



# AI Data Engine ドキュメント

## AI Data Engine

NetApp  
March 26, 2026

# 目次

AI Data Engineドキュメント	1
リリース ノート	2
AI Data Engineの新機能	2
AIDE 9.18.1初回リリースの新機能	2
AI Data Engineの既知の制限	3
AIDE 9.18.1初期リリースの既知の制限事項	4
はじめに	6
AI Data Engineシステムについて	6
AI Data Engineの詳細	6
AI Data Engineアーキテクチャ	7
役割別のAIDEコンポーネントと責任	11
AI Data Engineのクイックスタート	16
AIDEをインストール	17
AI Data Engineのインストール要件	17
AI Data Engine用のAFXストレージシステムをインストールする	18
データコンピューティングノードをインストールする	18
AIDEシステムをセットアップする	29
AI Data Engineのセットアップ	29
AFXシステムにAIDEライセンスをインストールする	31
ONTAP の AIDE に OpenID Connect を設定する	32
ワークスペースを設定する	36
AI Data Engineのデータを準備する	36
AI Data Engineでワークスペースを作成する	37
ワークスペースを作成する	38
ワークスペースの詳細を確認する	38
ワークスペースの更新とバージョン管理	39
AI Data Engineワークスペースへのユーザーアクセスの割り当て	39
管理と監視	41
クラスタプロセスを監視する	41
AI Data Engineシステムとクラスタのステータスの表示	41
インサイトを表示してAI Data Engineシステムを最適化します	43
AI Data Engineシステムイベント、ジョブ、および監査ログの表示	44
AI Data Engineワークスペースを管理	46
ワークスペースのステータスを確認する	46
ワークスペースのプロパティと更新スケジュールを編集する	47
既存のワークスペースにデータ コンテナを追加する	47
ワークスペースからデータ コンテナを削除する	48
ワークスペース ユーザーの管理	48
ワークスペースを削除する	48

AIDEシステムのアップグレードとメンテナンス	49
AI Data Engineシステムのアップデートと互換性	49
AI Data Engine softwareのアップデート	51
データコンピューティングノードをAI Data Engineクラスタに追加	52
AI Data Engineクラスタ内のノードを置き換える	54
ベクトル化とデータコレクションの管理	56
AI Data EngineのData-to-RAGクイック スタート	56
AI Data Engine Consoleでワークスペースのメタデータを探索する	57
データエンジニアまたはデータサイエンティストとして AI Data Engine Console に Sign in します	57
アクセス可能なワークスペースを表示する	57
次の手順	58
AI Data Engine Consoleでデータコレクションを作成する	58
ワークスペースのメタデータからデータコレクションを作成する	58
データ収集を公開する	59
データコレクションを更新または削除する	60
次の手順	60
AI Data Engineでデータコレクションを表示	61
クラスター全体のデータ コレクションを表示する	61
コレクション関連のジョブとイベントを監視する	62
AIDE Console からデータ コレクションを表示	62
次の手順	63
ガードレールを実装する	64
データ資産のためにAI Data Engineでガードレールポリシーを定義する	64
ポリシーのタイプを理解する	64
分類器を有効にする	64
分類カテゴリを管理する	65
ガードレール ポリシーの作成と管理	66
ポリシーとワークスペースの連携方法	67
関連情報	67
NetApp AI Data Engineに関するFAQ	68
AIDEの基本	68
ユーザーとロール	68
要件と導入	68
管理とインターフェイス	69
機能	69
統合と相互運用性	70
導入とライセンス	71
法律上の表示	72
著作権	72
商標	72
特許	72

プライバシー ポリシー .....	72
オープンソース .....	72
AI Data Engine .....	72

# AI Data Engine ドキュメント

# リリース ノート

## AI Data Engineの新機能

AI Data Engine (AIDE) 9.18.1は、NetAppのAIデータ管理のためのプラットフォームの初回リリースです。このリリースでは、組織がAIワークロードの非構造化データをカタログ化して整理できるようにするMetadata Engineと管理ワークフローが導入され、高度なガバナンスとベクトル化機能の基盤が提供されます。高度なガバナンス（ガードレール）とベクトル化は、適切なAI Data Engineライセンスを持つお客様が利用できます。

### AIDE 9.18.1初回リリースの新機能

AIDE 9.18.1 では、次の基本機能が導入されています。

#### "Metadata EngineによるAIデータカタログ化"

最初のリリースには、ONTAP クラスタ全体のファイルとオブジェクトをカタログ化する Metadata Engine が含まれています。

主な機能は次のとおりです：

- ピア クラスター上のローカルおよびリモート ONTAP ボリュームからのメタデータ（コア属性、拡張属性、オブジェクトタグ）の自動抽出。
- 企業データのグローバルビューを必要とするアプリケーション向けの集中クエリおよびフィルタリング REST API。
- スケーラブルな Metadata Engine ストレージ。
- ワークスペースの作成中にトリガーされる自動メタデータ抽出。

#### "ワークスペース管理"

ワークスペースは、AI プロジェクトのデータソース（ボリューム）の論理的なグループ化を提供します。

最初のリリースでは以下がサポートされます：

- ローカルおよびリモートのONTAPボリュームにまたがるワークスペースの作成（クラスターピアリングを使用）。
- ワークスペースへのアクセス制御の割り当て、マルチユーザーとマルチテナント環境のサポート。
- ワークスペースの作成時にメタデータを自動的に抽出し、カタログに入力します。

