carga de tráfego NFS nos sistemas de storage.
- Suporta grandes registros de metadados com recursos de indexação e pesquisa.
- Integra-se com abstrações de espaço de trabalho e coleta de dados para aplicar controle de acesso e governança.

Sincronização de dados

O Data Sync é um serviço automatizado em segundo plano que garante que o catálogo de metadados e as coleções de dados permaneçam atualizados e consistentes com as fontes de dados subjacentes, mesmo quando os dados de origem são alterados.

Licença e acesso

A funcionalidade de sincronização de dados está incluída na licença básica do ONTAP One e fica disponível após a instalação do AIDE.

Capacidades

- Sincroniza dados de clusters ONTAP remotos ou locais usando replicação SnapMirror voltada a políticas. Os dados dos clusters remotos são copiados para o cluster AFX local para processamento AIDE.
- Atualiza incrementalmente com base nas alterações detectadas, propagando apenas os dados modificados.
- Proporciona mobilidade de dados segura e incremental e sincronização em todo o ambiente de dados.

- Agenda e monitora intervalos de sincronização com taxas de atualização configuráveis por workspace.
- Integra-se aos fluxos de trabalho de criação de workspaces para extrair e atualizar metadados à medida que novas fontes de dados são adicionadas.

Data Guardrails

O serviço Data Guardrails oferece governança e proteção contínuas e automatizadas para dados confidenciais ao longo de todo o ciclo de vida da IA.

Licença e acesso

A funcionalidade Data Guardrails não está incluída na licença básica do ONTAP One e requer uma licença AIDE separada.

Você pode acessar a funcionalidade de Data Guardrails através do AI Data Engine Console.

Capacidades

- Analisa, classifica e categoriza dados continuamente.
- Identifica dados confidenciais e riscos usando classificadores incorporados e personalizáveis para tarefas como detecção de PII.
- Automatiza o tratamento de dados confidenciais por meio de redação, mascaramento e restrições de acesso voltados a políticas.
- Garante o cumprimento dos padrões da empresa e dos regulamentos por meio de políticas de Data Guardrails vinculadas aos workspaces.
- Restringe o acesso a arquivos ou volumes confidenciais conforme configurado, com registro de auditoria e relatórios de conformidade.
- Integra-se com o gerenciamento de espaço de trabalho e coleta de dados para aplicar guardrails de forma consistente em todos os fluxos de trabalho de dados de IA.

Data Curator

O serviço Data Curator permite a descoberta, busca, vetorização e recuperação rápidas de dados para aplicações de IA e GenAI.

Licença e acesso

A funcionalidade Data Curator não está incluída na licença básica do ONTAP One e requer uma licença AIDE separada.

Você pode acessar o Data Curator através do AI Data Engine Console.

Capacidades

- Pesquisa o storage por dados relevantes usando o catálogo de metadados centralizado.
- Fornece ferramentas para cientistas de dados criarem coleções de dados curadas.
- Gera representações vetoriais automaticamente na camada de storage.
- Fornece um ponto de extremidade de recuperação seguro para aplicações de IA, suportando busca semântica vetorial e reclassificação.
- Integra-se com ferramentas e tecnologias de IA, incluindo pipelines de Geração Aumentada por Recuperação (RAG) e frameworks de IA agentic.
- Fornece API REST para acesso programático a coleções de dados, pesquisa vetorial e endpoints de recuperação.

Segurança e alocação a vários clientes

A plataforma aplica tanto o controle de acesso baseado em funções (RBAC) quanto as listas de controle de acesso em nível de recurso (ACLs). Todas as ações de API e de usuários são auditadas, e todos os dados são criptografados em repouso e em trânsito. Os locatários individuais são isolados para dados e metadados.

Informações relacionadas

- ["Instale uma licença AIDE"](#)
- ["Início rápido de Data-to-RAG"](#)

Componentes e responsabilidades do AIDE por função

Componentes do AI Data Engine e interações baseadas em funções

AI Data Engine (AIDE) consiste em vários componentes principais que trabalham em conjunto para fornecer uma plataforma abrangente de gerenciamento e processamento de dados para cargas de trabalho de IA. Esses componentes incluem workspaces, coleções de dados, bancos de dados vetoriais, guardrails, catálogos de metadados, endpoints de recuperação e classificadores. Cada componente desempenha um papel específico para permitir a descoberta, curadoria, governança e integração eficientes de dados com aplicações de IA/ML.

Cada usuário do AIDE interage com os componentes do AIDE de maneira diferente, de acordo com sua função.

Funções de usuário focadas em storage e dados

AIDE introduz novas funções de usuário, mantendo o suporte às funções tradicionais de administração do sistema ONTAP:

Usuários de storage

- **Administrador de storage:** Gerencia a configuração do cluster AFX e AIDE, rede, provisionamento de storage e acesso do usuário.

Usuários de dados

- **Engenheiro de dados:** constrói e otimiza pipelines de IA/ML, gerencia coleções de dados e integra modelos de IA.
- **Cientista de dados:** Descobre, organiza e analisa conjuntos de dados, cria coleções de dados e utiliza pontos de acesso de recuperação para aplicações GenAI.

Função (nome RBAC)	Descrição
Administrador de storage (<code>admin</code>)	Gerencia a configuração do cluster AFX e AIDE, rede, provisionamento de storage e acesso do usuário. Atribui funções RBAC aos usuários que determinam o nível de acesso às interfaces e recursos do AIDE. Essa função de administrador tem acesso total de gerenciamento usando ONTAP System Manager e AI Data Engine Console.

Função (nome RBAC)	Descrição
Engenheiro de dados (data-engineer)	Cria e otimiza pipelines de IA/ML, gerencia coletas de dados e integra modelos de IA. Esta função tem acesso ao AI Data Engine Console para fluxos de trabalho de engenharia de dados.
Cientista de dados (data-scientist)	Descobre, organiza e analisa conjuntos de dados, cria coleções de dados e utiliza pontos de recuperação para aplicações GenAI. Esta função tem acesso ao AI Data Engine Console para fluxos de trabalho de ciência de dados.

Componentes do sistema AIDE

Cada usuário do AIDE (administradores de storage, engenheiros de dados e cientistas de dados) interage com os componentes do AIDE de acordo com sua função.

Espaços de trabalho

Um espaço de trabalho é um segmento lógico de dados dentro do cluster, agrupando volumes para um projeto, equipe ou fluxo de trabalho específico. Os espaços de trabalho definem o escopo da visibilidade, do acesso e da governança dos dados em AIDE.

Catálogo de metadados

Um banco de dados centralizado e escalável que armazena registros de metadados para todos os arquivos e objetos no cluster local, incluindo dados sincronizados de clusters ONTAP remotos usando ONTAP SnapMirror ou peering de cluster. Ele permite busca e filtragem interativas e avançadas.

Classificadores

Os classificadores são ferramentas (incorporado ou personalizadas) que analisam e etiquetam arquivos para identificar tipos específicos de dados confidenciais (por exemplo, PII, financeiro, saúde) ou categorizam documentos por tipo (por exemplo, legal, RH, vendas).

Coletas de dados

Uma coleta de dados é um grupo selecionado de arquivos ou objetos relacionados de um workspace, definido por uma consulta especificada pelo usuário para uso em fluxos de trabalho do GenAI. O conteúdo dos arquivos na coleta de dados, após a publicação, fica disponível para busca semântica por APIs para aplicações do GenAI.

Banco de dados vetorial

O banco de dados vetorial armazena embeddings gerados a partir de coleções de dados, permitindo busca e recuperação semântica de alto desempenho para aplicações de AI e GenAI.

Guardrails

Os guardrails são mecanismos voltado a políticas que aplicam governança de dados, classificação e proteção (como redação ou restrições de acesso) ao longo de todo o ciclo de vida dos dados de IA.

Endpoint de recuperação (endpoint RAG)

Um endpoint de recuperação (às vezes chamado de endpoint de Retrieval-Augmented Generation ou "RAG") é uma API segura que permite que aplicativos de AI e GenAI acessem dados, contexto ou embeddings

relevantes de coleções selecionadas e do banco de dados vetorial.

Os endpoints RAG são projetados para suportar fluxos de trabalho avançados de IA, como busca semântica e respostas contextuais em modelos generativos de IA. Ao conectar seus aplicativos de IA a um endpoint de recuperação, você pode aprimorar a precisão e a relevância do modelo, fornecendo acesso em tempo real a conjuntos de dados selecionados e prontos para IA, gerenciados pelo AIDE.

Informações relacionadas

- ["Como os administradores de storage do AIDE trabalham com os componentes do AIDE"](#)
- ["Como os engenheiros de dados da AIDE trabalham com os componentes da AIDE"](#)
- ["Como os cientistas de dados da AIDE trabalham com os componentes da AIDE"](#)

Interfaces do AI Data Engine

AI Data Engine (AIDE) fornece três interfaces principais para interação do usuário e automação. Cada função, como administradores de storage, engenheiros de dados e cientistas de dados, utiliza essas interfaces de acordo com suas tarefas e responsabilidades específicas.

ONTAP System Manager

ONTAP System Manager é uma interface Web projetada para administradores de storage. Ele fornece fluxos de trabalho para configuração de cluster, gerenciamento de espaço de trabalho, monitoramento de DCN e aplicação de políticas de guardrail.

AI Data Engine Console

AI Data Engine Console é uma interface dedicada para engenheiros de dados e cientistas de dados. Ela permite que os usuários explorem fontes de dados, criem e gerenciem coleções de dados, configurem pipelines de dados, apliquem classificadores e interajam com recursos de proteção e busca vetorial. O console fornece ferramentas avançadas para descoberta, curadoria e integração com fluxos de trabalho de IA/ML.

API REST

AIDE expõe a API REST do ONTAP para automação, integração e acesso programático. A API oferece suporte à configuração de clusters, gerenciamento de espaços de trabalho e coleções, consultas de metadados, busca vetorial e endpoints de recuperação.

Saiba como os administradores de storage do AI Data Engine trabalham com os componentes do AIDE

Como administrador de storage, você gerencia a infraestrutura do AIDE por meio do ONTAP e do AIDE Console, provisionando workspaces, aplicando políticas de guardrail e monitorando a integridade do sistema. Sua função se concentra em garantir um storage confiável, seguro e em conformidade para cargas de trabalho de IA.

Acesso do administrador de storage ao componente

Componente	Nível de acesso	Fluxo de trabalho do administrador de storage
ONTAP System Manager	Gerenciar (criar, editar, excluir)	Você utiliza ONTAP System Manager como sua interface principal para administração de cluster, provisionamento de workspace, gerenciamento de políticas de guardrail e monitoramento da integridade do sistema.
AI Data Engine Console	Gerenciar (criar, editar, excluir)	Você utiliza o AI Data Engine Console para monitorar espaços de trabalho, visualizar o status da coleta e supervisionar a atividade do sistema em todo o ambiente AIDE.
API REST do ONTAP	Gerenciar (criar, editar, excluir)	Você utiliza a API REST para automatizar tarefas de infraestrutura, gerenciar workspaces e políticas de guardrail de forma programática e integrar a administração do AIDE com ferramentas e fluxos de trabalho externos.
Espaços de trabalho	Gerenciar (criar, editar, excluir)	Você cria e gerencia espaços de trabalho usando ONTAP System Manager. Você seleciona quais fontes de dados são incluídas, atribui permissões a engenheiros de dados e cientistas de dados, e vincula políticas de proteção para impor governança e conformidade. Você também monitora a integridade do espaço de trabalho e o acesso.
Coletas de dados	Visualizar (somente leitura)	Você visualiza o status e a integridade das coleções de dados em cada espaço de trabalho usando System Manager. Você garante que as fontes de dados subjacentes estejam disponíveis e protegidas, mas não cria ou modifica coleções.
Guardrails	Gerenciar (criar, editar, excluir)	Você define e associa políticas de guardrail a workspaces usando System Manager. Você monitora o status do guardrail e os relatórios de conformidade. Você garante que as políticas sejam aplicadas e atualizadas conforme necessário.
Catálogo de metadados	Monitorar (visualizar saúde, status, atividade)	Você garante que o catálogo de metadados esteja preenchido e atualizado. Você monitora a integridade do catálogo e oferece suporte ao controle de acesso.
Banco de dados vetorial	Provisionar/Monitorar (implantar, configurar, visualizar status)	Você provisiona e monitora a infraestrutura do banco de dados vetorial, garantindo que os nós de computação de dados com recursos de GPU e licenciamento adequado estejam em vigor. Você dá suporte ao ambiente, mas não gerencia embeddings ou consultas diretamente.

Componente	Nível de acesso	Fluxo de trabalho do administrador de storage
Classificadores	Gerenciar (criar, editar, excluir)	Você cria, configura e gerencia classificadores e suas categorias. Você aplica classificadores a espaços de trabalho e monitora sua eficácia.

Aprenda como engenheiros de dados e cientistas de dados do AI Data Engine trabalham com os componentes do AIDE

Como engenheiro de dados ou cientista de dados, você usa o AI Data Engine Console para explorar workspaces aos quais você recebeu acesso, criar e gerenciar data collections, realizar buscas semânticas e integrar retrieval endpoints em workflows de IA/ML.

Os engenheiros de dados se concentram em transformar dados brutos em conjuntos de dados prontos para IA, criando coleções, configurando pipelines de incorporação e controlando quais usuários podem acessar as coleções publicadas. Os cientistas de dados se concentram em aproveitar conjuntos de dados selecionados para análise, treinamento de modelos e aplicações GenAI, sem gerenciar controle de acesso ou infraestrutura.

Acesso do usuário de dados ao componente

Componente	Nível de acesso	Fluxo de trabalho do engenheiro de dados	Fluxo de trabalho do cientista de dados
AI Data Engine Console	Gerenciar (criar, editar, excluir)	O AI Data Engine Console é sua interface principal para tarefas do dia a dia, incluindo descoberta de dados, gerenciamento de coleta, configuração de pipeline e publicação de endpoints RAG ou de recuperação, para os workspaces que você está autorizado a acessar.	O AI Data Engine Console é sua interface principal para exploração de dados, refinamento e versionamento de coleções dentro dos workspaces aos quais você pode acessar, além de conectar conjuntos de dados selecionados e endpoints de recuperação a fluxos de trabalho de análise, modelagem e GenAI.
API REST do ONTAP	Gerenciar (criar, editar, excluir)	Você utiliza a API REST para automatizar operações do ciclo de vida da coleta, acionar e monitorar pipelines de incorporação e integrar programaticamente fluxos de trabalho de dados com ferramentas externas.	Você utiliza a API REST para acessar programaticamente coleções de dados, executar consultas de busca vetorial e integrar endpoints de recuperação em aplicações de IA/ML e frameworks agentic.
Espaços de trabalho	Visualizar/usar (somente leitura)	Você explora seus espaços de trabalho atribuídos para identificar e compreender as fontes de dados disponíveis antes de criar coleções.	Você pesquisa em seus espaços de trabalho atribuídos para localizar arquivos e objetos relevantes para tarefas específicas de pesquisa ou modelagem.

Componente	Nível de acesso	Fluxo de trabalho do engenheiro de dados	Fluxo de trabalho do cientista de dados
Coletas de dados	Gerenciar (criar, editar, excluir)	Você cria coleções de dados selecionando e filtrando dados de origem usando tags, classificação e outros atributos, e gerencia todo o ciclo de vida da coleção, desde a criação e versionamento até a publicação como endpoints RAG para uso em IA. Você também gerencia quais cientistas de dados e outros usuários podem acessar cada coleção.	Você cria, seleciona, anota, versiona e refina coleções de dados dentro dos espaços de trabalho aos quais você tem acesso. Você usa essas coleções como base para busca semântica e fluxos de trabalho GenAI.
Catálogo de metadados	Consultar/utilizar (consumir para fluxos de trabalho)	Você utiliza o catálogo de metadados para avaliar e selecionar fontes de dados para ingestão, executando consultas para localizar arquivos relevantes e confirmar se eles atendem aos requisitos das coleções que você está criando em seus espaços de trabalho atribuídos.	Você pesquisa e filtra metadados nos workspaces aos quais tem acesso para localizar arquivos e objetos necessários para análise ou treinamento de modelos, contando com a estrutura de catálogo que foi criada e é mantida por engenheiros de dados.
Banco de dados vetorial	- Gerenciar embeddings/busca (engenheiro de dados) - Usar/pesquisar (cientista de dados)	Você aciona pipelines de incorporação, monitora o status da vetorização, configura parâmetros de fragmentação e incorporação e expõe endpoints de recuperação com suporte de busca vetorial. Aplicativos e agentes então consultam esses endpoints via a API para busca semântica e fluxos de trabalho RAG.	Você executa consultas de busca semântica em embeddings gerados por pipelines gerenciados por engenheiros de dados e integra os resultados da recuperação em fluxos de trabalho GenAI ou RAG para respostas de modelos sensíveis ao contexto. Você não configura o particionamento, os embeddings ou os parâmetros do pipeline.
Classificadores	Usar (consumir dados classificados)	Você utiliza os resultados da classificação para anotar e etiquetar os dados de origem durante a preparação da coleta, garantindo que o conteúdo que entra em seus pipelines esteja devidamente rotulado para fluxos de trabalho de IA downstream.	Você consome dados pré-classificados para garantir que apenas conteúdo em conformidade e relevante seja usado em suas análises e modelagem.