#### "自動データ更新のための Data Sync"

Data Sync により、ソース データが変更されても、手動による介入なしにメタデータ カタログとデータ コレクションが最新の状態に保たれます。

主な機能は次のとおりです：

- ポリシーベースの SnapMirror レプリケーションを使用した、リモートまたはローカルの ONTAP クラスターからのデータの自動同期。

- 変更されたデータのみを伝播し、オーバーヘッドを削減する増分更新。
- ワークスペースごとに設定可能な更新間隔。
- 同期ステータスとアクティビティのワークスペースレベルの監視。

### "クラスタのセットアップと管理"

最初のリリースには次のワークフローが含まれています：

- クラスタのセットアップ中にデータコンピューティングノード（DCN）を検出して追加します。
- Metadata Engine専用のメタデータストレージVMの作成。
- クラスタ全体のメタデータアクセス用の Data Engine サービスインターフェースの設定。
- 他の ONTAP クラスタとピアリングして、データ資産全体にメタデータカタログを拡張します。

### "OpenID Connect (OIDC) 認証"

- OIDC/OAuthベースの認証により、Microsoft Entra IDおよびActive Directory Federation Services（ADFS）を使用したONTAP System ManagerおよびData Engine Consoleへの安全なアクセスを実現。
- ワークスペースとメタデータ管理のためのロールベースのアクセス制御。

### "高度なデータキュレーションとガバナンス機能"

適切なAIDEライセンスをお持ちのお客様は、以下の機能をご利用いただけます：

- ベクトル化と**RAG**：AI Data Engine ConsoleでのデータコレクションAIDEワークスペースからのメタデータを使用した、埋め込み、および検索エンドポイントの作成。
- ガードレールベースのガバナンス：AI Data Engine Consoleでのガードレール ポリシーの定義、およびONTAP System Managerでのそれらのポリシーとワークスペースの関連付け。

サポートされているハードウェアとプラットフォーム

AIDE 9.18.1 は、以下の要素を組み合わせた ONTAP AI データ プラットフォーム クラスタで実行されます：

- AFX 1K ストレージ ノード
- NetApp データコンピューティングノード

関連情報

- ["AIDEの既知の制限事項"](#)
- ["AIDE のアーキテクチャとコンポーネントについて"](#)

## AI Data Engineの既知の制限

今回のリリースでサポートされていない、または今回のリリースでは正常に機能しないプラットフォーム、デバイス、機能が記載されています。これらの制限事項をよく確認してください。

## AIDE 9.18.1初期リリースの既知の制限事項

これらの制限は、AIDE 9.18.1 の Metadata Engine、データコンピューティングノード、および管理ワークフローに適用されます。

### コンピューティングノードの要件と管理

- 最小データコンピューティングノード要件

AIDE クラスタには、Metadata Engine 機能のために最小 3 台、最大 3 台のデータコンピューティングノード (DCN) が必要です。3 台未満の DCN を持つクラスタでは、Metadata Engine 機能を有効にできません。

- **DCN** ノードの追加はサポートされていません**NetApp Console** を使用

DCN ノードのアップグレードと追加は、NetApp Console ではなく、ONTAP System Manager を使用して実行する必要があります。

### サポートされているデータソース

- **ONTAP S3**バケットまたは**StorageGRID**のデータソースとしてのサポートはありません

ONTAPボリューム (ローカルまたはリモート) のみが、ワークスペースおよびメタデータカタログのデータソースとしてサポートされます。ONTAP S3バケットとStorageGRIDオブジェクトはワークスペースに追加できず、このリリースではMetadata Engineによってインデックス化されません。

- **FlexCache** ボリュームを使用したワークスペースの作成はサポートされていません

FlexCacheボリュームをデータソースとしてワークスペースに追加することはできません。

### ソフトウェアの更新と復元の制限

- **DCN** ノードのみの手動ソフトウェアアップデート

DCNクラスタの自動ソフトウェアアップデートは、AIDE 9.18.1ではサポートされていません。DCNノードソフトウェアは、ローカルクライアントからイメージをアップロードすることによってのみ更新できます。外部サーバー (HTTP/FTP) からのイメージのダウンロードはサポートされていません。

- **DCN** クラスターソフトウェアの復元は不可

DCN クラスターソフトウェアを以前のバージョンに戻すことはできません。以降のバージョンへのアップグレードのみが許可されます。

- **AFX** ストレージクラスタの **ONTAP** リバートなし

AFXストレージクラスタを以前のONTAPバージョンに戻すことはできません。以降のバージョンへのアップグレードのみが許可されます。

### ワークスペースのライフサイクルとアクセス構成

- ワークスペースのソフト削除または復元はできません

ワークスペースの削除は永続的です。削除されたワークスペースを復元するオプションはありません。

- 初期クラスタセットアップ中の **OIDC** 構成はサポートされません

OIDC/OAuthの設定は、クラスタ作成後にONTAP System Managerを使用して実行する必要があります。

#### 関連情報

- ["AI Data Engineの新機能"](#)

# はじめに

## AI Data Engineシステムについて

### AI Data Engineの詳細

NetApp AI Data Engine (AIDE) は、AI主導のデータ処理、管理、ガバナンスを加速し、簡素化するために設計されたエンタープライズクラスのプラットフォームです。AIDE は、大量の非構造化データを構造化されたAI対応データセットに変換するのに役立ちます。最新の機械学習 (ML) および生成AI (GenAI) ワークロードの需要を満たすように設計されており、従来のIT運用と新しいAI中心の役割の両方をサポートします。

### AIDE は AI の課題に対処する

AIDEは、組織がAIワークロードのデータを管理できるように設計されており、次の主要な機能を提供します：

- 集中メタデータ管理：AIDEはONTAPボリュームからメタデータを収集してカタログ化し、データセットの検索、分類、ガバナンス ポリシーの適用を可能にします。
- 自動化されたデータ処理：AIDE は、セマンティック検索用のベクトル埋め込みを生成する機能（適切なライセンスが必要）を含む、AI および ML ワークロード用のデータ パイプラインの作成をサポートします。
- データの分離とアクセス制御：AIDEは、複数のチームまたはプロジェクトに対してアクセス制御と基本的なデータ分離を適用します。
- **NetApp**ツールとの統合：AIDEは、ストレージ管理用のONTAP System Managerと連携し、データ エンジニアやデータ サイエンティストがデータ コレクションとワークフローを管理するための専用インターフェイス (AI Data Engine Console) を提供します。

### 高レベルの設計特性

以下の設計特性は、AIDE が AI ワークロードのニーズを満たすように構築されている方法を定義しています。

- マイクロサービス ベースのサービス：Kubernetes を使用して、メタデータのカタログ化、ベクター検索、インフラストラクチャ管理のためのモジュール式の回復力のあるサービスを調整します。
- エンタープライズクラスのセキュリティ：すべてのデータとメタデータに対して、暗号化、ロールベースのアクセス制御 (RBAC) 、および監査を実装します。
- マルチプロトコル データ アクセス：柔軟なデータの取り込みと取得のために NFS と SMB をサポートします。
- 自動化されたデータ パイプライン：データの変更を追跡し、埋め込みを作成し、AI アプリケーションのベクトル データベースを管理します。

### データが AIDE を流れる仕組み

データが AIDE を通じてどのように流れるかを理解することは、AI/ML チームにとってのプラットフォームの価値を説明するのに役立ちます。

1. データ取り込み：ファイルは標準プロトコル（NFSおよびSMB）を使用してONTAPボリュームに保存されます。データは、ローカルAIDEストレージ（AIDE導入環境内のAFXクラスタ）またはリモートONTAPクラスタに配置できます。リモートクラスタからのデータは、ONTAP SnapMirrorを使用してローカルAFXクラスタに同期されるため、AIDEで処理されるすべてのデータは最終的にローカルに保存され、アクセスされます。



S3 バケットは、ワークスペースまたはデータ コレクションのデータ ソースとしてサポートされていません。

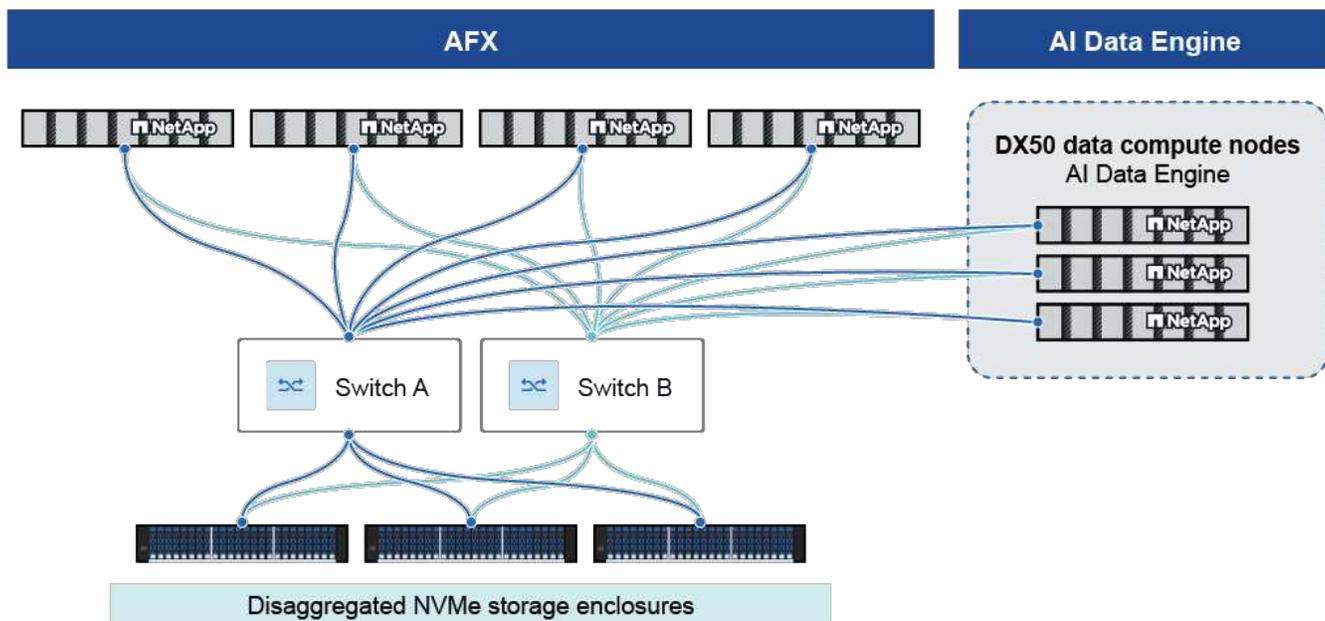
1. ワークスペースの作成：ストレージ管理者は ONTAP System Manager でワークスペースを定義し、特定のプロジェクト、チーム、またはワークフロー用に関連する ONTAP ボリュームをグループ化します。アクセス権限とガバナンス ポリシーはワークスペース レベルで割り当てられます。
2. メタデータ抽出：AIDE はワークスペース内のファイルとオブジェクトを自動的にスキャンし、メタデータ（ファイルの種類、サイズ、タイムスタンプ、カスタム属性）を抽出して、集中カタログに保存します。これは、データが変更されるたびに継続的に発生します。
3. 分類とガバナンス：分類子は、機密情報（PII、財務データ）またはドキュメントの種類（法務、人事）についてデータをスキャンします。ガードレール ポリシーは、編集またはアクセス制限を自動的に適用します。
4. データ収集の作成：データエンジニアとデータサイエンティストは、AIDE Consoleを使用してメタデータカタログを照会し、結果をフィルタリングし、特定のAIタスク向けに厳選されたデータコレクションを作成します。
5. ベクトル化：セマンティック検索を必要とするコレクションの場合、AIDEは選択したAIモデルを使用して埋め込みを生成します。ベクトルはベクトルデータベースに保存され、ハイパフォーマンス検索が可能になります。
6. AI/ML消費：アプリケーションは複数のパスを通じてデータにアクセスします：
  - NFSまたはSMBを使用した直接ファイル/オブジェクト アクセス
  - ベクター データベースに対するセマンティック検索クエリ
  - データ取得と GenAI モデル統合を組み合わせた RAG エンドポイント
  - プログラムによるワークフローのためのREST APIアクセス