Guia rápido para AI Data Engine

Para colocar seu AI Data Engine system em funcionamento, você precisa instalar os

componentes de hardware, configurar seu cluster, configurar o acesso aos dados dos seus hosts ao sistema de storage e provisionar seu armazenamento.

1

AIDE será instalado em um cluster AFX novo ou já existente?

Você precisa decidir se AIDE e AFX serão instalados juntos ao mesmo tempo ou se AIDE será integrado a um cluster AFX existente.

2

Instale e configure seu hardware

"[Instale e configure](#)" seus nós de computação do cluster AIDE. Dependendo do ambiente de instalação, certifique-se também de "[instalar](#)" o hardware AFX.

3

Configure seu cluster

Use ONTAP System Manager para guiá-lo por um processo rápido e fácil para "[Configure o AIDE com um cluster AFX](#)".

4

Configure os espaços de trabalho e o acesso aos dados

"[Configure um espaço de trabalho e usuários para esse espaço de trabalho que podem acessar os dados do AI Data Engine](#)".

Qual é o próximo passo?

Agora você pode usar ONTAP System Manager para gerenciar seu AI Data Engine e fazer com que seus engenheiros de dados e cientistas de dados comecem com seus workspaces e configurações.

Instale o AIDE

Requisitos para instalação do AI Data Engine

Analise os requisitos para a instalação do AI Data Engine. AIDE requer um sistema de storage AFX, no mínimo três Data Compute Node, switches de rede e cabos.

Requisitos de hardware

O AI Data Engine requer um sistema de storage AFX e um mínimo de três Data Compute Nodes. O sistema AFX fornece a infraestrutura de storage, enquanto os Data Compute Nodes hospedam os componentes de software AIDE que permitem o gerenciamento, a curadoria e os recursos de IA dos dados.

- **AFX sistema de storage:** Inclui um controlador AFX, um compartimento para discos e um switch de rede. O sistema de storage AFX é necessário para a implantação do AIDE.
- **Nós de computação de dados:** É necessário um mínimo de três nós de computação de dados. Os nós de computação de dados são nós de hardware fornecidos pela NetApp que hospedam o software AIDE, incluindo o Metadata Engine, Data Sync, Data Curator e Data Guardrails.
 - Cada Data Compute Node possui os slots de E/S 4 e 5 disponíveis para conectividade. O slot 3 é reservado para a GPU. Os slots 1 e 2 não estão ocupados nem acessíveis.

- As portas e4a e e5a são usadas para conexões de cluster.
- As portas e4b e e5b são usadas para conexões de rede do host.

Requisitos do switch de rede

AI Data Engine requer switches de rede para habilitar a conectividade de rede do host e a comunicação entre os Data Compute Node.

- Switches de cliente para conectividade de rede do host (Cisco Nexus 9332D-GX2B ou Cisco Nexus 9364D-GX2A)
- Switches de cluster para comunicação entre nós (Cisco Nexus 9332D-GX2B ou Cisco Nexus 9364D-GX2A)
- Switches de gerenciamento opcionais para gerenciamento de rede

Requisitos de cabeamento

Os seguintes cabos são necessários para conectar os Data Compute Node aos switch de rede e à rede de gerenciamento.

- Cabos breakout de 400-GbE para 100-GbE (4x100GbE) para conectar nós a switches de cliente e de cluster

Suporte a múltiplos clusters

Após o AIDE ser implantado com o AFX, ele pode se conectar e gerenciar dados de outros clusters ONTAP 9.18.1 e posteriores usando SnapMirror e peering de cluster.

Instale o sistema de storage AFX para AI Data Engine

Instale o sistema de storage AFX como primeiro passo na implantação do AI Data Engine. O sistema de storage AFX fornece a infraestrutura de storage fundamental e é necessário antes de instalar os data compute nodes.

Siga o ["Documentação de instalação do AFX 1K"](#) para instalar o sistema de storage AFX.

O que vem a seguir

Após concluir a instalação do sistema de storage AFX, ["instale os Data Compute Node"](#).

Instale seus nós de computação de dados

Fluxo de trabalho de instalação e configuração para nós de computação de dados para AI Data Engine

Para instalar e configurar seus nós de computação de dados (DCN), você revisa os requisitos de hardware, prepara seu local, instala e conecta os componentes de hardware, liga o sistema e configura seu cluster ONTAP.



"Revise os requisitos de instalação do hardware"

Certifique-se de que você já possui um sistema de storage AFX 1K instalado e, em seguida, revise os requisitos de hardware para instalar os nós de computação de dados para AIDE. Para obter informações

sobre a instalação do sistema de storage AFX 1K, consulte ["Documentação de instalação do sistema de storage AFX 1K"](#).

2

"Prepare-se para instalar seus nós de computação de dados"

Para se preparar para instalar seus nós de computação de dados, você precisa organizar o local, verificar os requisitos ambientais e elétricos e garantir que haja espaço suficiente no gabinete. Em seguida, desembale o equipamento, compare o conteúdo com o packing slip e registre o hardware para acessar os benefícios de suporte. São necessários no mínimo três nós de computação de dados para implantar o AI Data Engine.

3

"Instale o hardware para seus nós de computação de dados"

Instale os kits de trilhos para seus nós de computação de dados. Fixe seus nós de computação de dados dentro do gabinete. Por fim, conecte os dispositivos de gerenciamento de cabos na parte traseira do sistema para organização dos cabos.

4

"Conecte seus nós de computação de dados"

Para conectar o hardware, primeiro conecte os nós de computação de dados à sua rede de dados e depois conecte os nós de computação de dados aos switches do cluster.

5

"Ligue seus nós de computação de dados"

Após instalar o hardware do rack e conectar os nós de computação de dados, você deve ligar os DCNs e os nós controladores do sistema de storage AFX, caso ainda não estejam ligados.

Requisitos de instalação para nós de computação de dados para AI Data Engine

Analise o equipamento necessário e as precauções de elevação para os nós de computação de dados do AI Data Engine.

Pré-requisitos

Antes de instalar os nós de computação de dados para AIDE, certifique-se de que você tem:

- Um sistema de storage AFX 1K.



Para obter informações sobre a instalação do sistema de storage AFX 1K, consulte ["Documentação de instalação do sistema de storage AFX 1K"](#).

Equipamentos necessários para instalação

Para instalar os nós de computação de dados para AIDE, você precisa dos seguintes equipamentos e ferramentas.

- Acesso a um navegador web para configurar seus nós de computação de dados
- Pulseira de descarga eletrostática (ESD)
- Lanterna

- Laptop ou console com uma conexão USB/serial
- Chave de fenda Phillips #2

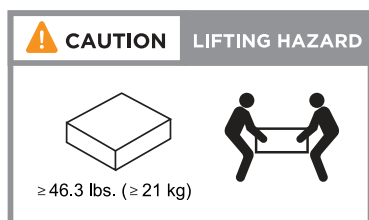
Precauções de levantamento

Os nós de computação de dados são pesados. Tenha cuidado ao levantar e mover esses itens.

Pesos dos nós de computação de dados

Tome as precauções necessárias ao mover ou levantar seu data compute node.

Um nó de computação de dados pode pesar até 46,3 libras (21 kg). Para levantar o nó de computação de dados, use duas pessoas ou um elevador hidráulico.



Informações relacionadas

- ["Informações de segurança e avisos regulamentares"](#)

Qual é o próximo passo?

Após analisar os requisitos de hardware, ["prepare-se para instalar seus nós de computação de dados"](#).

Prepare-se para instalar seus nós de computação de dados para AI Data Engine

Prepare-se para instalar seus nós de computação de dados para AI Data Engine, preparando o local, desembalando as caixas e comparando o conteúdo das caixas com a nota de remessa, e registrando o sistema para acessar os benefícios de suporte.

Etapa 1: preparar o local

Para instalar seus nós de computação de dados, certifique-se de que o local e o gabinete ou rack que você planeja usar atendam às especificações da sua configuração.

Passos

1. Use ["NetApp Hardware Universe"](#) para confirmar se o seu site atende aos requisitos ambientais e elétricos para os seus nós de computação de dados.
2. Certifique-se de que você possui espaço suficiente no gabinete para seus nós de computação de dados dentro da instalação existente do sistema de storage AFX 1K.
 - 1U para cada nó de computação de dados
 - 2U para cada nó controlador AFX 1K
 - 2U para cada shelf NX224
 - 1U ou 2U por switch, dependendo do modelo de switch.

Passo 2: desembale as caixas

Após garantir que o local e o gabinete ou rack que você planeja usar para seus nós de computação de dados atendam às especificações exigidas, desembale todas as caixas e compare o conteúdo com os itens da nota de entrega.

Passos

1. Abra cuidadosamente todas as caixas e disponha o conteúdo de forma organizada.
2. Compare os itens desembalados com a nota de embalagem e observe quaisquer discrepâncias.

Você pode obter sua lista de embalagem ao escanear o código QR na lateral da caixa de envio.

Os seguintes itens são alguns dos conteúdos que você pode ver nas caixas.

Hardware	Cabos	
• Painel • Kits de trilhos com instruções • Nó(s) de computação de dados	• Cabos de alimentação	

Etapa 3: registre seus nós de computação de dados

Após garantir que seu site atende aos requisitos das especificações do seu data compute node e verificar se você possui todas as partes que encomendou, você deve registrar seu sistema.

Passos

1. Localize os números de série dos seus nós de computação de dados.

Você pode encontrar os números de série nos seguintes locais:

- Na nota de embalagem
- No seu e-mail de confirmação
- Em cada nó de computação de dados ou, em alguns sistemas, no módulo de gerenciamento do sistema de cada nó de computação de dados.

SSN: XXXXXXXXXXXXX



2. Navegue até o ["NetApp site de suporte"](#).
3. Determine se você precisa registrar seu sistema de storage:

Se você é um(a)...	Siga estes passos...
Cliente existente NetApp	- Sign in com seu nome de usuário e senha. - Selecione Systems > My Systems. - Confirme se o novo número de série está listado. - Se o número de série não estiver listado, siga as instruções para novos NetApp customers.
Novo NetApp cliente	- Clique em Register Now para criar uma conta. - Selecione Systems > Register Systems. - Insira o número de série do sistema de storage e os detalhes solicitados. Depois que a NetApp aprovar seu cadastro, você poderá baixar o software necessário. A aprovação leva até 24 horas.

Qual é o próximo passo?

Depois de se preparar para instalar seus nós de computação de dados, ["instale os Data Compute Node"](#).

Instale seus nós de computação de dados para AI Data Engine

Instale e proteja seus nós de data compute no gabinete.

Antes de começar

- Certifique-se de que você tenha as instruções embaladas com o kit de trilhos.
- Compreenda as questões de segurança relacionadas ao peso do nó de computação de dados, sistema de storage e storage shelf.
- Entenda que o fluxo de ar através do sistema de storage entra pela frente, onde o painel ou as tampas laterais estão instaladas, e sai pela parte traseira, onde as portas estão localizadas.



Em geral, os switches devem ser instalados no centro do gabinete. As prateleiras de armazenamento devem ser instaladas abaixo do switch e acima de um segundo switch instalado. Os nós controladores podem ser instalados acima ou abaixo dos switches dentro do gabinete. Os nós de computação de dados podem ser instalados acima ou abaixo dos nós controladores dentro do gabinete.

Passos

1. Instale os kits de trilhos para seus data compute nodes, conforme necessário, usando as instruções incluídas com os kits.
2. Instale e proteja seus nós de computação de dados no gabinete:
 - a. Posicione o nó de computação de dados nos trilhos no meio do gabinete e, em seguida, apoie o appliance pela parte inferior e deslize-o para o lugar.
 - b. Prenda o nó de computação de dados ao gabinete usando os parafusos de montagem incluídos.
3. Fixe o painel na parte frontal do nó de computação de dados.

Qual é o próximo passo?

Depois de instalar os nós de computação de dados, ["Conecte os nós de computação de dados para AIDE"](#).

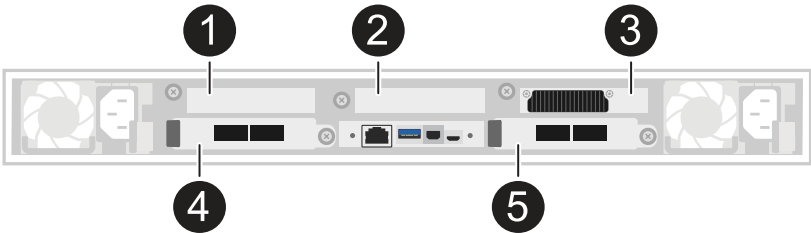
Requisitos para cabeamento de Data Compute Node para AI Data Engine

Os nós de Data Compute se integram ao seu sistema de storage AFX 1K por meio de conexões de rede do host e de rede do cluster. Revise a configuração dos slots de E/S, os tipos de cabos e os requisitos de conexão para sua implantação.

Configuração de cabeamento

Os nós de computação de dados se conectam aos mesmos switches de cluster que os nós controladores do AFX 1K, estendendo seu sistema de storage com recursos de computação otimizados para cargas de trabalho de AI e machine learning.

A configuração inicial do AI Data Engine (AIDE) suporta um mínimo de três nós de computação de dados. Para obter detalhes completos sobre a configuração e as prioridades dos slots, consulte ["NetApp Hardware Universe"](#).



1	Slot não utilizado no Data Compute Node.
2	Slot não utilizado no Data Compute Node.
3	Slot de GPU no Data Compute Node.
4	Slot de E/S no Data Compute Node.
5	Slot de E/S no Data Compute Node.

Configuração do slot de E/S

O Data Compute Node utiliza um esquema de numeração de slots específico que difere das configurações de servidor padrão. Compreender o layout dos slots é essencial para a correta instalação dos cabos.

- **Slot 3:** reservado para GPU (não acessível para cabos de E/S)
- **Slots 4 e 5:** slots de E/S usados para conexões de rede
 - Porta a: conexões de rede do cluster
 - Porta b: conexões de rede do host
- **Slots 1 e 2:** desocupados e inacessíveis para uso

Conexões de rede

Os nós de computação de dados requerem dois tipos de conexões de rede para se integrarem ao sistema de storage AFX 1K.

• Conexões de rede do host

As conexões de rede do host fornecem acesso aos dados do cliente e permitem que os nós de computação de dados processem as cargas de trabalho. Cada nó de computação de dados usa as portas e4b e e5b para conexões redundantes com switches de rede de host separados.

Atribuições de portas:

- e4b: conecta-se ao switch de rede A
- e5b: conecta-se ao switch de rede B

• Conexões de rede do cluster

As conexões de rede do cluster permitem a comunicação entre os nós Data Compute Node e os nós AFX 1K controller dentro do storage cluster. Cada Data Compute Node utiliza as portas e4a e e5a para conexões redundantes com switches de rede de cluster separados.

Atribuições de portas:

- e4a: conecta-se ao switch de rede A
- e5a: conecta-se ao switch de rede B

Componentes de hardware suportados

Os nós de computação de dados requerem cabos e switches específicos para garantir a conectividade e o desempenho adequados com o sistema de storage AFX 1K.