この自動化されたポリシーベースのワークフローにより、AI用データの準備に必要な時間と手作業が削減され、チームはデータの処理ではなくモデルの開発と分析に集中できるようになります。

## AI Data Engineアーキテクチャ

AI Data Engine (AIDE) は、ストレージとコンピューティングを分離した、拡張性と耐障害性に優れたアーキテクチャに基づいて構築されており、AIワークロードに対して高いパフォーマンスと柔軟性を提供します。

物理コンポーネント



### AFX コントローラー ノード

AFX コントローラー ノードは、AFX 環境の要件をサポートするように設計された ONTAP ソフトウェアの特殊なパーソナリティを実行します。クライアントは、NFS や SMB などの複数のプロトコルを通じてノードにアクセスします。各ノードはストレージの完全なビューを持ち、クライアントの要求に基づいてアクセスできます。ノードは、重要な状態情報を保持するための不揮発性メモリを備えたステートフルであり、ターゲットワークロードに固有の追加の拡張機能も備えています。

高可用性とパフォーマンスを確保するために、AIDE の導入には少なくとも 4 つの AFX コントローラー ノードが必要です。

### データ コンピューティング ノード

データ コンピューティング ノード (DCN) は、AI データ処理タスク専用の高性能な CPU、RAM、GPU リソースを備えた Linux ベースのサーバーです。メタデータのカタログ化、ベクトル検索、埋め込みパイプラインなどの AI 固有のサービスをホストします。

AIDE デプロイメントには、正確に 3 つの DCN が必要です。

### クラスター/ストレージ スイッチ

冗長な高速 (100GbE 以上) スイッチが ONTAP と DCN を接続し、低レイテンシのデータ転送と高可用性を実現します。

### ストレージ シェルフ

高密度 SSD を搭載した NVMe-oF シェルフは、超低レイテンシと冗長性を提供し、PB 規模のストレージをサポートします。

### ネットワーク

すべての DCN と ONTAP ストレージ ノードは、冗長化された高速クラスター スイッチ (最小 100GbE) を介して接続されています。このアーキテクチャは、コンピューティング リソースとストレージ リソースを分離し、

それぞれを独立して拡張できるようにして、パフォーマンスとリソース使用率の両方を最適化します。

DCN と ONTAP ノード間のネットワークは、クラスター スイッチ上の専用 VLAN と IPspace を使用して分離されます。これにより、データ アクセス、管理 API、内部サービス トラフィックなどのすべての通信が安全かつ効率的に維持され、他のネットワーク操作に干渉することがなくなります。

## AI Data Engineの主な機能

AI Data Engine (AIDE) の主要機能が連携して、AIデータのライフサイクルの自動化、セキュリティ強化、および高速化を実現します。各機能は、DCN上で動作するマイクロサービスのセットとして実装され、ONTAP ストレージと統合され、REST APIおよび管理インターフェイスを通じて公開されます。

### Metadata Engine

Metadata Engineは、NetAppデータ資産の構造化された最新のインタラクティブなビューを自動的に生成します。

#### ライセンスとアクセス

Metadata Engine は、ベースの ONTAP One ライセンスに含まれており、AIDE のインストール時に利用できます。

ONTAP System Managerからアクセスできます。

#### 機能

- AFXクラスターにローカルに保存されたボリュームやリモートONTAPクラスターから同期されたボリュームを含む、すべてのデータソースのメタデータをカタログ化します。
- メタデータを自動的に抽出し、データが取り込まれたり変更されたりするとカタログに入力します。
- REST APIアクセスを提供してメタデータを照会し、データ実践者とストレージ管理者がデータを検出、分類、理解できるようにします。
- メタデータ クエリをデータ パスからオフロードし、ストレージ システム上の NFS トラフィック負荷を軽減します。
- インデックス作成および検索機能を備えた大規模なメタデータ レコードをサポートします。
- ワークスペースとデータ収集の抽象化と統合して、アクセス制御とガバナンスを適用します。

#### データ同期

Data Syncは、ソース データが変更された場合でも、メタデータ カタログとデータ コレクションが最新の状態に保たれ、基礎となるデータ ソースと一貫性が保たれることを保証する自動化されたバックグラウンド サービスです。

#### ライセンスとアクセス

データ同期機能は、基本の ONTAP One ライセンスに含まれており、AIDE のインストール時に利用可能になります。

#### 機能

- ポリシーベースのSnapMirrorレプリケーションを使用して、リモートまたはローカルのONTAPクラスターからデータを同期します。リモート クラスターからのデータは、AIDE処理のためにローカルのAFXクラスターにコピーされます。
- 検出された変更に基づいて増分的に更新し、変更されたデータのみを伝播します。

- データ資産全体にわたって、安全で増分的なデータの移動と同期を提供します。
- ワークスペースごとに設定可能な更新レートを使用して、同期間隔をスケジュールおよび監視します。
- ワークスペース作成ワークフローと統合して、新しいデータソースが追加されたときにメタデータを抽出および更新します。

## Data Guardrails

Data Guardrails サービスは、AI ライフサイクル全体にわたって機密データに対する継続的かつ自動化されたガバナンスと保護を提供します。

### ライセンスとアクセス

Data Guardrails機能は、ベースのONTAP Oneライセンスには含まれておらず、別途AIDEライセンスが必要です。

ガードレール機能には、AIDE Consoleからアクセスできます。

### 機能

- データを継続的にスキャン、分類、カテゴリ化します。
- PII 検出などのタスク用に、組み込みのカスタマイズ可能な分類子を使用して機密データとリスクを識別します。
- ポリシーベースの編集、マスキング、アクセス制限により、機密データの処理を自動化します。
- ワークスペースに添付された Data Guardrails ポリシーを通じて、会社および規制の標準を適用します。
- 監査ログとコンプライアンス レポートを使用して、構成された機密ファイルまたはボリュームへのアクセスを制限します。
- ワークスペースおよびデータ収集管理と統合して、AI データ ワークフロー全体に一貫してガードレールを適用します。

## Data Curator

Data Curatorサービスにより、AIおよびGenAIアプリケーションでの高速なデータ検出、検索、ベクトル化、取得が可能になります。

### ライセンスとアクセス

Data Curator機能は、ベースのONTAP Oneライセンスには含まれておらず、別途AIDEライセンスが必要です。

AIDE Console を通じて Data Curator にアクセスできます。

### 機能

- 集中化されたメタデータ カタログを使用して、ストレージ内の関連データを検索します。
- データサイエンティストがキュレーションされたデータコレクションを作成するためのツールを提供します。
- ストレージレイヤでベクトル埋め込みを自動的に生成します。
- AI アプリケーションに安全な検索エンドポイントを提供し、ベクトルセマンティック検索と再ランク付けをサポートします。
- 検索拡張生成 (RAG) パイプラインやエージェント AI フレームワークなどの AI ツールおよびテクノロジー

ーと統合します。

- データ コレクション、ベクター検索、および取得エンドポイントへのプログラムによるアクセス用の REST API を提供します。

## セキュリティとマルチテナンシー

プラットフォームは、ロールベースのアクセス制御 (RBAC) とリソース レベルのアクセス制御リスト (ACL) の両方を適用します。すべての API とユーザー アクションは監査され、すべてのデータは保存時および転送時に暗号化されます。個々のテナントは、データとメタデータで分離されています。

### 関連情報

- ["AIDEライセンスをインストールする"](#)
- ["データからRAGへのクイックスタート"](#)

## 役割別のAIDEコンポーネントと責任

### AI Data Engine コンポーネントとロールベースの相互作用

AI Data Engine (AIDE) は、AIワークロード向けの包括的なデータ管理および処理プラットフォームを提供するために連携する多くのコアコンポーネントで構成されています。これらのコンポーネントには、ワークスペース、データコレクション、ベクターデータベース、ガードレール、メタデータカタログ、取得エンドポイント、分類子が含まれます。各コンポーネントは、効率的なデータの検出、キュレーション、ガバナンス、AI/MLアプリケーションとの統合を可能にする特定の役割を果たします。

各AIDEユーザーは、役割に応じてAIDEコンポーネントとやりとりします。

### ストレージとデータに重点を置いたユーザーロール

AIDEは、従来のONTAPシステム管理ロールを引き続きサポートしながら、新しいユーザーロールを導入します：

#### ストレージ ユーザー

- ストレージ管理者：AFXとAIDEクラスタのセットアップ、ネットワーク、ストレージ プロビジョニング、およびユーザー アクセスを管理します。

#### データ利用者

- データ エンジニア：AI/ML パイプラインを構築および最適化し、データ収集を管理し、AI モデルを統合します。
- データ サイエンティスト：データセットを検出、キュレート、分析し、データ コレクションを作成し、GenAI アプリケーションの取得エンドポイントを活用します。

ロール (RBAC名)	概要
ストレージ管理者(admin)	AFXおよびAIDEクラスタのセットアップ、ネットワーク、ストレージ プロビジョニング、ユーザーアクセスを管理します。RBACロールをユーザーに割り当て、AIDEインターフェイスと機能へのアクセスレベルを決定します。この管理者ロールには、ONTAPシステムマネージャーとAIDE Consoleを使用した完全な管理アクセス権限があります。
データエンジニア(data-engineer)	AI/MLパイプラインの構築と最適化、データ収集の管理、AIモデルの統合を行います。この役割では、データエンジニアリングワークフロー用のAIDE Consoleにアクセスできます。
データサイエンティスト(data-scientist)	GenAIアプリケーション向けに、データセットの発見、キュレーション、分析、データコレクションの作成、および検索エンドポイントの活用を行います。この役割では、データサイエンスワークフロー用のAI Data Engine Consoleにアクセスできます。

## AIDE システム コンポーネント

各AIDEユーザー（ストレージ管理者、データエンジニア、データサイエンティスト）は、それぞれの役割に応じてAIDEコンポーネントを操作します。

### ワークスペース

ワークスペースはクラスタ内のデータの論理セグメントであり、特定のプロジェクト、チーム、またはワークフローのボリュームをグループ化します。ワークスペースは、AIDEにおけるデータの可視性、アクセス、ガバナンスの範囲を定義します。

### メタデータ カタログ

ローカル クラスタ全体のすべてのファイルとオブジェクトのメタデータ レコード（ONTAP SnapMirrorまたはクラスタ ピアリングを使用してリモート ONTAP クラスタから同期されたデータを含む）を格納する、集中型のスケーラブルなデータベース。豊富でインタラクティブな検索とフィルタリングが可能になります。