Data Compute Node	Switches compatíveis	Cabos compatíveis
Data compute nodes (mínimo de três necessários)	• Cisco Nexus 9332D-GX2B (400GbE) • Cisco Nexus 9364D-GX2A (400GbE)	• Cabos breakout 400GbE QSFP-DD para 4x100GbE QSFP56 para conexões com Data Compute Node: ◦ 100GbE para portas de rede do cluster Data Compute Node (e4a, e5a) ◦ 100GbE para portas de rede do host do Data Compute Node (e4b, e5b) • Cabos RJ-45 para conexões de gerenciamento



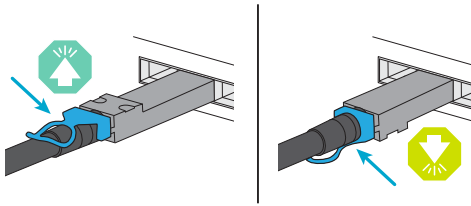
Os cabos breakout fornecem quatro conexões de 100GbE a partir de cada porta do switch de 400GbE. Conecte a extremidade de 400GbE aos switches e a extremidade de 100GbE às portas de E/S do Data Compute Node.

Orientação do cabo

Ao conectar cabos aos Data Compute Node, a orientação correta dos cabos garante conexões confiáveis.

Os diagramas de cabeamento nos procedimentos de instalação mostram ícones de seta indicando a

orientação correta (para cima ou para baixo) da aba de puxar do conector do cabo ao inserir um conector em uma porta. Ao inserir o conector, você deve sentir um clique ao encaixá-lo no lugar. Se você não sentir o clique, remova-o, vire-o e tente novamente.



Manuseie cuidadosamente os componentes delicados do conector ao encaixá-los no lugar.

Qual é o próximo passo?

Após revisar a configuração de cabeamento, ["conecte o hardware"](#) para seus Data Compute Nodes.

Conecte seus nós de computação de dados para AI Data Engine

Conecte seus nós de Data Compute Node à rede do host e aos switches de rede do cluster para habilitar o processamento de cargas de trabalho de IA e a integração com seu sistema de storage AFX 1K. Este procedimento utiliza conexões 100GbE tanto para acesso à rede do host quanto para comunicação do cluster, permitindo que os nós aproveitem a infraestrutura de cluster existente sem desligar o sistema AFX.

Sobre esta tarefa

Esses procedimentos mostram configurações comuns. O cabeamento específico depende dos componentes encomendados para o seu sistema de storage. Para obter detalhes completos sobre a configuração e as prioridades dos slots, consulte ["NetApp Hardware Universe"](#).



Não é necessário desligar o sistema de storage AFX 1K ao conectar os Data Compute Nodes. Você pode adicionar os Data Compute Nodes a um sistema de storage AFX 1K existente que já esteja ligado e configurado.

Antes de começar

- Você já possui um sistema de storage AFX 1K instalado. Para obter informações sobre a instalação do sistema de storage AFX 1K, consulte ["Documentação de instalação do sistema de storage AFX 1K"](#).
- Você já instalou e configurou os switches de rede necessários. Entre em contato com o administrador de rede para obter informações sobre como conectar o sistema aos switches de rede.
- Você revisou o ["Requisitos de cabeamento para os Data Compute Node"](#).



São necessários no mínimo três nós de computação de dados para implantar o AI Data Engine.

Etapa 1: conecte os Data Compute Node à rede do host

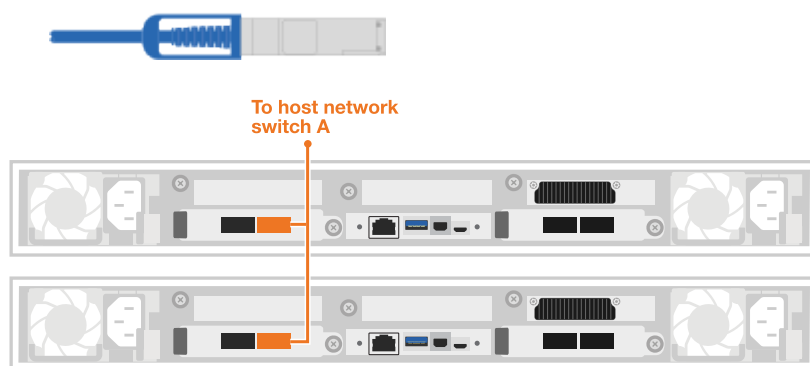
Você pode conectar as portas do Data Compute Node à sua rede host.

Passos

1. Conecte a porta e4b dos seguintes Data Compute Node ao switch de rede Ethernet A:
 - Data Compute Node 1, porta e4b

- Data Compute Node 2, porta e4b

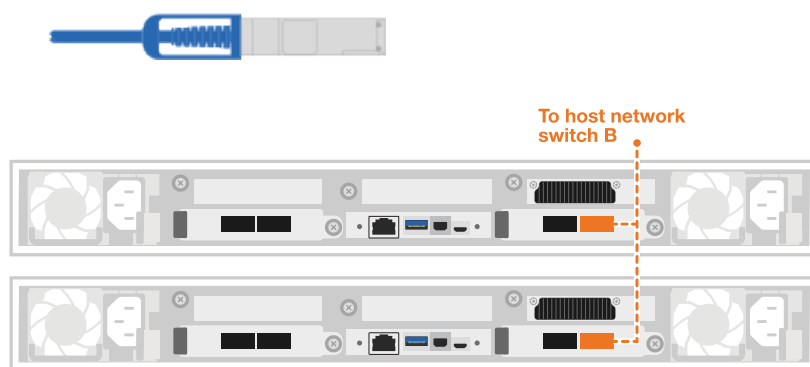
Cabos 100GbE



2. Conecte a porta e5b dos seguintes Data Compute Node ao switch de rede Ethernet B:

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

Cabos 100GbE



Etapa 2: conecte os cabos do cluster de Data Compute Node

Para os nós de computação de dados, use cabos breakout 4x100GbE para conectar as portas e4a/e5a para as conexões do cluster.

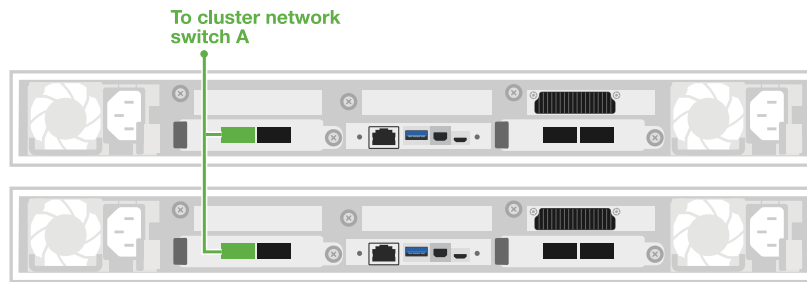
Passos

1. Conecte a porta e4a dos seguintes Data Compute Node a uma porta não ISL no switch de rede do cluster A:

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

4x100GbE cabos breakout

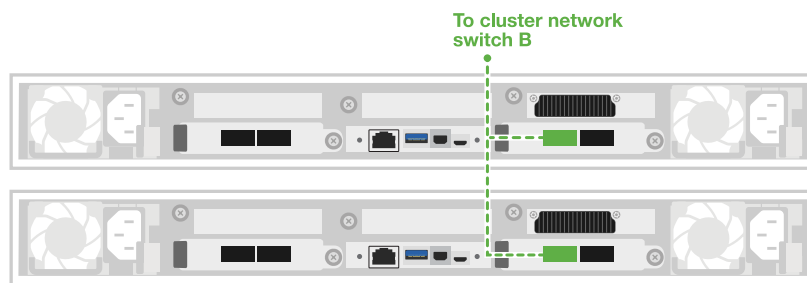




2. Conecte a porta e5a dos seguintes Data Compute Node a uma porta não ISL no switch de rede do cluster B:

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

4x100GbE cabos breakout



Qual é o próximo passo?

Depois de conectar os cabos do hardware, ["Ligue seus data compute nodes"](#).

Ligue seus nós de computação de dados para AI Data Engine

Após instalar o hardware do rack e conectar os nós de computação de dados, você deve ligar os DCNs e os nós controladores do sistema de storage AFX, caso ainda não estejam ligados.

Antes de começar

- Certifique-se de que suas prateleiras estejam ligadas e que cada uma tenha um ID exclusivo atribuído. Para obter informações sobre como atribuir IDs de prateleira para o sistema de storage AFX, consulte o ["documentação sobre a atribuição de IDs de shelf exclusivos"](#).

Passos

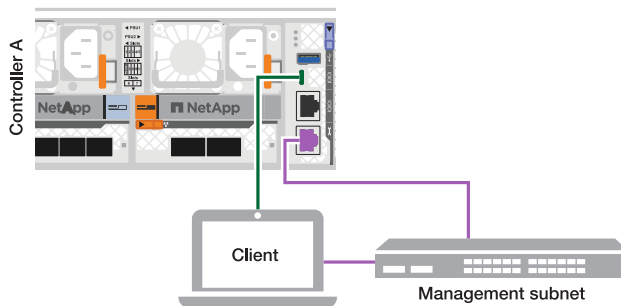
Após ligar as suas prateleiras de armazenamento e atribuir os IDs exclusivos, ligue os seus DCNs e ligue os nós controladores de armazenamento, caso ainda não estejam ligados.

1. Conecte seu laptop à porta de console serial. Isso permite que você monitore a sequência de inicialização quando os controladores são ligados.
 - a. Configure a porta de console serial do laptop para 115,200 baud com N-8-1.

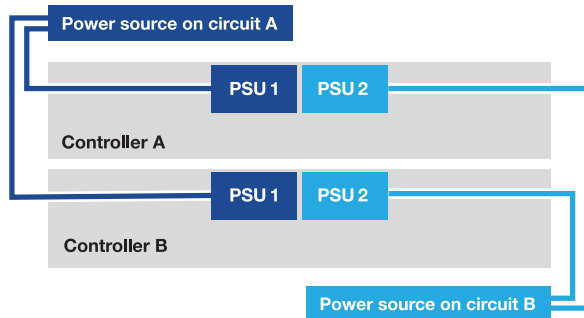
Consulte a ajuda online do seu laptop para obter instruções sobre como configurar a porta de console

serial.

- b. Conecte o cabo de console ao laptop e conecte a porta de console serial no controlador usando o cabo de console que veio com o seu sistema de storage.
- c. Conecte o laptop ao switch na sub-rede de gerenciamento.



2. Atribua um endereço TCP/IP ao laptop, usando um que esteja na sub-rede de gerenciamento.
3. Conecte os cabos de alimentação às fontes de alimentação do controlador e, em seguida, conecte-os a fontes de alimentação em circuitos diferentes.



- O sistema começa a inicializar. A inicialização inicial pode levar até oito minutos.
 - Os LEDs piscam e as ventoinhas ligam, o que indica que os controladores estão sendo ligados.
 - Os ventiladores podem fazer barulho ao iniciar, o que é normal.
4. Conecte os cabos de alimentação às fontes de alimentação do nó de computação de dados e, em seguida, conecte-os a fontes de alimentação em circuitos diferentes.
 5. Prenda os cabos de alimentação usando o dispositivo de fixação em cada fonte de alimentação.
 6. Ligue os nós de computação de dados.

Pode ser necessário remover o painel para acessar o botão liga/desliga; nesse caso, lembre-se de reinstalá-lo posteriormente.

Qual é o próximo passo?

Depois de ativar seus nós de computação de dados, ["Configurar um cluster ONTAP AIDE"](#).

Configure seu sistema AIDE

Configurar AI Data Engine

Como administrador de storage ONTAP, você pode integrar AI Data Engine (AIDE) à sua

implantação do sistema AFX. Você pode configurar AIDE como parte de uma implantação inicial do sistema de storage AFX ou adicionar AIDE a um cluster AFX que já foi implantado e está servindo dados. Em ambos os casos, você também precisará concluir a configuração do AIDE após finalizar a configuração básica.

1. Configure AIDE e integre-o a um sistema de storage AFX

Existem duas maneiras de configurar AIDE e integrá-lo a um sistema de storage AFX. Selecione a opção apropriada para o seu ambiente.



Antes de configurar AIDE, certifique-se de que os data compute nodes estejam instalados e conectados aos switches do cluster. Consulte ["Instale o AIDE"](#) para mais informações.

Configurar AIDE com um novo cluster AFX

Se você estiver configurando AIDE como parte da implantação inicial de um sistema AFX, o processo de configuração do AIDE faz parte do fluxo de trabalho padrão de configuração do cluster AFX. Durante a configuração do cluster ONTAP, o System Manager descobre automaticamente os DCNs conectados e os inclui como parte da configuração do cluster. Depois disso, você estará pronto para concluir a configuração do AIDE, incluindo a configuração de licença necessária.

Configurar AIDE com um cluster AFX existente

Se você estiver adicionando um cluster de Data Compute Node a um sistema de storage AFX existente, será necessário configurar os DCNs e finalizar a configuração do seu cluster AIDE.

Adicione os Data Compute Nodes ao cluster

ONTAP detecta dinamicamente os novos Data Compute Node conectados à sua rede de cluster e os exibe no System Manager. Você pode ["Adicione nós de computação de dados ao seu cluster AIDE"](#). Observe também o seguinte:

- Os nós de Data Compute devem estar fisicamente conectados aos switches do cluster. Exatamente três DCNs são necessários ao criar um AIDE cluster.
- Você precisa ter endereços IP suficientes para a rede DCN e outras informações de configuração de rede necessárias.

Finalize a configuração do seu cluster AIDE

Após integrar os nós de Data Compute e concluir a configuração inicial da rede, você pode finalizar a configuração do seu novo cluster AIDE.

Sobre esta tarefa

Uma janela de boas-vindas com orientações de configuração é exibida no lado direito da página com os passos necessários para a configuração. Uma marca de seleção é exibida à esquerda de cada passo ou item de ação que já foi concluído. Uma seta (→) é exibida para os itens que ainda não foram concluídos. Para cada uma das ações incompletas:

- Selecione o link do título ativado e conclua a ação de configuração para o seu ambiente.
- Se o link estiver desativado, passe o cursor sobre o título para exibir a ação necessária para ativá-lo e continue com a ação de configuração.

Cada etapa possui um link para documentação adicional, quando possível. Parte da configuração pode estar incluída como parte da configuração do cluster AFX. Consulte o ["Documentação da AFX"](#) para mais informações.

Passos

1. Selecione [Configurar agregação de links e VLANs](#).

Caso seja necessário para o seu ambiente, você pode configurar LAGs e VLANs em nível de cluster. Selecionar esta opção abrirá a página de portas Ethernet.

2. Selecione [Configurar protocolos de rede](#).

Você pode configurar os protocolos de rede da storage VM (SVM).

3. Selecione [Update data compute software](#).

Se uma imagem de software de Data Compute estiver disponível para atualização, esta opção será habilitada. Selecione-a para abrir a página de atualização do software de Data Compute. Para obter mais informações, consulte ["Atualizar o AI Data Engine software"](#).

4. Selecione [Configurar rede do mecanismo de dados](#).

Esta opção exibe a página para configurar o endereço de rede do AI Data Engine.

5. Selecione [Criar interfaces de rede entre clusters](#)

Esta opção abre a página para configurar as interfaces de rede intercluster como parte de um procedimento de duas etapas, caso você pretenda catalogar dados remotos.

6. Selecione [Interagir com outros sistemas de storage ONTAP](#).

Após criar as interfaces de rede entre clusters, você pode estabelecer relações de peering com outros sistemas de storage ONTAP.

7. Selecione [Adicionar um contêiner de dados](#).

Você pode adicionar um volume para uso pelo AIDE e associá-lo a uma SVM específica.

8. Selecione [Add Workspaces](#).

Isso inicia a página de espaços de trabalho do Data Engine, onde você pode ["criar um workspace"](#).

9. Selecione [Configurar OpenID Connect](#).

Você precisará ["configurar autenticação OpenID"](#) habilitar o acesso ao AI Data Engine Console.

Depois que você terminar

Após a conclusão de todas as tarefas, a janela de orientação fica oculta. Você pode fechá-la e reabri-la manualmente conforme necessário.

2. Concluir configuração do AIDE

Após o seu AIDE cluster estar configurado e integrado ao sistema de storage AFX, você precisa concluir a configuração do AIDE.

Passos

1. "Adicionar as licenças AIDE necessárias".

Há uma licença do AI Data Engine que você deve instalar para a funcionalidade completa do AIDE, incluindo vetorização e recursos de proteção.

2. "Configurar autenticação OIDC".

Certifique-se de configurar o OIDC para todas as implementações. Você deve configurar o OIDC para acessar o AI Data Engine Console.

Instale uma licença AIDE em seu sistema AFX

Você precisará instalar uma licença NetApp para acessar todos os recursos do AI Data Engine (AIDE). Como administrador de cluster ONTAP, você pode gerenciar as licenças usando ONTAP System Manager.