### 分類器

分類子は、特定の種類の機密データ（PII、財務、医療など）についてファイルをスキャンしてタグ付けしたり、ドキュメントを種類（法務、人事、販売など）別に分類したりするツール（組み込みまたはカスタム）です。

### データ コレクション

データ コレクションは、ワークスペースからの関連ファイルまたはオブジェクトのキュレーションされたグループであり、GenAI ワークフローで使用するためにユーザー指定のクエリによって定義されます。データ コレクション内のファイルの内容は、公開後、GenAI アプリケーションの API によるセマンティック検索に利用できるようになります。

### ベクター データベース

ベクターデータベースは、データコレクションから生成された埋め込みを保存し、AIおよびGenAIアプリケーション向けのハイパフォーマンスなセマンティック検索と取得を可能にします。

## ガードレール

ガードレールは、AIデータのライフサイクル全体にわたってデータ ガバナンス、分類、保護（編集やアクセス制限など）を強制するポリシーベースのメカニズムです。

## 取得エンドポイント（**RAG**エンドポイント）

検索エンドポイント（検索拡張生成エンドポイントまたは「RAG」エンドポイントと呼ばれることもあります）は、AI および GenAI アプリケーションが、キュレーションされたコレクションとベクター データベースから関連するデータ、コンテキスト、または埋め込みにアクセスできるようにする安全な API です。

RAG エンドポイントは、生成 AI モデルにおけるセマンティック検索やコンテキスト認識応答などの高度な AI ワークフローをサポートするように設計されています。AI アプリケーションを検索エンドポイントに接続することで、AIDE によって管理されるキュレーション済みの AI 対応データセットへのリアルタイム アクセスを提供し、モデルの精度と関連性を高めることができます。

## 関連情報

- ["AIDE ストレージ管理者が AIDE コンポーネントを操作する方法"](#)
- ["AIDE データ エンジニアが AIDE コンポーネントを使用する方法"](#)
- ["AIDE データ サイエンティストが AIDE コンポーネントを使用する方法"](#)

## AI Data Engine インターフェイス

AI Data Engine（AIDE）は、ユーザーとの対話と自動化のための3つの主要なインターフェイスを提供します。ストレージ管理者、データ エンジニア、データ サイエンティストなどの各役割は、それぞれの特定のタスクと責任に基づいてこれらのインターフェイスを利用します。

### ONTAP System Manager

ONTAP System Managerは、ストレージ管理者向けに設計されたWebベースのインターフェイスです。クラスタのセットアップ、ワークスペースの管理、DCNの監視、ガードレール ポリシーの適用のワークフローを提供します。

### AI Data Engine Console

AIDE Consoleは、データ エンジニアとデータ サイエンティスト専用のインターフェイスです。これにより、ユーザーはデータ ソースの探索、データ コレクションの作成と管理、データ パイプラインの設定、分類器の適用、ガードレールやベクトル検索機能との連携が可能になります。このコンソールは、データ探索、キュレーション、およびAI/MLワークフローとの統合のための高度なツールを提供します。

### REST API

AIDEは、自動化、統合、プログラムによるアクセスのためにONTAP REST APIを公開します。このAPIは、クラスタのセットアップ、ワークスペースとコレクションの管理、メタデータ クエリ、ベクター検索、および取得エンドポイントをサポートします。

## AI Data Engine ストレージ管理者がAIDEコンポーネントを操作する方法を学ぶ

ストレージ管理者として、ONTAPおよびAI Data Engine（AIDE） Consoleを通じてAIDE

インフラを管理し、ワークスペースのプロビジョニング、ガードレールポリシーの適用、およびシステムの状態監視を行います。あなたの役割は、AIワークロードのための信頼性が高く、安全で、コンプライアンスに準拠したデータストレージを確保することに重点を置いています。

#### ストレージ管理者コンポーネント アクセス

コンポーネント	アクセス レベル	ストレージ管理者のワークフロー
<b>ONTAP System Manager</b>	管理（作成、編集、削除）	ONTAP System Managerは、クラスター管理、ワークスペースのプロビジョニング、ガードレールポリシー管理、およびシステムヘルスの監視のための主要なインターフェイスとして使用します。
<b>AIDEコンソール</b>	管理（作成、編集、削除）	AIDE Consoleを使用して、ワークスペースを監視し、コレクションのステータスを表示し、AIDE環境全体のシステムアクティビティを監視します。
<b>ONTAP REST API</b>	管理（作成、編集、削除）	REST APIを使用して、インフラストラクチャタスクを自動化し、ワークスペースとガードレールポリシーをプログラムで管理し、AIDEの管理を外部ツールやワークフローと統合します。
ワークスペース	管理（作成、編集、削除）	ワークスペースの作成と管理には、ONTAP System Managerを使用します。含めるデータソースを選択し、データエンジニアとデータサイエンティストに権限を割り当て、ガードレールポリシーをアタッチしてガバナンスとコンプライアンスを強化します。また、ワークスペースの健全性とアクセスも監視します。
データ コレクション	表示（読み取り専用）	System Manager を使用して、各ワークスペース内のデータ収集のステータスと健全性を表示します。基礎となるデータソースが利用可能であり保護されていることを確認しますが、コレクションを作成または変更することはありません。
ガードレール	管理（作成、編集、削除）	System Manager を使用して、ガードレールポリシーを定義し、ワークスペースにアタッチします。ガードレールのステータスとコンプライアンス レポートを監視します。必要に応じてポリシーが確実に適用され、更新されます。
メタデータ カタログ	モニター（ヘルス、ステータス、アクティビティを表示）	メタデータ カタログが入力され、最新の状態であることを確認します。カタログの健全性を監視し、アクセス制御をサポートします。