Prepare-se para licenciar o AI Data Engine

Uma licença é um registro de um ou mais direitos de uso de software. Todas as licenças AFX são fornecidas como um NetApp license file (NLF), que é um único arquivo que habilita vários recursos. Há vários aspectos que você deve considerar antes de instalar uma licença em seu sistema de storage para dar suporte ao AIDE.

Tipos de licenças

Existem dois tipos principais de licenças necessárias para começar a usar AIDE através do seu sistema de storage AFX.

Licença ONTAP One

A licença Metadata Engine Basic geralmente vem instalada de fábrica como parte da licença ONTAP One. Ela permite o acesso à funcionalidade do Metadata Engine, que fornece a base essencial para as operações do AIDE. Todas as funcionalidades principais necessárias para administrar seu sistema ONTAP também estão incluídas.

AI Data Engine

Você precisa adquirir e instalar a licença do AI Data Engine para acessar os serviços premium necessários para ativar todos os recursos do AIDE. A licença desbloqueia seus nós de computação de dados, habilitando recursos de IA como vetorização, mecanismos de governança, inferência e uma experiência de interface de usuário integrada. A licença inclui uma quantidade de GPUs e uma data de expiração.

Requisitos de instalação da licença

Você precisará adquirir a licença do AIDE e baixar o arquivo NLF correspondente para o seu sistema local. Em seguida, você poderá carregar o arquivo para o seu sistema de storage através do System Manager. Certifique-se também de ter o seguinte:

- Credenciais de administrador para fazer login no ONTAP System Manager
- Um cluster AFX executando ONTAP 9.18.1 ou posterior

Instale uma licença no seu sistema AFX

Você pode instalar uma licença AIDE para ativar os recursos adicionais do AIDE necessários para o seu

sistema de storage AFX.

Passos

1. No System Manager, selecione **Cluster** e depois **Settings**.
2. Ao lado de **Licenses**, selecione ➔.
3. Selecione a guia **Features** para exibir os recursos ONTAP disponíveis.
4. Para instalar uma licença, selecione a guia **Licenças instaladas**.
5. Selecione **+ Add**.
6. Selecione um arquivo de licença local e selecione **Adicionar**.

Informações relacionadas

- ["Visão geral do licenciamento ONTAP"](#)
- ["Como baixar licenças NLF do site de suporte da NetApp"](#)
- ["ONTAP CLI: adicionar licença do sistema"](#)

Configurar OpenID Connect para AIDE no ONTAP

Como administrador de cluster ONTAP, você pode usar ONTAP System Manager para configurar a autenticação OpenID Connect (OIDC) para um cluster AI Data Engine (AIDE). Isso proporciona um login seguro e centralizado por meio de um provedor de identidade (IdP) externo.



Você deve configurar o OIDC para acessar o AI Data Engine Console. Quando configurado, toda autenticação passa pelo OIDC. Se o OIDC não estiver configurado, o console ficará indisponível para administradores, engenheiros de dados e cientistas de dados. Nesse caso, o login no System Manager volta a utilizar autenticação local.

Observe também o seguinte sobre a configuração OIDC para acesso ao AIDE:

- Não é possível modificar uma configuração OIDC existente. Se precisar fazer uma alteração, primeiro exclua a configuração e crie uma nova com as configurações desejadas.
- Se você desativar ou remover o OIDC, System Manager voltará para a autenticação local do ONTAP.

Visão geral do OIDC

OpenID Connect (OIDC) é um protocolo de autenticação construído sobre a estrutura OAuth 2.0. Ele estende o OAuth 2.0, usado principalmente para autorização, adicionando uma camada de identidade. OIDC introduz o conceito de um token de ID, que é um JSON Web Token (JWT) contendo declarações sobre o evento de autenticação e a identidade do usuário.

Você precisa selecionar e configurar um provedor de identidade (IdP) externo compatível com AFX e AIDE. O IdP autentica os usuários e emite tokens que o AFX, por meio do System Manager, pode usar para conceder acesso ao AIDE Console.

Configurar provedores de identidade de terceiros

Para autenticar usando OIDC, primeiro você precisa configurar um IdP externo. A implementação do OIDC no ONTAP usa declarações de função nos tokens para impor o RBAC. Ao configurar um IdP, certifique-se de que

ele esteja configurado para retornar declarações de função no token de id e no token de acesso. ONTAP oferece suporte a dois IdPs para autenticação OIDC: Entra ID e Active Directory Federation Services (AD FS).

Entra ID

Você pode configurar o Entra ID usando as seguintes etapas de alto nível:

1. Crie um novo registro de aplicativo na página de configuração do Entra ID.
2. Defina o valor do URI de redirecionamento (Web) para `https://$CLUSTER_MGMT_IP/oidc/callback`, substituindo o endereço IP de gerenciamento do cluster ou FQDN apropriado.
3. Crie as funções necessárias em App Roles e atribua-as aos seus usuários.
4. Atualize as declarações do token em Token Configuration para retornar as funções em id-token e access-token.

Veja ["Configure um provedor OpenID Connect com Entra ID"](#) para mais informações.

Active Directory Federation Services

Você pode configurar o AD FS usando as seguintes etapas de alto nível:

1. Crie um novo Grupo de Aplicação e selecione **Aplicativo de servidor acessando uma API da web**.
2. Defina o valor do URI de redirecionamento (Web) para `https://$CLUSTER_MGMT_IP/oidc/callback`, substituindo o endereço IP de gerenciamento do cluster ou FQDN apropriado.
3. Configure as declarações para retornar funções em tokens.

Veja ["Adicione AD FS como um provedor de identidade OpenID Connect"](#) para mais informações.

Configurar OIDC no System Manager

Após configurar seu IdP, você pode configurar autenticação OIDC no System Manager para habilitar o acesso seguro ao AIDE Console.

Antes de começar

- Você precisa ter acesso de administrador ao System Manager.
- Seu provedor de identidade OIDC deve estar configurado e acessível.

Passos

1. No System Manager, selecione **Cluster** e depois **Settings**; localize o cartão OpenID Connect.
2. Se o OIDC já estiver configurado, você pode editar ou desativar a configuração. Se o OIDC não estiver configurado, selecione  para iniciar o processo de configuração.
3. Em Configurar OpenID Connect, forneça valores para os seguintes campos:
 - Fornecedor
 - Emissor
 - URI do conjunto de chaves web JSON
 - Endpoint de autorização
 - Endpoint de token

- Encerrar sessão endpoint
 - Emissor do Token de Acesso (opcional)
4. Em Configuração do cliente, forneça valores para os seguintes campos:
 - ID do cliente
 - Reivindicação de usuário remoto
 - Intervalo de atualização
 5. Em detalhes da conexão, forneça valores para os seguintes campos:
 - Endereço IP do cluster ou FQDN
 - Proxy de saída (opcional)
 6. Em Mapeamento de função externa, selecione um mapeamento de função existente ou defina uma nova função para o ONTAP `admin` usuário.
 7. Selecione **Ativar agora** e depois **Salvar**. System Manager será atualizado para aplicar as novas configurações de autenticação.
 8. Faça login com suas credenciais do IdP; após autenticação bem-sucedida, você retornará ao System Manager.

Informações relacionadas

- ["OpenID Foundation"](#)
- ["Implementação do OAuth 2.0 no ONTAP"](#)

Configurar espaços de trabalho

Preparar dados para AI Data Engine

Após criar um cluster, estabeleça um contêiner de dados que contenha os dados que você pretende usar com AI Data Engine (AIDE). Esse contêiner de dados deve ser um volume ONTAP, seja um volume local ou um volume de um cluster ONTAP emparelhado executando ONTAP 9.

Em vez de carregar manualmente os dados do cluster ONTAP para o AIDE, você precisará emparelhar os clusters e SVMs de destino com o cluster AIDE e então decidir quais volumes NFS deseja usar com o catálogo de metadados do AIDE. Depois de criar o contêiner de dados, você pode criar um workspace e associar o contêiner de dados a esse workspace. Os usuários desse workspace poderão acessar e interagir com as coleções de dados e recursos afiliados ao workspace para suas cargas de trabalho de IA.

Sobre esta tarefa

Você deve estabelecer uma conexão entre cada SVM que contenha os dados que deseja usar com o AI Data Engine. Estabelecer apenas a conexão do cluster não é suficiente. Isso garante que o AI Data Engine possa acessar, integrar e indexar seus dados conforme o esperado.

Você deve fazer o peering da SVM no AIDE cluster que atuará como o SnapMirror destino.

Você não precisa criar relações de SnapMirror entre os clusters ONTAP e o cluster AIDE. Essas relações serão criadas automaticamente durante a criação do espaço de trabalho.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para emparelhar clusters e SVMs e selecionar contêineres de dados.
- Você identificou os clusters ONTAP e as SVMs que contêm os dados que deseja usar com AI Data Engine.
- Você confirmou que os volumes da fonte de dados atendem aos seguintes requisitos:
 - Volumes estão online e acessíveis.
 - O protocolo NFS está ativado. Somente volumes com NFS habilitado são suportados como contêineres de dados para AI Data Engine. Volumes SMB e CIFS não são suportados.
 - Volumes não são volumes FlexCache.
 - As fontes de dados são volumes de leitura e gravação. Volumes de proteção de dados não são suportados.

Passos

1. ["Interligue cada ONTAP cluster e SVM"](#) que contém os dados que você deseja usar com AI Data Engine.
2. ["Selecione os volumes que deseja usar com AI Data Engine"](#).

Para cada volume, observe as seguintes informações:

- Nome do volume
- UUID
- Nome da SVM e UUID

- Nome do cluster e UUID

Qual é o próximo passo?

"[Criar um espaço de trabalho](#)" e associe os contêineres de dados que você criou a um espaço de trabalho.

Informações relacionadas

- "[Opções de migração de dados](#)"

Crie um workspace no AI Data Engine

Após configurar um cluster, você pode criar um espaço de trabalho. Os espaços de trabalho permitem segmentar dados no cluster, controlar o acesso aos dados para usuários individuais e excluir dados que AI Data Engine (AIDE) não deve acessar.

Se você administra storage, usará ONTAP System Manager para criar e gerenciar workspaces.

As organizações criam espaços de trabalho com base em equipes, projetos, níveis de sensibilidade de dados ou outros critérios relevantes. Por exemplo, se você trabalha em saúde, pode segmentar dados clínicos em um espaço de trabalho, mas deixar de fora dados referentes à TI, jurídico ou outros departamentos.

Sobre esta tarefa

Os limites de processamento do sistema afetam a criação de espaços de trabalho (normalmente até 15 GB por dia por cluster). Se você criar vários espaços de trabalho em paralelo ou em rápida sucessão, cada espaço de trabalho poderá demorar mais para ser processado e você poderá experimentar atrasos significativos.

Acompanhe o status da criação de espaços de trabalho na página de inventário de Espaços de Trabalho. Para obter melhores resultados, evite criar muitos espaços de trabalho de uma vez se precisar de acesso imediato a esses recursos.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para criar workspaces e associar coletas de dados.
- Você já definiu as fontes de dados remotas (com peering) e locais que pretende usar com o workspace e com AI Data Engine.
- Você "[criou pelo menos um contêiner de dados](#)" que o espaço de trabalho pode usar, como um volume local ou um volume de um cluster emparelhado.



Adicione um volume a um espaço de trabalho que você não excluirá durante o tempo de vida esperado desse espaço de trabalho. Se você excluir um volume depois de adicioná-lo a um espaço de trabalho, o espaço de trabalho entrará em estado de falha. Confirme a viabilidade a longo prazo do volume antes de estabelecer um espaço de trabalho.

- Certifique-se de que NFS esteja habilitado no volume, mas que CIFS não esteja habilitado. Os workspaces suportam apenas volumes com NFS. Volumes com CIFS (SMB) não são suportados.

Criar um espaço de trabalho

Crie um espaço de trabalho e associe contêineres de dados que contenham os dados que você deseja usar com AI Data Engine.

Passos

1. No ONTAP System Manager, navegue até **Data Engine > Workspaces**.
2. Selecione **Add**.
3. Na caixa de diálogo **Adicionar Espaço de Trabalho**, selecione pelo menos um contêiner de dados disponível para associar ao espaço de trabalho.
4. Configurar "**clusters pareados**" para que os dados desses clusters possam ser acessados dentro do espaço de trabalho
5. Se você quiser configurar o acesso do usuário ao espaço de trabalho, pode fazer isso agora ou "**Aguarde até que o espaço de trabalho seja criado**".
6. Configurar um intervalo de atualização para a frequência com que o workspace sincroniza com os contêineres de dados associados para capturar dados novos ou atualizados (por exemplo, seis horas).



Escolha um intervalo que equilibre a atualização dos dados com o desempenho do sistema. Se você adicionar um contêiner de dados a vários espaços de trabalho, o sistema usará automaticamente o intervalo mais agressivo (mais curto). Para saber mais, consulte a documentação sobre [Atualizações e versionamento de workspace](#).

7. Selecione **Continue**.
8. Na caixa de diálogo **Finalizar espaço de trabalho**, insira um nome e uma descrição para o espaço de trabalho.
9. Selecione **Add** para criar o espaço de trabalho.

Resultado

O processo de criação do espaço de trabalho leva de vários minutos a horas para ser concluído, dependendo do conjunto de dados associado e de sua contagem de arquivos, tamanho de arquivos e outros fatores.

O sistema extrai automaticamente metadados de todas as fontes de dados e os armazena em um catálogo de metadados que os usuários podem usar para localizar os arquivos necessários para seus projetos. Depois de atribuir usuários ao espaço de trabalho, os engenheiros de dados podem configurar e interagir com os componentes associados ao espaço de trabalho a partir do AI Data Engine Console.

O novo espaço de trabalho aparece na página Espaços de trabalho em `Creating` estado até que o processo seja concluído e o estado mude para `ready`.

Revise os detalhes do espaço de trabalho

Após a criação do espaço de trabalho, revise os detalhes do espaço de trabalho.

Passos

1. Analise os detalhes do espaço de trabalho, incluindo o tamanho total, porcentagem da capacidade do cluster utilizada e a data da atualização mais recente do espaço de trabalho.
2. Selecione o nome do workspace para abrir a página de detalhes.
3. Na guia Visão geral, visualize os detalhes do espaço de trabalho que incluem os contêineres de dados, usuários e atividades associados.

Atualizações e versionamento do espaço de trabalho

Cada atualização do espaço de trabalho cria uma versão imutável que captura o estado atual de todos os arquivos e objetos no espaço de trabalho. As versões incluem metadados completos, referências aos

snapshots usados durante a extração e um ID de tarefa para rastreabilidade. Isso oferece suporte à linhagem de dados, reprodutibilidade e auditoria.

As atualizações ocorrem de acordo com a programação que você configurar (como a cada seis horas) ou quando você as aciona manualmente. O intervalo mínimo de atualização suportado é de uma hora; o máximo é de um ano. Se um contêiner de dados estiver incluído em vários espaços de trabalho, o sistema usa o intervalo de atualização mais frequente e de menor duração para agendar a extração de metadados.

Por padrão, o sistema mantém as versões anterior, atual e seguinte (em andamento). O sistema mantém versões mais antigas de acordo com a política da sua organização e pode excluí-las conforme necessário.

Você pode listar todas as versões de um espaço de trabalho e visualizar as diferenças entre versões para identificar quais arquivos ou objetos foram adicionados, modificados ou excluídos. Isso permite acompanhar as alterações ao longo do tempo e compreender a evolução dos dados do seu espaço de trabalho.

Atribuir acesso de usuário aos workspaces do AI Data Engine

Como administrador de storage, você atribui usuários a um espaço de trabalho com base em suas funções: engenheiros de dados, cientistas de dados ou outras funções, dependendo da estrutura e das necessidades da sua organização. Os usuários fazem login no AI Data Engine (AIDE) Console usando suas credenciais e acessam os recursos do contêiner de dados dentro do espaço de trabalho atribuído.

ONTAP System Manager permite gerenciar quais usuários têm acesso aos workspaces do AI Data Engine. Você pode adicionar ou remover usuários para controlar quem pode visualizar, modificar e interagir com os dados e atividades do workspace.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para gerenciar o acesso de usuários aos workspaces.
- Confirme se você criou o espaço de trabalho e se ele está ativo no inventário de Workspaces.
- Confirme que todos os contêineres de dados pertinentes foram adicionados ao espaço de trabalho e estão acessíveis.
- "[Confirme se o OIDC está habilitado e configurado para o cluster](#)". O mapeamento de funções do IdP para funções do ONTAP deve ser concluído para cada usuário ou grupo relevante de engenheiro de dados e cientista de dados do IdP.

Passos

- Adicionar um usuário a um workspace:
 - a. No ONTAP System Manager, navegue até **Data Engine > Workspaces**.
 - b. Selecione o nome do workspace para abrir a página de detalhes.
 - c. Vá para a aba Usuários.
 - d. Selecione o botão **Adicionar** para abrir a caixa de diálogo de adicionar usuários.
 - e. Insira os dados de um ou mais usuários. Insira os dados como uma lista de usuários OIDC separados por vírgulas.
 - f. Selecione **Add** para conceder ao usuário acesso ao espaço de trabalho.

- Remova um usuário de um espaço de trabalho:
 - a. Na guia Usuários da página de detalhes do workspace, localize e selecione o usuário que deseja remover.
 - b. Selecione o botão **Remove**.
 - c. Confirme a remoção na caixa de diálogo.
 - d. O sistema remove imediatamente o usuário e revoga o acesso dele ao espaço de trabalho.

Resultado

Somente os usuários listados na guia Usuários do espaço de trabalho podem acessar e interagir com os dados e atividades do espaço de trabalho.

Administrar e monitorar

Monitorar processos de cluster

Visualize o status do sistema AIDE e do cluster

Como administrador de storage, você pode usar ONTAP System Manager para acessar o dashboard e exibir o status de cluster. Este é um bom primeiro passo antes de iniciar suas tarefas administrativas no AIDE ou se você suspeitar de um problema operacional.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para executar tarefas administrativas relacionadas ao AIDE ONTAP.

Monitore a integridade e a capacidade do AIDE a partir do painel de controle

1. Conecte-se ao ONTAP System Manager usando o endereço de gerenciamento do cluster:

```
https://$FQDN_OR_IP/
```

2. Sign in com uma conta de administrador.
3. Selecione **Dashboard** no painel de navegação à esquerda.
4. Revise o bloco **Health**:
 - Confirme a saúde geral do cluster.
 - Verifique a contagem e o status dos **Data compute nodes**.
 - Verifique se há alertas:
 - Problemas com nós da DCN ou problemas de conectividade
 - Espaços de trabalho ou coleta de dados com erro (por exemplo, falhas na publicação de coleta)
5. Analise o bloco **Capacity**:
 - Observe a capacidade total do cluster e a capacidade utilizada.
 - Para clusters AIDE, verifique:
 - Capacidade utilizada pelos metadados do AIDE e volumes de aplicativos (metadata Storage VM)
 - Capacidade utilizada pelos workspaces e coleta de dados (se disponível)
6. Opcionalmente, revise os blocos **Rede** e **Desempenho** para entender o comportamento em todo o cluster que pode afetar as cargas de trabalho do AIDE (por exemplo, congestionamento de rede ou atraso de proteção).

Visualizar dados DCN sobre a saúde e utilização

1. No painel de navegação, selecione **Cluster** e depois **Overview**.
2. Selecione a guia **Data compute**.

Esta aba mostra todos os nós DCN no cluster com:

- Nome do nó, modelo, número de série e versão do software

- Estado geral do nó
 - Utilização de CPU e memória
 - Utilização da GPU (se GPUs estiverem presentes)
 - Quaisquer indicadores de erro em nível de nó
3. Expanda um nó DCN para abrir a visualização detalhada e verifique:
- Uso de CPU e memória do sistema
 - Uso de memória da GPU
 - Problemas de hardware ou serviço relatados
4. Selecione **Cabeamento** na página **Cluster > Visão geral** para verificar se os nós do DCN estão corretamente cabeados aos switches do cluster e para identificar quaisquer problemas de porta ou link.

Monitorar espaços de trabalho e a pegada de metadados

1. No painel de navegação, selecione **Data engine** e, em seguida, **Workspaces**.
2. Revise o resumo do espaço de trabalho na parte superior da página:
 - Contagem de espaços de trabalho e seus estados (por exemplo, `Processing`, `Healthy`, `Error`).
 - Tamanho total do workspace.
 - Porcentagem da capacidade do cluster consumida por todos os workspaces.
3. Analise a grade do espaço de trabalho:
 - Confirme se os espaços de trabalho críticos apresentam um estado **Healthy**.
 - Verifique os tamanhos dos workspaces e o consumo de capacidade.
 - Procure por quaisquer espaços de trabalho em `Error` ou em estados de longa execução `Processing`.
4. Para visualizar os detalhes de um espaço de trabalho específico, selecione seu nome:
 - Na guia **Overview**, confirme:
 - Estado e tamanho do espaço de trabalho
 - Contêineres de dados (volumes) incluídos e suas respectivas contagens de itens
 - Última hora de atualização para cada fonte de dados
 - Na guia **Coleta de dados**, confirme:
 - Quais coleções de dados existem para esse espaço de trabalho (as coleções de dados são somente leitura no System Manager)
 - Seu estado, tamanho e hora da última atualização
 - Na aba **Usuários**, verifique quais usuários do AI Data Engine Console têm acesso.

Monitorar metadados Storage VM e proteção gerenciada pelo AIDE

1. No painel de navegação, selecione **Cluster** e depois **Storage VMs**.
2. Localize a Storage VM com o subtipo `data-engine` (a SVM de metadados):
 - Confirme se o SVM de metadados está online.
 - Opcionalmente, abra os detalhes para ver as contagens de:

- Volumes
- LIFs com tipo `Data compute network` (usado para comunicação DCN-ONTAP)

3. Selecione **Proteção** e depois **Relacionamentos** para visualizar a proteção das fontes de dados remotas usadas nos workspaces:

- Identifique os relacionamentos SnapMirror criados pelo AIDE por meio de padrões de nomenclatura:
 - Volume de destino: `<source_volume_name>_dest_<source_volume_UUID>`
 - Política: `<source_volume_name>_dest_aide_policy_<source_volume_UUID>`
- Use esta visualização para verificar se os relacionamentos estão saudáveis e se o tempo de atraso está alinhado com as expectativas de atualização do workspace.



Não modifique os metadados da Storage VM, os relacionamentos SnapMirror criados pelo AIDE ou os snapshots gerenciados pelo AIDE (ou seus agendamentos) diretamente no ONTAP. As alterações podem interromper o histórico de versões do AIDE. ["Ajuste as configurações de atualização do espaço de trabalho"](#) se precisar ajustar o comportamento de atualização.

Analise os alertas e notificações relacionados ao AIDE

1. No painel de navegação, selecione **Eventos & Jobs** e depois **System alerts**.
2. Analise quaisquer alertas ativos relacionados a:
 - Saúde ou conectividade do nó DCN
 - Problemas de rede do mecanismo de dados
 - Erros no espaço de trabalho ou na coleta de dados
 - Incompatibilidade de versões de software entre ONTAP e DCN cluster
3. Conforme necessário, configure os destinos de notificação (por exemplo, email, syslog) em **Cluster > Settings > Notification management** para garantir que os alertas relacionados ao AIDE sejam encaminhados para suas ferramentas operacionais.

Informações relacionadas

["Prepare-se para administrar seu sistema de storage AFX"](#)

Veja Insights para otimizar seu sistema AIDE

Como administrador de storage, você pode usar o recurso *Insights* do ONTAP System Manager para exibir atualizações de configuração sugeridas que estejam alinhadas com as melhores práticas da NetApp. Essas alterações podem otimizar a segurança e o desempenho do seu AIDE cluster.

Sobre esta tarefa

Cada insight é apresentado como um bloco ou cartão separado na página que você pode optar por implementar ou descartar. Você também pode selecionar o link da documentação associada para saber mais sobre uma tecnologia específica.

Passos

1. No System Manager, selecione **Analysis** e depois **Insights**.
2. Analise as recomendações disponíveis.

O que vem a seguir

Execute qualquer uma das ações recomendadas para implementar as melhores práticas de configuração.

Visualizar eventos do AIDE, tarefas e log de auditoria

Como administrador de storage, você pode revisar os eventos, trabalhos e mensagens de log de auditoria gerados pelo AIDE para acompanhar o processamento interno e diagnosticar possíveis problemas. O sistema AIDE pode ser configurado para encaminhar essas informações, juntamente com outros dados relacionados, para processamento e arquivamento adicionais.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para executar tarefas administrativas relacionadas ao AIDE ONTAP.

Monitorar as atividades, eventos e trabalhos do AIDE

Você pode usar a visualização centralizada **Activity** para monitorar eventos e tarefas específicos do AIDE em todos os espaços de trabalho ou revisar a atividade restrita a espaços de trabalho individuais.

Visualizar a atividade do AIDE em todo o cluster

Monitore as operações do workspace, solucione problemas de extração de metadados e acompanhe a publicação da coleta de dados em toda a sua implementação do AIDE.

1. No ONTAP System Manager, no painel de navegação selecione **Data engine** e depois **Activity**.
2. Selecione a guia **Eventos**:
 - Revise eventos recentes específicos da AIDE, como os seguintes:
 - Criação, atualização ou exclusão de workspace
 - Operações de adição/remoção de contêiner de dados
 - Publicação da coleta de dados (se presente)
 - Use filtros (por gravidade, tipo de objeto, espaço de trabalho ou intervalo de tempo) para se concentrar em eventos ativos ou críticos.
3. Selecione um evento individual para abrir e revisar:
 - Descrição e timestamp
 - Espaço de trabalho, coleta de dados ou fonte de dados afetada
 - Ação recomendada, se fornecida
4. Selecione a aba **Jobs**:
 - Monitore tarefas de longa duração, como:
 - Extração inicial de metadados para um espaço de trabalho
 - Tarefas de atualização do espaço de trabalho / atualização do catálogo
 - Publicação de coleta de dados ou trabalhos de atualização
 - Verifique o status e o progresso do trabalho.
5. Selecione uma tarefa para abrir a pré-visualização e revisar:

- Horário de início e término
- Percentual de progresso e fase (por exemplo, **Scanning**, **Publishing**)
- Espaço de trabalho, coleta de dados ou fonte de dados afetada
- Mensagens de erro para jobs com falha

Visualizar atividade específica do espaço de trabalho

Para solucionar problemas em um espaço de trabalho específico, abra os detalhes do espaço de trabalho (**Data Engine > Workspaces**, selecione um espaço de trabalho) e use a guia **Atividade**:

- Analise eventos e tarefas restritos apenas a esse workspace.
- Use esta visualização para isolar problemas, como um único espaço de trabalho travado em `Processing` estado.

Visualizar eventos em todo o cluster

Analise as mensagens de eventos para obter um registro valioso da atividade do sistema. Cada evento inclui uma descrição e um identificador exclusivo junto com uma ação recomendada.

1. No ONTAP System Manager, selecione **Events & jobs** e, em seguida, **Events**.
2. Analise e responda às ações recomendadas na parte superior da página, como ativar a atualização automática.
3. Selecione a aba **Events log** para exibir uma lista das mensagens.
4. Selecione uma mensagem de evento para examiná-la em mais detalhes, incluindo o número de sequência, descrição, evento e ação recomendada.
5. Opcionalmente, selecione a guia **Sugestões do Active IQ** e registre-se no Active IQ para obter informações detalhadas sobre os riscos do cluster.

Visualizar tarefas em todo o cluster

Veja todos os trabalhos em execução no cluster AIDE, incluindo trabalhos específicos do AIDE e trabalhos gerais do ONTAP.

1. No ONTAP System Manager, selecione **Events & Jobs** e depois **Jobs**.
2. Personalize a exibição, além de pesquisar e baixar as informações do trabalho conforme necessário.

Visualizar log de auditoria

Utilize o log de auditoria para revisar um registro da atividade do sistema com base no uso de protocolos de acesso como HTTP.

1. No ONTAP System Manager, selecione **Eventos e tarefas** e, em seguida, **log de auditoria**.
2. Selecione **Configurações** para ativar ou desativar as operações que são rastreadas.

Gerenciar notificações

Configurar os destinos de notificação para encaminhar automaticamente eventos AIDE e logs de auditoria.

Passos

1. No ONTAP System Manager, selecione **Cluster** e depois **Settings**.
2. Acesse **Gerenciamento de notificações** e selecione .
3. Selecione a ação apropriada para visualizar ou configurar os destinos usados pelo AIDE:
 - a. Destinos do evento: selecione **Ver destinos do evento**
 - b. Destinos do log de auditoria: selecione **Exibir destinos de auditoria**
4. Selecione **Adicionar** conforme apropriado e forneça as informações de destino.
5. Selecione **Save**.

Informações relacionadas

- ["Monitoramento de eventos, desempenho e saúde do ONTAP"](#)

Gerenciar workspaces do AI Data Engine

Um workspace é um conjunto de fontes de dados (volumes) que o AI Data Engine (AIDE) usa para construir e atualizar um catálogo de metadados para um projeto ou caso de uso. Como administrador de storage, você pode usar o ONTAP System Manager para monitorar a integridade do workspace, ajustar a configuração, controlar as fontes de dados, gerenciar usuários e excluir workspaces quando eles não forem mais necessários.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para gerenciar workspaces.

Revisar o status do workspace

Analise a integridade do workspace, a utilização da capacidade e o status dos metadados para garantir que o Metadata Engine esteja operando conforme o esperado e não esteja consumindo recursos inesperados.

Passos

1. No ONTAP System Manager, no painel de navegação selecione **Data engine > Workspaces**.
 2. Consulte o resumo na parte superior da página para total de espaços de trabalho, integridade geral dos espaços de trabalho e uso da capacidade.
 3. Para obter informações específicas do espaço de trabalho, selecione um nome de espaço de trabalho. Na guia **Visão geral**, confirme:
 - Estado e tamanho do espaço de trabalho.
 - Contêineres de dados (volumes) incluídos no workspace.
 - Contagem de itens e horário da última atualização por fonte de dados.
 - Quaisquer avisos em nível de workspace.
- 
- Se um workspace ou coleta de dados apresentar um estado de erro, verifique se todos os volumes de origem estão online e acessíveis.
4. Selecione a aba **Coleta de dados** para ver:
 - Todas as coleções de dados associadas a este workspace.

- Estado (como `Published` ou `Error`), tamanho e hora da última atualização.



System Manager é somente leitura para coleções de dados. Engenheiros de dados podem criar e gerenciar coleções de dados no AI Data Engine Console.

5. Selecione a aba **Usuários** para visualizar:

- A lista de usuários com acesso a este espaço de trabalho.

6. Selecione a aba **Atividade** para ver eventos e tarefas relacionados somente a este espaço de trabalho.

Editar propriedades do espaço de trabalho e agendamento de atualização

Você pode ajustar o nome, a descrição, o intervalo de atualização e (se licenciado) a política de guardrail de um espaço de trabalho.

Passos

1. Em **Data engine > Workspaces**, selecione  ao lado de um workspace e selecione **Edit**.
2. Editar propriedades do espaço de trabalho:
 - Atualize **Name** e **Description** conforme necessário.
 - Ajuste o **intervalo de atualização** (frequência de atualização de metadados) dentro do intervalo permitido (horas e dias).
 - Se uma licença AIDE estiver instalada, você pode selecionar **Guardrail policy**.
3. Selecione **Save**.



Alterações no intervalo de atualização ou no processamento de metadados podem afetar a frequência com que relações remotas do SnapMirror são atualizadas para este workspace.

Adicionar contêineres de dados a um espaço de trabalho existente

Você pode adicionar volumes montados adicionais (locais ou de um cluster remoto emparelhado) para que seus metadados sejam incluídos no catálogo do workspace.

1. Em **AI Data Engine > Espaços de trabalho**, faça um dos seguintes procedimentos:
 - Selecione  ao lado da área de trabalho e selecione **Adicionar contêineres de dados**.
 - Abra o workspace, selecione a guia **Overview** e, em seguida, selecione **Add** na seção de data containers.
2. Na caixa de diálogo **Adicionar contêineres de dados ao espaço de trabalho**:
 - Localize os volumes locais no cluster AIDE.
 - Expanda os clusters emparelhados para selecionar volumes remotos (volumes remotos exigem emparelhamento de cluster e SVM).



Somente volumes online elegíveis, que não estejam globalmente excluídos e que ainda não façam parte do workspace, podem ser selecionados.

3. Se você for solicitado a mapear um volume remoto:

- Selecione a máquina virtual de armazenamento de destino no cluster AIDE para receber destinos do SnapMirror para os volumes remotos escolhidos.