コンポーネント	アクセス レベル	ストレージ管理者のワークフロー
ベクター データベース	プロビジョニング/監視（導入、設定、ステータスの表示）	ベクター データベース インフラストラクチャをプロビジョニングおよび監視し、GPU リソースと適切なライセンスを備えたデータ コンピューティング ノードが確実に配置されていることを確認します。環境をサポートしますが、埋め込みやクエリを直接管理しません。
分類器	管理（作成、編集、削除）	分類子とそのカテゴリを作成、構成、管理します。分類子をワークスペースに適用し、その有効性を監視します。

**AI Data Engine**のデータエンジニアとデータサイエンティストが **AIDE** コンポーネントを使用する方法を学ぶ

データ エンジニアまたはデータ サイエンティストは、AI Data Engine (AIDE) Console を使用して、アクセス権が付与されているワークスペースを探索したり、データ コレクションを作成および管理したり、セマンティック検索を実行したり、取得エンドポイントを AI/ML ワークフローに統合したりできます。

データ エンジニアは、コレクションの構築、埋め込みパイプラインの設定、公開されたコレクションにアクセスできるユーザーの制御を行うことで、生データを AI 対応データセットに変換することに注力します。データ サイエンティストは、アクセス制御やインフラストラクチャを管理することなく、分析、モデルトレーニング、GenAI アプリケーション用にキュレーションされたデータセットを活用することに重点を置いています。

データ ユーザー コンポーネント アクセス

コンポーネント	アクセス レベル	データ エンジニアのワークフロー	データサイエンティストのワークフロー
<b>AIDE</b> コンソール	管理（作成、編集、削除）	AIDE Consoleは、アクセス権限を持つワークスペースのデータ検出、コレクション管理、パイプライン構成、RAGまたは取得エンドポイントの公開など、日常的なタスクを実行するための主要なインターフェースです。	AIDE Consoleは、アクセス可能なワークスペース内でのデータ探索、コレクションの精緻化とバージョン管理、およびキュレーションされたデータセットと取得エンドポイントを分析、モデリング、およびGenAIワークフローに接続するための主要なインターフェースです。
<b>ONTAP REST API</b>	管理（作成、編集、削除）	REST API を使用して、コレクションのライフサイクル操作を自動化し、埋め込みパイプラインをトリガーおよび監視し、データ ワークフローを外部ツールとプログラマ的に統合します。	REST APIを使用して、プログラムでデータ コレクションにアクセスし、ベクトル検索クエリを実行し、取得エンドポイントをAI/MLアプリケーションおよびエージェント フレームワークに統合します。
ワークスペース	表示/使用（読み取り専用）	コレクションを構築する前に、割り当てられたワークスペースを調査して、利用可能なデータソースを識別して理解します。	割り当てられたワークスペースを検索して、特定の調査またはモデリング タスクに関連するファイルとオブジェクトを見つけます。

コンポーネント	アクセス レベル	データ エンジニアのワークフロー	データサイエンティストのワークフロー
データ コレクション	管理（作成、編集、削除）	タグ、分類、その他の属性を使用してソース データを選択およびフィルタリングすることでデータ コレクションを構築し、AI で使用するための RAG エンドポイントとして作成、バージョン管理から公開まで、コレクションのライフサイクル全体を管理します。また、各コレクションにアクセスできるデータ サイエンティストやその他のユーザーも管理します。	アクセス権が付与されたワークスペース内で、データ コレクションを作成、選択、注釈付け、バージョン管理、および調整できます。これらのコレクションは、セマンティック検索と GenAI ワークフローの基盤として使用されます。
メタデータ カタログ	クエリ/使用（ワークフローで使用）	メタデータ カタログを使用して、取り込み用のデータ ソースを評価および選択し、クエリを実行して関連ファイルを検索し、割り当てられたワークスペース内で構築しているコレクションの要件を満たしていることを確認します。	データ エンジニアによって構築および管理されているカタログ構造に依存して、アクセスできるワークスペース全体でメタデータを検索およびフィルター処理し、分析やモデルのトレーニングに必要なファイルとオブジェクトを見つけます。
ベクター データベース	<ul style="list-style-type: none"> <li>埋め込み/検索の管理（データ エンジニア）</li> <li>使用/検索（データサイエンティスト）</li> </ul>	埋め込みパイプラインをトリガーし、ベクトル化ステータスを監視し、チャンクと埋め込みパラメータを設定し、ベクトル検索でサポートされる取得エンドポイントを公開します。アプリケーションとエージェントは、セマンティック検索と RAG ワークフローのために API を介してこれらのエンドポイントをクエリします。	データ エンジニアが管理するパイプラインによって生成された埋め込みに対してセマンティック検索クエリを実行し、検索結果を GenAI または RAG ワークフローに統合して、コンテキスト認識モデル応答を実現します。チャンク、埋め込み、またはパイプライン パラメータは構成しません。
分類器	使用（分類されたデータの消費）	分類結果を使用して、コレクションの準備中にソース データに注釈を付けてタグ付けし、パイプラインに入るコンテンツがダウンストリームの AI ワークフローに対して適切にラベル付けされるようになります。	事前に分類されたデータを使用することで、分析とモデリングでは準拠した関連コンテンツのみが使用されるようになります。

## AI Data Engine のクイックスタート

AI Data Engine (AIDE) システムをすぐに使い始めるには、ハードウェア コンポーネントをインストールし、クラスタをセットアップし、ホストからストレージ システムへのデータ アクセスをセットアップし、ストレージをプロビジョニングする必要があります。