4. Selecione **Add**.
5. Utilize a guia **Atividade** do espaço de trabalho ou **Data engine > Atividade** para acompanhar a extração de metadados e qualquer inicialização do SnapMirror para novas fontes de dados.

Remova contêineres de dados de um espaço de trabalho

Você pode remover um contêiner de dados quando ele não for mais relevante para a finalidade do espaço de trabalho ou se desejar reduzir o escopo do gerenciamento de metadados para esse espaço de trabalho. Remover um contêiner de dados interrompe a atualização de metadados para esse volume e remove seus metadados do catálogo de metadados.



Não exclua um volume de origem do ONTAP que tenha sido adicionado a um espaço de trabalho. Se você excluir o volume, o espaço de trabalho entrará em estado de falha. Sempre remova o contêiner de dados do espaço de trabalho antes de excluir quaisquer volumes ONTAP subjacentes.

Passos

1. Navegue até **Data Engine > Workspaces** e selecione o workspace que contém o contêiner de dados.
2. Na guia **Visão geral**, localize o contêiner de dados a ser removido.
3. Selecione **Remove** ao lado do contêiner de dados.
4. Analise a caixa de diálogo de confirmação e selecione **Remove**.



Remover um contêiner de dados de um workspace não exclui o volume ONTAP subjacente nem sua relação do SnapMirror. Isso afeta apenas o uso de metadados dentro do AI Data Engine.

Gerenciar usuários do workspace

Você pode conceder ou revogar o acesso de usuários engenheiros de dados e cientistas de dados a um workspace. Esses usuários são definidos em seu provedor de identidade (OIDC) e mapeados para funções do ONTAP. Consulte "[atribuir usuários a workspaces](#)" a documentação para saber como gerenciar o acesso de usuários.

Excluir um workspace

Você pode excluir um espaço de trabalho para remover a definição do espaço de trabalho e os metadados do AIDE. Quaisquer coleções de dados e incorporações vetoriais vinculadas ao espaço de trabalho também serão removidas.



Os dados subjacentes do ONTAP (volumes, SnapMirror relationships) não são excluídos.

Passos

1. Em **Data Engine > Workspaces**, faça um dos seguintes procedimentos:
 - Exclua um único espaço de trabalho, selecione  e selecione **Excluir**.
 - Excluir vários espaços de trabalho, selecione as caixas de seleção dos espaços de trabalho e, em seguida, selecione **Excluir**.
2. Na caixa de diálogo de confirmação, revise os impactos da ação antes de prosseguir:

- Os metadados do workspace são excluídos permanentemente.
- As coleções de dados e as incorporações associadas ao espaço de trabalho são excluídas permanentemente.



Não existe opção de exclusão lógica ou de restauração.

3. Selecione a caixa de seleção para confirmar seu entendimento e selecione **Excluir**.

Informações relacionadas

- ["Atribuir usuários a espaços de trabalho"](#)

Atualize e faça a manutenção do seu sistema AIDE

Atualizações e compatibilidade do sistema AI Data Engine

Mantenha os componentes do sistema AI Data Engine (AIDE) atualizados para garantir o desempenho ideal e o acesso a novos recursos. Atualize os componentes após a implantação, quando novas versões de software ou firmware estiverem disponíveis, ao adicionar ou substituir nós ou periodicamente para atualizações de recursos.

Componentes do sistema AIDE

Existem dois componentes principais que requerem atualizações em um sistema AIDE: ONTAP software e AIDE software, que inclui DCN firmware.

Software ONTAP

ONTAP é o sistema operacional que roda nos sistemas de armazenamento NetApp, incluindo aqueles usados em implantações do AIDE. Mantenha o ONTAP atualizado para preservar a estabilidade, segurança e compatibilidade do sistema com os componentes do AIDE. Os componentes do AIDE são atualizados separadamente.

Atualizações do software AIDE

As atualizações de software do AI Data Engine Console e de firmware do DCN são distribuídas juntas como um único pacote (.tgz) e não estão incorporadas nas imagens do ONTAP. As atualizações garantem o funcionamento adequado dos componentes de hardware dentro do sistema AI Data Engine e fornecem novos recursos, melhorias de desempenho e correções de erros.

Entenda o processo de atualização

As atualizações do software AIDE podem ser gerenciadas através do ONTAP System Manager.

AIDE não suporta a funcionalidade de atualizações automáticas de software do ONTAP. Você pode se inscrever para receber notificações de ["Downloads do site de suporte da NetApp"](#), mas todas as atualizações do AI Data Engine software são realizadas manualmente pelo administrador.

Tipos e escopo de release:

- As versões principais do ONTAP (9.x.x) e as versões principais do AIDE (9.x.x U0) introduzem novos recursos, APIs ou alterações que afetam a integração do ONTAP.

- As versões de correção do ONTAP (9.x.x Px) e as versões de atualização do AIDE (9.x.x Ux) contêm correções e atualizações que não afetam a integração do ONTAP.

Matriz de compatibilidade

Ao planejar atualizações, assegure-se da compatibilidade entre ONTAP e AI Data Engine software.

O software AIDE é lançado como versões "U". Uma versão principal do AIDE é uma versão "U0" e as versões secundárias subsequentes serão "U1" e posteriores.

Compatibilidade entre ONTAP e AIDE

Versões do AIDE	Versões do ONTAP suportadas
9.18.1 U0	9.18.1 GA e todos 9.18.1 Px
9.18.1 U1 e posteriores	9.18.1 GA e todos 9.18.1 Px



"Px" indica todas as versões de correção (patches) do ONTAP na versão principal (por exemplo, 9.18.1 P1, 9.18.2 e outras).

Caminhos de atualização do AIDE

Os exemplos a seguir utilizam versões futuras hipotéticas para mostrar os caminhos de upgrade e atualização permitidos a partir do AIDE 9.18.1 U0 e 9.18.1 U1.

Se a sua versão atual do AIDE for...	E a sua versão alvo do AIDE é...	Seu caminho de upgrade ou atualização é...
9.18.1 U0	9.18.1 U1	Direto
9.18.1 U0	9.18.1 U3	Direto (você pode atualizar de qualquer Ux para qualquer Ux posterior na 9.18.1)
9.18.1 U1	9.18.1 U3	Direto (você pode atualizar de qualquer Ux para qualquer Ux posterior na 9.18.1)

Reverter limitações

Os sistemas AIDE não suportam operações de reversão para firmware DCN, atualizações de software AIDE ou ONTAP em sistemas de storage AFX. Após instalar uma atualização ou upgrade, você não pode reverter para uma versão anterior. Revise [requisitos de compatibilidade](#) antes de atualizar ou fazer upgrade.

Informações relacionadas

- ["Atualize o software ONTAP para sistemas AFX"](#)
- ["Atualizar o software AIDE"](#)

Atualizar o AI Data Engine software

Como administrador de storage, você pode atualizar o software do AI Data Engine (AIDE), o firmware do data compute node (DCN) e outros arquivos de sistema no seu sistema AIDE usando o ONTAP System Manager.

AIDE não suporta a funcionalidade de atualizações automáticas de software do ONTAP. Você pode se inscrever para receber notificações de ["Downloads do site de suporte da NetApp"](#), mas todas as atualizações do AI Data Engine software são realizadas manualmente pelo administrador.

Sobre esta tarefa

O pacote de software combinado do AI Data Engine é significativamente maior do que um pacote de atualização típico do ONTAP (o pacote AIDE tem aproximadamente 40GB). Planeje tempos de upload e instalação mais longos ao atualizar o software AIDE.

Antes de começar

- Você precisa de privilégios de *administrador de storage* para atualizar DCN firmware e AI Data Engine software.
- Você precisa das credenciais do site de suporte da NetApp para sua conta ativa.
- ["Assegure a compatibilidade entre ONTAP, firmware DCN e AI Data Engine software ao planejar atualizações"](#).



A funcionalidade de reversão não é compatível com DCN firmware ou AI Data Engine software. Após a instalação de uma atualização, não é possível reverter para uma versão anterior.

Passos

1. ["Baixe o arquivo combinado de firmware DCN e atualização de software AIDE para um cliente local"](#).
2. No System Manager, selecione **Cluster > Settings > Software updates**.
3. Ao lado de **Software updates**, selecione ➔.
4. Em atualizações do AI Data Engine, selecione **Adicionar arquivos de software do AI Data Engine** e selecione o pacote de atualização.
5. Após a conclusão do upload do pacote, selecione **Atualizar** para começar a instalar as atualizações nos nós do DCN.



AIDE Console não está disponível ou acessível quando a atualização do nó DCN está em andamento.

Resultado

As DCNs são atualizadas com AI Data Engine software e a versão atualizada é exibida para cada nó.

Informações relacionadas

- ["Atualize o software ONTAP para sistemas AFX"](#)

Adicione nós de computação de dados ao seu cluster AIDE

Você pode adicionar nós de computação de dados (DCNs) como parte da criação de um novo cluster AI Data Engine (AIDE) ou da expansão de um cluster existente. O fluxo de trabalho consiste em descobrir e configurar os nós usando o ONTAP System Manager.

Prepare-se para adicionar nós

Existem várias considerações ao adicionar nós DCN.

Ao criar um novo AIDE cluster

São necessários exatamente três nós DCN para um novo cluster AIDE.

Instalação e endereçamento de hardware

Certifique-se de que os seguintes pré-requisitos sejam atendidos:

- O novo DCN hardware está instalado no rack, ligado e conectado aos switches do cluster.
- Você tem um intervalo de endereços IP disponível para a subnet de backend do DCN para o ONTAP.
- O cluster ONTAP foi inicializado e está acessível a partir da LIF de gerenciamento do cluster.

Credenciais do System Manager

Você precisa de privilégios de *administrador de storage* para executar tarefas de criação ou expansão de cluster AIDE.

Compatibilidade de software

Consulte a seguinte documentação para confirmar se o hardware e as versões de software do seu DCN são compatíveis com o seu cluster ONTAP:

- ["O software DCN é compatível com a versão do ONTAP"](#).

Durante a operação de adição de nós, System Manager confirmará se os novos nós executam uma versão de software compatível com:

- A versão efetiva do ONTAP (ECV) do cluster, caso esta seja a primeira junção do DCN.
- A versão do cluster DCN existente, caso já existam DCNs presentes.

Se um nó for incompatível:

- Um erro é exibido ao lado do DCN afetado na caixa de diálogo **Adicionar**.
- Primeiro, você deve atualizar o DCN (ou ONTAP, conforme o caso) para uma versão compatível.

Adicione os nós de computação de dados

Você adiciona nós DCN como parte da criação de um novo cluster AIDE ou da expansão de um cluster existente.

Passos

1. No System Manager, selecione **Dashboard** no painel de navegação e, em seguida, o cartão **Health**.
2. Confirme se há nós para adicionar e selecione **View details** para exibir a lista.

A lista contém nós descobertos que ainda não fazem parte do AIDE cluster

3. Alternativamente, você pode selecionar **Cluster** e **Overview** e a guia **Data compute** para ver a lista.
4. Na parte inferior da página de computação de dados, selecione **Adicionar** acima da lista de nós.
5. Na caixa de diálogo **Adicionar nós de computação de dados**, selecione os nós DCN que deseja adicionar.

Você pode opcionalmente renomear nós individuais antes de adicioná-los.

6. Se esta for a primeira vez que você está adicionando nós e não houver uma sub-rede de backend, selecione **Adicionar sub-rede** e forneça:

- Nome da sub-rede (para uso interno)
- Endereço e máscara de sub-rede
- Intervalo de endereços IP para DCN e nós ONTAP nesta rede de back-end

System Manager valida se o intervalo inclui endereços IP livres suficientes para todos os DCNs que estão sendo adicionados, todos os nós ONTAP no cluster e IPs flutuantes adicionais no nível do cluster usados para comunicação entre o DCN e o ONTAP.

7. Se você adicionou a sub-rede de backend ou se ela já existe:

- Analise os endereços IP disponíveis.
- Se necessário, selecione **Editar sub-rede** e amplie o intervalo de IP.
 - Você só pode aumentar o intervalo. Reduzir ou alterar a sub-rede não é compatível.
 - Alterar a sub-rede ou o intervalo de IP pode exigir a recriação do cluster Kubernetes subjacente nos DCNs e pode levar vários minutos.

8. Opcionalmente, configure a interface do serviço Data Engine fornecendo:

- Endereço IP do serviço
- Máscara de rede
- Gateway (se necessário para o seu ambiente)

O endereço IP único será balanceado entre as DCNs e usado como endereço de frontend para o AI Data Engine Console e APIs relacionadas.

9. Analise os nós selecionados, a sub-rede de back-end e as configurações da interface de serviço do Data Engine.

10. Selecione **Add** e aguarde a conclusão da operação. System Manager executará as seguintes ações:

- Adicione os nós selecionados ao cluster DCN
- Provisione a rede de backend e adicione nós ao cluster DCN baseado em Kubernetes
- Atualiza os metadados internos para descoberta de DCN

11. Após a conclusão, selecione **Cluster** e **Overview** e confirme:

- Em **Computação de dados**, os novos DCNs aparecem como parte do cluster
- Todos os nós estão **saudáveis**
- Verifique se o cartão **Saúde** do dashboard mostra a contagem de nós atualizada

Substitua um nó no seu AIDE cluster

Você precisa substituir um nó de computação de dados (DCN) em seu AI Data Engine (AIDE) cluster se ele parar de funcionar ou precisar ser trocado devido a falha de hardware, upgrade ou manutenção. Isso garante que o cluster do AI Data Engine (AIDE) permaneça íntegro e operacional. O procedimento pode ser realizado sem interromper os serviços em andamento.

Prepare-se para substituir um nó

Há vários fatores a considerar antes de substituir um nó no seu AIDE cluster.

Credenciais do System Manager

Você precisa de privilégios de *administrador de storage* para executar tarefas de substituição de nó de cluster AIDE.

Restrições

Você deve estar ciente das seguintes restrições ao substituir um nó no AIDE cluster:

- A substituição de nó só é suportada usando a CLI e, opcionalmente, a API REST.
- Não é possível realizar a substituição de nó usando System Manager.
- O novo nó deve corresponder à versão do software do cluster; ONTAP atualizará se necessário.
- O nó com falha não deve ser ligado enquanto estiver conectado à rede do cluster para evitar conflitos de endereço IP.

Requisitos

Você precisará do seguinte:

- Número de série do novo nó de substituição

Substitua um nó DCN no seu cluster AIDE

Você pode substituir um nó DCN em seu AIDE cluster usando o seguinte procedimento.

Passos

1. Remova fisicamente o nó com defeito

Desligue e desconecte o nó da rede do cluster. Certifique-se de que o nó não esteja inicializado na rede durante o processo de substituição.

2. Exclua o nó com falha do cluster usando o seguinte comando:

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

Forneça o nome real para o valor <node_name>.

3. Conecte fisicamente o novo nó ao cluster

Certifique-se de que o nó esteja conectado por cabo, ligado e detectável.

4. Visualize os nós detectáveis e não configurados para verificar se o novo nó está online:

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

5. Adicione o nó ao cluster usando o seguinte comando:

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

ONTAP alocará um endereço IP para o novo nó. Se a versão do software do nó não corresponder à do cluster, ONTAP atualizará automaticamente o nó.

6. Verifique a integridade do cluster e a integração do nó com um dos seguintes comandos:

```
dcn cluster node show
```

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

Informações relacionadas

- ["Expanda seu cluster de computação"](#)

Gerenciar vetorização e coleções de dados

Início rápido Data-to-RAG para AI Data Engine

Vá de um sistema AI Data Engine (AIDE) recém-implantado para um endpoint de geração aumentada por recuperação (RAG) em funcionamento usando este fluxo de trabalho. Entenda como administradores de storage, engenheiros de dados e cientistas de dados colaboram usando ONTAP System Manager e AIDE Console.

Antes de começar

- Você instalou e adicionou nós de computação de dados (DCNs) ao cluster ONTAP.
- Você instalou e licenciou AI Data Engine software para vetorização e guardrails.
- Você configurou ["OpenID Connect \(OIDC\)"](#) e mapeou as funções de administrador, engenheiro de dados e cientista de dados.

1

Defina o escopo e a governança dos dados

Como administrador de storage ou administrador de segurança, você deseja preparar o ambiente no AIDE Console e ONTAP System Manager:

- ["Crie um ou mais espaços de trabalho"](#) de fontes de dados locais e remotas.
- ["Configure classificadores e políticas de proteção"](#) no AIDE Console.
- ["Atribua acesso de engenheiro de dados e cientista de dados aos espaços de trabalho"](#).

2

Explorar metadados do workspace

Como engenheiro de dados ou cientista de dados, você deseja explorar os metadados do workspace usando AIDE Console:

- ["Explorar metadados do workspace"](#) para compreender o conteúdo disponível.
- Defina um ou mais subconjuntos lógicos de dados que devem alimentar RAG (por exemplo, artigos de suporte, manuais de produtos ou notas clínicas anonimizadas).

3

Criar e publicar uma coleta de dados

Como engenheiro de dados ou cientista de dados, você deseja transformar o subconjunto escolhido em uma coleção pronta para RAG:

- ["Criar uma coleta de dados"](#) a partir do espaço de trabalho usando os filtros selecionados.
- ["Publicar a coleta de dados"](#) e monitorar a indexação até que ela atinja Ready estado.
- Copie o URI do endpoint de recuperação para a coleção escolhida e forneça aos cientistas de dados ou desenvolvedores de aplicativos.
- ["Visualizar o status da coleta de dados e a pegada vetorial"](#) conforme necessário.

Qual é o próximo passo?

- ["Defina seu ambiente de dados e políticas de proteção no AI Data Engine"](#)
- ["Explore metadados de workspace no AI Data Engine Console"](#)
- ["Crie coleções de dados no AI Data Engine Console"](#)

Explore metadados de workspace no AI Data Engine Console

Como engenheiro de dados ou cientista de dados, sua primeira tarefa no AI Data Engine (AIDE) é entender quais dados estão disponíveis em seus workspaces. Você usa o AI Data Engine Console para consultar o catálogo de metadados, buscar arquivos relevantes e identificar subconjuntos de dados para transformar em data collections.

Antes de começar

- Você precisa de privilégios de *data engineer* ou *data scientist* no AI Data Engine Console e acesso a pelo menos um espaço de trabalho.
- Um administrador de storage tem:
 - Criou um ou mais workspaces no ONTAP System Manager.
 - Atribuído acesso de usuário ou grupo aos espaços de trabalho relevantes.
- A extração de metadados para o espaço de trabalho foi concluída e o espaço de trabalho está em Ready estado.
- Os classificadores estão ativados, de forma que os metadados incluam etiquetas de classificação (por exemplo, indicadores de PII).

Sign in no AIDE Console como engenheiro de dados ou cientista de dados

Passos

1. Em um navegador, acesse a URL do AIDE Console:

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

2. Autentique-se através do provedor OIDC da sua organização.
3. Confirme se sua função é reconhecida como engenheiro de dados ou cientista de dados (por exemplo, pelo espaço de trabalho disponível e ações de coleta de dados). Para mais informações, veja ["AIDE documentação da função"](#) para saber como engenheiros de dados e cientistas de dados trabalham com os componentes do AIDE.

Resultado

Você está conectado ao AIDE Console e vê apenas os espaços de trabalho aos quais você tem acesso.

Veja seus espaços de trabalho acessíveis

Passos

1. No AIDE Console, navegue até **Data Curator > Workspaces**.
2. Revise a lista de espaços de trabalho que você pode acessar.
3. Selecione um workspace para abrir seus detalhes.

Resultado

Agora você tem uma visão com escopo no espaço de trabalho do conjunto de dados que os administradores de storage disponibilizaram para seus projetos.

Qual é o próximo passo?

- ["Crie coleções de dados para RAG a partir de um workspace"](#)

Crie coleções de dados no AI Data Engine Console

As coleções de dados são os blocos de construção RAG essenciais no AI Data Engine (AIDE). Como engenheiro de dados ou cientista de dados, você define quais arquivos pertencem a uma coleção, configura as opções de incorporação e indexação e publica a coleção para que os aplicativos possam consultá-la por meio de um endpoint de recuperação.

Você executará todas as tarefas de coleta de dados no AI Data Engine Console.

Antes de começar

- Você precisa de privilégios de *engenheiro de dados* ou *cientista de dados* no AI Data Engine Console (https://<cluster_management_ip>/console).
- Você tem acesso a pelo menos um espaço de trabalho com metadados extraídos e em Ready estado.
- Você explorou os metadados do workspace e identificou consultas ou filtros que definem subconjuntos de dados relevantes.
- A licença do AI Data Engine software está instalada e os recursos de inferência estão ativados.

Criar uma coleta de dados a partir de metadados do workspace

Passos

1. Acesse **Data Curator > Workspaces** e selecione o workspace que contém seus dados de destino.
2. Selecione **Add data collection**.
3. Na página **Criar nova coleta de dados**, faça o seguinte:
 - a. Insira um nome e uma descrição para a coleção (por exemplo, `Support_KB_RAG_EN`).
 - b. Escolha se a coleta deve ser:
 - **Dinâmico**: Novos arquivos são identificados e adicionados automaticamente à coleta de dados com base nos critérios de filtragem que você define. Isso ocorre durante as atualizações do espaço de trabalho.
 - **Estático**: Você escolhe quais arquivos são incluídos na coleção. Você pode editar os arquivos se a coleta de dados estiver no `draft` estado. Depois que a coleta de dados passa para `Published` estado, ela não pode ser editada.
4. Especifique o subconjunto de origem:
 - a. Utilize palavras-chave e filtros (tipo de arquivo, carimbos de data e hora e outros atributos) para encontrar os arquivos relevantes a serem incluídos.



Você pode selecionar um nome de arquivo para abrir uma janela de visualização prévia do conteúdo.

5. Adicione esses arquivos à coleta de dados.
6. Selecione **Save** para finalizar a coleta.

Resultado

Você definiu o escopo da coleta de dados e adicionou os arquivos necessários a ela. AIDE gera embeddings e constrói o índice vetorial quando você publica a coleção.



Crie coleções pequenas e focadas (por exemplo, por caso de uso ou domínio) em vez de uma única coleção "para tudo". Isso melhora a relevância da recuperação e a capacidade de gerenciamento.

Publicar uma coleta de dados

Publique a coleta de dados para torná-la consultável por aplicações de IA através de um endpoint de recuperação RAG. A publicação gera embeddings vetoriais a partir dos arquivos selecionados e os indexa para busca semântica. Após a coleção atingir *Ready* estado, seu endpoint fica disponível para que cientistas de dados a integrem em notebooks, pipelines e aplicações de IA para geração aumentada por recuperação (RAG) e busca.



Para coleções grandes, considere agendar a publicação inicial e as principais republicações em horários de menor movimento para minimizar a disputa por recursos.

Passos

1. Navegue até **Data Curator > Data collections** e selecione o menu de opções (⋮) para sua coleta de dados.
2. Selecione **Publish**.
3. Selecione uma configuração de otimização padrão ou personalizada.
4. Selecione **Publicar** para iniciar a transformação de dados.
5. No AIDE Console, abra a visualização de detalhes da coleta (**Data Curator > Coletas de dados**) para atualizações de status.

Resultado

A coleção chega ao estado *Ready* e fica disponível para uso por aplicações subsequentes e cientistas de dados.

Em **Data Curator > Data collections**, você pode selecionar **Copy URI** para obter as informações necessárias para acessar a coleta de dados usando uma API.

Atualizar ou excluir uma coleta de dados

Com o tempo, você pode precisar refinar ou desativar coleções de dados. Refinar uma coleção pode envolver ajustar filtros para adicionar ou remover arquivos, alterar configurações de incorporação ou atualizar a descrição da coleção. Excluir uma coleção a remove permanentemente e torna seu endpoint de recuperação indisponível.

Atualizar uma coleta de dados

Você pode atualizar uma coleta de dados quando ela estiver em `draft` estado.

Passos

1. Acesse **Data Curator > Data collections**.
2. Selecione a coleção que você deseja modificar.
3. Escolha **Editar**.
4. Ajuste qualquer um dos seguintes:
 - Nome e descrição
 - Filtros (caminhos, tipos de arquivo, classification tags).
 - Configurações de embedding e chunking.
5. Salve suas alterações.
6. Publique a coleção novamente para que a nova definição e os embeddings entrem em vigor.

Resultado

Uma nova tarefa de indexação é executada com a configuração atualizada e a coleção retorna a `Ready` um estado quando concluída.

Excluir uma coleção

A exclusão de uma coleção é permanente. Certifique-se de que nenhum aplicativo em produção ainda dependa do endpoint de recuperação da coleção antes de excluí-la.

Passos

1. Navegue até **Data Curator > Data collections** e selecione o menu de opções (**...**) para a coleção.
2. Escolha **Excluir**.
3. Confirme a exclusão.

Resultado

A definição da coleção e seus embeddings são removidos do AI Data Engine. Aplicativos que tentarem consultar o antigo endpoint de recuperação falharão após a remoção da coleção.

Qual é o próximo passo?

- ["Visualizar coletas de dados"](#)

Visualizar coleções de dados no AI Data Engine

Depois que engenheiros de dados ou cientistas de dados criam e publicam coleções de dados a partir de workspaces, você precisa de visibilidade sobre o status, tamanho e impacto dessas coleções no AI Data Engine cluster.

Se você for um administrador de storage, engenheiro de dados ou cientista de dados, poderá visualizar data collections no ONTAP System Manager e AIDE Console.

Antes de começar

- Você precisa de privilégios de *administrador de storage* no ONTAP System Manager ou de *engenheiro de dados* ou *cientista de dados* no AI Data Engine Console (https://<cluster_management_ip>/console para visualizar coleções de dados).
- Existe pelo menos um espaço de trabalho com metadados extraídos com sucesso.
- Engenheiros de dados ou cientistas de dados criaram e publicaram pelo menos uma data collection no AI Data Engine Console.
- A licença do AI Data Engine software está instalada e os recursos de inferência estão ativados, de modo que os endpoints de vetorização e recuperação estejam ativos.

Visualizar coleta de dados de todo o cluster

Para administradores de storage, ONTAP System Manager fornece uma visão de todo o cluster das coletas de dados e seu footprint, mas não permite que os administradores as criem ou modifiquem.

Passos

1. No System Manager, navegue até **Data Engine > data collections**.
2. Confira o resumo do inventário no topo da página:
 - Número total de coleta de dados por status
 - Espaço total consumido pelo banco de dados de vetores em todas as coleções
 - Espaço vetorial como porcentagem da capacidade total do cluster
3. Selecione uma coleta de dados e revise:
 - Nome e descrição da coleção
 - UUID
 - Espaço de trabalho associado
 - Status
 - Tamanho da coleção
 - Criador
 - Última atualização

Resultado

Agora você tem uma visão de alto nível de todas as coletas de dados no cluster e seu impacto no armazenamento. Use essa visão para identificar coletas que são grandes, desatualizadas ou que estão em um estado não pronto.

Você também pode verificar se uma coleta de dados está sendo atualizada ativamente e se alguma falha está bloqueando o uso do RAG.

Monitorar trabalhos e eventos relacionados à coleta

Como administrador de storage, você pode monitorar tarefas que criam e atualizam coleções na página **Activity** de todo o cluster e nos detalhes do espaço de trabalho.

Passos

1. No System Manager, navegue até **AI Data Engine > Activity**.
2. Na aba **Eventos**:

- a. Filtre por tipo (por exemplo, workspace, coleta de dados) ou gravidade.
 - b. Expanda qualquer evento relacionado à coleta de dados (por exemplo, "Data collection publish failed") para ver mais detalhes.
3. Na aba **Jobs**:
- a. Filtre para focar em coleta de dados, indexação e publicação.
 - b. Para cada tarefa, abra a pré-visualização para ver:
 - Porcentagem de progresso.
 - Horários de início e término.
 - Quaisquer mensagens de erro ou avisos relatados.
4. Opcionalmente, volte ao espaço de trabalho afetado (**Data Engine > Workspaces**) e abra a guia **Atividade** para ver os eventos e trabalhos restritos a esse espaço de trabalho.

Resultado

Você pode acompanhar o ciclo de vida de coletas de dados, identificar trabalhos paralisados ou com falha e reunir informações contextuais para repassar a data engineers, data scientists ou suporte.



Quando uma coleta de dados permanece em **Publishing** estado por um período prolongado, verifique se há uma tarefa de longa duração correspondente na página **Activity** antes de presumir uma falha.

Visualize as coleções de dados do AIDE Console

Engenheiros de dados e cientistas de dados normalmente monitoram coleta de dados diretamente do AIDE Console, onde são criadas e publicadas.

Passos

1. Faça login no AIDE Console como data engineer ou data scientist.
2. Navegue até **Coleções de Dados** e selecione a data collection desejada.
3. Para cada coleção:
 - a. Verifique o estado (**Draft**, **Publishing**, **Ready**, ou **Failed**).
 - b. Selecione o nome da coleta de dados para revisar os detalhes da definição (filtros, tipos de arquivo incluídos, opções de classificação, configurações de incorporação).
 - c. Verifique os registros de data e hora da última publicação ou atualização.
4. Se necessário, abra os detalhes ou logs da tarefa (quando disponíveis) para entender falhas ou execuções incompletas.

Resultado

Engenheiros de dados e cientistas de dados podem iterar nas definições de coleção e publicá-las novamente enquanto monitoram o status e a integridade, sem envolver o administrador de storage.

Qual é o próximo passo?

- ["Crie coleções de dados para RAG no AIDE Console"](#)

Implementar guardrails

Defina suas políticas de Data Guardrails no AI Data Engine para seu data estate

Como proprietário de dados ou plataforma, você usa o AI Data Engine Console para definir quais dados estão dentro do escopo para IA, quais dados são sempre proibidos e quais regras de segurança se aplicam quando esses dados são usados para classificação e geração aumentada por recuperação (RAG).

Utilize esses procedimentos para definir essas políticas no AIDE Console para que o ONTAP System Manager possa aplicá-las a todos os dados nos workspaces.

Antes de começar

- Você precisa de privilégios de *administrador de armazenamento* no AI Data Engine Console (https://<cluster_management_ip>/console) para criar e gerenciar políticas globais.
- Você possui um AIDE cluster com nós de computação de dados implantados e íntegros.
- "OpenID Connect (OIDC)" está configurado e sua função de IdP está mapeada para uma função de administrador do AIDE que permite o gerenciamento de políticas de dados.
- A licença do AI Data Engine software está instalada para que os recursos de Data Guardrails e inferência estejam ativados.
- Existe pelo menos um espaço de trabalho, ou você se coordenou com o administrador para entender quais fontes de dados (volumes) serão usadas nos espaços de trabalho.

Entenda os tipos de políticas

AIDE Console expõe estes tipos de políticas que moldam seu data estate:

- **Classificadores:** Habilite classificadores para detectar PII, problemas de segurança ou outros padrões em todos os workspaces.
- **Categorias de classificadores:** Agrupe os classificadores em categorias de conformidade para organização e gerenciamento.
- **Políticas de Data Guardrails:** Regras de segurança e redação aplicadas no momento da recuperação ou inferência.

Não é possível usar ONTAP System Manager para criar ou gerenciar essas políticas de Data Guardrails. Ele apenas as lê e as aplica quando um administrador de storage as define para workspaces. Toda a definição e manutenção de políticas ocorre no AIDE Console.

Ativar classificadores

Os classificadores analisam tanto metadados quanto conteúdo para anotar arquivos e objetos (por exemplo, detectando PII ou categorias sensíveis). Antes que os classificadores possam ser executados nos dados do espaço de trabalho, você deve habilitá-los no AIDE Console.

Sobre esta tarefa

O comportamento dos classificadores é controlado globalmente no AIDE Console. Todos os classificadores habilitados são executados em todos os workspaces. Como são aplicados globalmente, não podem ser

habilitados ou desabilitados para um workspace individual. Eles só podem ser habilitados ou desabilitados globalmente.

Passos

1. No AIDE Console, navegue até **Data Guardrails > Classifiers**.
2. Selecione uma categoria de classificador para revelar os classificadores que ela contém.
3. Selecione as caixas de seleção dos classificadores que deseja ativar ou selecione todas as linhas para ativar classificadores em massa.
4. Selecione **Enable**.



Use a opção de seleção em massa para ativar vários classificadores simultaneamente. Cada vez que você ativa um classificador, uma atualização do espaço de trabalho é acionada em todos os espaços de trabalho. Para minimizar atualizações desnecessárias, ative vários classificadores simultaneamente em vez de um de cada vez.

Resultado

Todos os espaços de trabalho recém-criados e existentes executam os classificadores ativados durante o processamento de metadados.

As etiquetas de classificação são gravadas no catálogo de metadados e ficam disponíveis para engenheiros de dados para filtragem durante a criação de coleções de dados.

Gerenciar categorias de classificação

Os classificadores são organizados em categorias (como "PII" ou "Financial data"). As categorias ajudam a agrupar classificadores relacionados para facilitar o gerenciamento e a visibilidade da conformidade. Você pode usar as categorias padrão que a AIDE fornece ou criar categorias personalizadas para atender aos seus requisitos de conformidade.

Passos

1. No AIDE Console, navegue até **Data Guardrails > Classifiers**.
2. Veja as categorias de classificação existentes. Há duas categorias principais de classificação:
 - Conteúdo ou dados: detecta tipos específicos de dados dentro dos arquivos.
 - Documento: classifica o tipo de documento com base no conteúdo.
3. Determine se as subcategorias do classificador padrão são suficientes ou se você deseja criar sua própria subcategoria.
 - Se você estiver usando uma subcategoria de classificador padrão (por exemplo, **General Privacy**):
 - i. Selecione o nome da categoria em Classifier categories para revelar os classificadores associados.
 - ii. Examine a lista de classificadores.
 - iii. Selecione **Adicionar** para encontrar e adicionar classificadores não listados da lista completa de classificadores disponíveis.
 - Se você deseja criar uma categoria personalizada, selecione **+ Add**.
 - i. Adicione um nome exclusivo, uma descrição e atribua os classificadores disponíveis à categoria.
 - ii. Selecione **Add**
4. Para desativar um classificador dentro de uma categoria, selecione **...** para o classificador e escolha

Desativar. Você também pode selecionar todas as linhas para fazer alterações de estado em massa.

Resultado

Categorias organizam classificadores para visibilidade da conformidade. Engenheiros de dados podem usar tags de classificação ao filtrar e criar coletas de dados.

Criar e gerenciar políticas de Data Guardrails

As políticas de Data Guardrails determinam como AIDE responde quando os classificadores detectam conteúdo sensível ou quando prompts e resultados da recuperação violam as regras de conteúdo.

Comportamentos típicos de Data Guardrails incluem:

- Mascaram ou redigir PII de trechos recuperados.
- Bloqueando respostas que violem as regras de conformidade.
- Registrar ou etiquetar violações para auditoria.

Sobre esta tarefa

Você cria e gerencia políticas de proteção somente no AIDE Console.

Você só pode associar workspaces no ONTAP System Manager a uma única política de guardrail por vez.

Passos

1. No AIDE Console, navegue até **Data Guardrails > Guardrail policies**.
2. Selecione **Add**.
3. Insira um nome e uma descrição que descrevam claramente o escopo (por exemplo, Customer PII redaction for support KB).
4. Configure as condições voltadas ao classificador de dados necessárias para a ativação do guardrail:
 - a. Defina as condições para ativação do Data Guardrails:
 - i. Escolha a categoria de classificador ou o tipo de classificador para cada condição.
 - ii. Adicione e defina condições adicionais conforme necessário.
 - iii. Defina critérios de pesquisa específicos em **Pesquisar**, depois selecione **Aceitar**.
 - b. Defina ações para a política de guardrail, como anonimizar conteúdo ou bloquear e remover um arquivo de uma coleta de dados.
5. Selecione a área de trabalho à qual o guardrail será aplicado.
6. Defina o estado da política:
 - **Ativado:** Ativa a política imediatamente.
 - **Modo de teste:** Permite validar o impacto da política antes de ativá-la.
 - **Desativado:** salva o Data Guardrails sem aplicá-lo.
7. Selecione **Adicionar** para salvar a política e aplicá-la ao workspace.



Utilize o **Modo de Teste** com um ambiente de trabalho piloto e uma coleta de dados não-produtiva para entender quantas respostas seriam afetadas antes de habilitar a aplicação rigorosa.

Resultado

A nova política de Data Guardrails está ativa e restrita ao espaço de trabalho selecionado.

Como as políticas interagem com os workspaces

Após a definição das políticas:

- O administrador de armazenamento usa ONTAP System Manager para criar espaços de trabalho, selecionar contêineres de dados e associar uma política de guardrail.
- Os classificadores são executados automaticamente no conteúdo do espaço de trabalho com base no que você ativou.
- Os Data Guardrails associados ao workspace influenciam como os endpoints de recuperação se comportam.

Para engenheiros de dados e cientistas de dados:

- O conjunto de dados visíveis (workspaces e coleções de dados) já está filtrado por atribuição de função.
- Os metadados que você consulta (por exemplo, tags PII) são definidos pelos classificadores que estão habilitados.
- As respostas que seus pipelines RAG recebem são limitadas pelos guardrails configurados no nível do espaço de trabalho.

Informações relacionadas

- ["Visualizar coleções de dados no AI Data Engine"](#)
- ["Início rápido Data-to-RAG para AI Data Engine"](#)

Perguntas frequentes sobre NetApp AI Data Engine

Este FAQ aborda dúvidas comuns sobre NetApp AI Data Engine (AIDE), incluindo sua arquitetura, implantação, tipos de usuários, recursos técnicos, integração e licenciamento.

Noções básicas de AIDE

O que é NetApp AI Data Engine (AIDE)?

NetApp AI Data Engine (AIDE) é um serviço de dados de IA integrado ao storage que abrange todo o ciclo de vida de IA, desde a descoberta e preparação de dados brutos até o fornecimento de pontos de recuperação para impulsionar IA generativa (GenAI), Retrieval-Augmented Generation (RAG), agentic AI e fábricas de IA. AIDE automatiza a sincronização e a detecção de alterações, fornecendo uma visão unificada e atualizada dos dados selecionados para descoberta e curadoria de dados.

Como o AIDE funciona?

AIDE integra-se diretamente com os sistemas de armazenamento NetApp ONTAP para criar uma visão global e estruturada de todo o conjunto de dados NetApp, com detecção e sincronização automatizadas de alterações. AIDE oferece vetorização em tempo real com compressão e deduplicação, Data Guardrails voltado a políticas e integração com ferramentas de IA.

Usuários e funções

Quem usa o AI Data Engine?

Os principais usuários do AIDE incluem:

1. **Administradores de storage ONTAP:** Gerenciam a infraestrutura, as necessidades de storage específicas para IA, a segurança e a conformidade.
2. **Engenheiros de dados:** gerenciam movimentação, preparação e integração de dados entre ambientes.
3. **Cientistas de dados:** preparar e transformar os dados relevantes para o consumo por IA.

Requisitos e implantação

Que hardware é necessário?

AIDE requer sistemas AFX para implantação (incluindo um controlador AFX, um gabinete de discos e um switch de rede), mas pode usar dados de cluster de clusters executando ONTAP 9 usando SnapMirror e peering de cluster. Pelo menos quatro nós de controlador AFX são necessários para implantações do AIDE, a fim de garantir alta disponibilidade e desempenho.

AIDE é executado em um nó de computação de dados NetApp (DCN). Três DCNs são necessários. O DCN hospeda o software AIDE, que inclui o Metadata Engine, Data Sync, Data Curator e Data Guardrails.

Posso usar minha própria DCN?

Não. O DCN é um nó de hardware de computação de dados fornecido pela NetApp e é o único mecanismo de implantação para o AI Data Engine.

Qual é o número mínimo de DCNs necessário?

São necessários exatamente três DCNs.

Qual sistema operacional é executado nos DCNs?

Os DCNs executam uma pilha de software fornecida pela NetApp com AIDE.

O AIDE pode ser implantado sem o AFX?

Não. AIDE requer AFX para implantação. AIDE usa Trident para consumir os volumes do AFX para storage interno (volumes persistentes). O cluster AFX que fornece storage para AIDE pode ser emparelhado com um sistema ou cluster ONTAP 9. Ele usa o emparelhamento de clusters e SnapMirror para sincronizar dados do cluster ONTAP remoto para o sistema AFX.

Gerenciamento e interfaces

O AI Data Engine Console faz parte do NetApp Console ou é uma interface separada?

O AIDE Console é uma interface de gerenciamento separada que é executada em DCNs. Você usa o AIDE Console para gerenciar serviços do AIDE, como Data Guardrails e Data Curator. Você também pode usar o ONTAP System Manager para monitorar o cluster AIDE.

Recursos e funcionalidades

Quais são os principais recursos do AIDE?

Existem quatro principais recursos do AIDE:

Metadata Engine

- Gera automaticamente uma visualização estruturada, atualizada e interativa dos seus dados.
- Funciona com dados armazenados no ONTAP.
- Permite que profissionais de dados colaborem com administradores de storage para encontrar e compreender dados.
- As APIs consultam metadados para fornecer funcionalidades enquanto reduzem a carga de tráfego NFS nos sistemas de armazenamento.
- A capacidade de extração e catalogação de metadados foi desenvolvida especificamente para AIDE, funciona de forma contínua e aproveita recursos do ONTAP, como snapshots.

Sincronização de dados

- Mantém a atualização dos dados automaticamente conforme os dados de origem são alterados sem intervenção manual.
- Os administradores definem o intervalo de atualização de dados em dias ou horas.
- Proporciona mobilidade de dados incremental e sincronização em todos os dados para eliminar cópias redundantes de dados de IA.

Data Guardrails

- Identifica e protege automaticamente dados confidenciais ao longo de todo o ciclo de vida da IA. Está acessível através do AI Data Engine Console.
- Analisa, classifica e categoriza dados continuamente.
- Identifica dados confidenciais (como PII) e riscos.
- Facilita a criação de políticas para o tratamento automático de dados confidenciais em conformidade com os padrões da empresa e normas regulamentares.

- Oferece redação automática de dados confidenciais para proteção de dados.
- Restringe o acesso a arquivos confidenciais conforme necessário.

Data Curator

- Permite que cientistas de dados pesquisem em todo o storage em busca de dados relevantes.
- Cria coleções de dados selecionados com dados existentes nos volumes do AFX.
- Gera representações vetoriais na camada de storage para reduzir dados brutos e aumentar o desempenho.
- Fornece um endpoint de recuperação para aplicações de IA com busca semântica vetorial e reclassificação.

Integração e interoperabilidade

AIDE oferece suporte a metadados federados em vários clusters ONTAP?

AIDE pode se conectar a vários clusters ONTAP usando SnapMirror e peering de clusters, permitindo visibilidade centralizada dos metadados.

Onde os metadados são armazenados?

AIDE armazena metadados no cluster AFX conectado usando um volume persistente fornecido pelo AFX. Os DCNs usam armazenamento local para operações internas.

O AIDE Metadata Engine classifica dados?

Não. O Metadata Engine cataloga metadados do sistema de arquivos e fornece APIs para consultar esses metadados catalogados.

Quais fontes de dados são suportadas?

AIDE suporta volumes ONTAP (locais ou remotos) como fontes de dados. Clusters ONTAP remotos devem executar ONTAP 9 e estar conectados via peering de cluster e SnapMirror.

Os buckets ONTAP S3 e os objetos StorageGRID não são suportados como fontes de dados no AIDE 9.18.1.

Quais tipos de arquivos a AIDE pode processar para classificação, vetorização e busca semântica?

AIDE suporta uma ampla variedade de tipos de arquivo, incluindo PDF, DOCX, PPTX, TXT e arquivos de imagem com recursos de OCR.

AIDE oferece suporte à classificação de dados em idiomas diferentes do inglês?

AIDE suporta apenas dados em inglês.

Quais integrações o AIDE suporta?

AIDE fornece um endpoint de API RAG acessível por meio de chamadas diretas à API ou por meio de um servidor de Model Context Protocol (MCP). Isso permite a integração com frameworks e ferramentas de IA agentic.

Implantação e licenciamento

Quais são as opções de implantação?

AIDE é implementado localmente na infraestrutura AFX com DCNs. Ele se integra diretamente com instalações AFX do NetApp ONTAP.

Como o AIDE é licenciado?

AIDE requer uma licença de software para executar Data Guardrails e Data Curator.

Se você precisar apenas do Metadata Engine, a licença ONTAP One, que está incluída em todos os sistemas AFX, concede direito apenas às funcionalidades do Metadata Engine.

Informações relacionadas

- ["Instale licenças AIDE no ONTAP System Manager"](#)
- ["Saiba mais sobre a arquitetura e os componentes do AIDE"](#)

Avisos legais

Avisos legais fornecem acesso a declarações de direitos autorais, marcas registradas, patentes e mais.

Direitos autorais

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

Marcas registradas

NETAPP, o logotipo da NETAPP e as marcas listadas na página de Marcas Registradas da NetApp são marcas registradas da NetApp, Inc. Outros nomes de empresas e produtos podem ser marcas registradas de seus respectivos proprietários.

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

Patentes

A lista atual de patentes detidas pela NetApp pode ser encontrada em:

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

Política de privacidade

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

Código aberto

Os arquivos de notificação fornecem informações sobre direitos autorais e licenças de terceiro usados no software NetApp.

AI Data Engine

["Aviso para AIDE 9.18.1"](#)

Informações sobre direitos autorais

Copyright © 2026 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

O software derivado do material da NetApp protegido por direitos autorais está sujeito à seguinte licença e isenção de responsabilidade:

ESTE SOFTWARE É FORNECIDO PELA NETAPP "NO PRESENTE ESTADO" E SEM QUAISQUER GARANTIAS EXPRESSAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO E ADEQUAÇÃO A UM DETERMINADO PROPÓSITO, CONFORME A ISENÇÃO DE RESPONSABILIDADE DESTES DOCUMENTOS. EM HIPÓTESE ALGUMA A NETAPP SERÁ RESPONSÁVEL POR QUALQUER DANO DIRETO, INDIRETO, INCIDENTAL, ESPECIAL, EXEMPLAR OU CONSEQUENCIAL (INCLUINDO, SEM LIMITAÇÕES, AQUISIÇÃO DE PRODUTOS OU SERVIÇOS SOBRESSALENTE; PERDA DE USO, DADOS OU LUCROS; OU INTERRUPÇÃO DOS NEGÓCIOS), INDEPENDENTEMENTE DA CAUSA E DO PRINCÍPIO DE RESPONSABILIDADE, SEJA EM CONTRATO, POR RESPONSABILIDADE OBJETIVA OU PREJUÍZO (INCLUINDO NEGLIGÊNCIA OU DE OUTRO MODO), RESULTANTE DO USO DESTES DOCUMENTOS, MESMO SE ADVERTIDA DA RESPONSABILIDADE DE TAL DANO.

A NetApp reserva-se o direito de alterar quaisquer produtos descritos neste documento, a qualquer momento e sem aviso. A NetApp não assume nenhuma responsabilidade nem obrigação decorrentes do uso dos produtos descritos neste documento, exceto conforme expressamente acordado por escrito pela NetApp. O uso ou a compra deste produto não representam uma licença sob quaisquer direitos de patente, direitos de marca comercial ou quaisquer outros direitos de propriedade intelectual da NetApp.

O produto descrito neste manual pode estar protegido por uma ou mais patentes dos EUA, patentes estrangeiras ou pedidos pendentes.

LEGENDA DE DIREITOS LIMITADOS: o uso, a duplicação ou a divulgação pelo governo estão sujeitos a restrições conforme estabelecido no subparágrafo (b)(3) dos Direitos em Dados Técnicos - Itens Não Comerciais no DFARS 252.227-7013 (fevereiro de 2014) e no FAR 52.227- 19 (dezembro de 2007).

Os dados aqui contidos pertencem a um produto comercial e/ou serviço comercial (conforme definido no FAR 2.101) e são de propriedade da NetApp, Inc. Todos os dados técnicos e software de computador da NetApp fornecidos sob este Contrato são de natureza comercial e desenvolvidos exclusivamente com despesas privadas. O Governo dos EUA tem uma licença mundial limitada, irrevogável, não exclusiva, intransferível e não sublicenciável para usar os Dados que estão relacionados apenas com o suporte e para cumprir os contratos governamentais desse país que determinam o fornecimento de tais Dados. Salvo disposição em contrário no presente documento, não é permitido usar, divulgar, reproduzir, modificar, executar ou exibir os dados sem a aprovação prévia por escrito da NetApp, Inc. Os direitos de licença pertencentes ao governo dos Estados Unidos para o Departamento de Defesa estão limitados aos direitos identificados na cláusula 252.227-7015(b) (fevereiro de 2014) do DFARS.

Informações sobre marcas comerciais

NETAPP, o logotipo NETAPP e as marcas listadas em <http://www.netapp.com/TM> são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.