1

AIDEは新しいAFXクラスターまたは既存のAFXクラスターとともにインストールされますか？

AIDE と AFX を同時にインストールするか、AIDE を既存の AFX クラスターと統合するかを決定する必要があります。

2

ハードウェアのインストールとセットアップ

"インストールとセットアップ" AIDE クラスターのコンピューティングノード。設置環境に応じて、"インストール" AFX ハードウェアも確認してください。

3

クラスターをセットアップする

ONTAP System Managerを使用すると、"AFX クラスターを使用した AIDE のセットアップ"の迅速で簡単なプロセスをガイドします。

4

ワークスペースとデータアクセスを設定する

"ワークスペースを設定し、そのワークスペースのAIDEデータにアクセスできるユーザーを設定します"。

次の手順

これで ONTAP System Manager を使用して AIDE を管理し、データ エンジニアとデータ サイエンティストがワークスペースと構成を使い始められるように準備できます。

## AIDEをインストール

### AI Data Engineのインストール要件

AI Data Engine (AIDE) のインストール要件を確認してください。AIDEには、AFXストレージシステム、最低3つのデータ処理ノード、ネットワークスイッチ、およびケーブルが必要です。

#### ハードウェア要件

AIDEにはAFXストレージシステムと、最低3つのデータコンピューティングノードが必要です。AFXシステムはストレージインフラストラクチャを提供し、データコンピューティングノードはデータ管理、キューレーション、およびAI機能を実現するAIDEソフトウェアコンポーネントをホストします。

- **AFXストレージシステム**：AFXコントローラ、ディスク シェルフ、ネットワーク スイッチが含まれます。AFXストレージシステムはAIDE導入に必要です。
- **Data Compute Node**：少なくとも3つのData Compute Nodeが必要です。Data Compute Node は、NetAppが提供するハードウェアノードで、Metadata Engine、Data Sync、Data Curator、Data GuardrailsなどのAIDEソフトウェアをホストします。
  - 各 Data Compute Node には、接続に使用できる I/O スロット 4 と 5 があります。スロット 3 は GPU 用に予約されています。スロット 1 と 2 には何も装着されていないか、アクセスできません。
  - ポート e4a と e5a はクラスター接続に使用されます。

- ポート e4b と e5b はホスト ネットワーク接続に使用されます。

## ネットワーク スイッチの要件

AIDE では、データ処理ノードのホスト ネットワーク接続とノード間通信を有効にするために、ネットワーク スイッチが必要です。

- ホスト ネットワーク接続用のクライアント スイッチ (Cisco Nexus 9332D-GX2B または Cisco Nexus 9364D-GX2A)
- ノード間通信用のクラスタ スイッチ (Cisco Nexus 9332D-GX2B または Cisco Nexus 9364D-GX2A)
- ネットワーク管理用のオプションの管理スイッチ

## ケーブル要件

Data Compute Node をネットワーク スイッチおよび管理ネットワークに接続するには、次のケーブルが必要です。

- ノードをクライアントおよびクラスタ スイッチに接続するための400GbEから100GbE (4x100GbE) へのブレイクアウト ケーブル

## マルチクラスタサポート

AIDE を AFX で展開した後、SnapMirror とクラスタ ピアリングを使用して、他の ONTAP 9.18.1 以降のクラスタに接続し、データを管理できます。

## AI Data Engine用のAFXストレージシステムをインストールする

AI Data Engine (AIDE) を導入する最初のステップとして、AFXストレージシステムをインストールします。AFXストレージシステムはストレージインフラストラクチャの基盤を提供するものであり、データ処理ノードをインストールする前に必要となります。

["AFX 1K インストールドキュメント"](#)に従って、AFX ストレージ システムをインストールします。

### 次の手順

AFXストレージシステムのインストールが完了したら、["データ計算ノードをインストールする"](#)。

## データコンピューティングノードをインストールする

### AI Data Engineのデータコンピューティングノードのインストールとセットアップのワークフロー

データコンピューティングノード (DCN) をインストールして構成するには、ハードウェア要件を確認し、サイトを準備し、ハードウェアコンポーネントをインストールしてケーブルで接続し、システムの電源を入れ、ONTAPクラスタをセットアップします。

1

#### "ハードウェアのインストール要件を確認する"

既存のAFX 1Kストレージシステムがインストールされていることを確認し、AIDEのデータコンピューティングノードをインストールするためのハードウェア要件を確認します。AFX 1K ストレージ システムのインスト

ールについては、"[AFX 1K ストレージシステムのインストールドキュメント](#)"を参照してください。

2

### "データコンピューティングノードのインストールの準備"

データ処理ノードを設置する準備として、設置場所の準備、環境および電気に関する要件の確認、そして十分なキャビネットスペースの確保が必要です。次に、機器を開梱し、内容物を梱包明細書と照合し、ハードウェアを登録してサポート特典を受けられるようにしてください。AIDEを導入するには、最低3つのデータ処理ノードが必要です。

3

### "データコンピューティングノード用のハードウェアをインストールする"

データコンピューティングノード用のレールキットをインストールします。キャビネット内のデータコンピューティングノードを固定します。最後に、ケーブル配線を整理するために、システムの背面にケーブル管理デバイスを取り付けます。

4

### "データ計算ノードをケーブルで接続する"

ハードウェアをケーブル接続するには、まずデータ処理ノードをデータネットワークに接続し、次にデータ処理ノードをクラスタスイッチに接続します。

5

### "データコンピューティングノードの電源をオンにする"

ラックハードウェアをインストールし、データコンピューティングノードをケーブル接続した後、まだ電源がオンになっていない場合は、DCN と AFX ストレージシステムのコントローラーノードの電源をオンにする必要があります。

## AI Data Engineのデータコンピューティングノードのインストール要件

AI Data Engine (AIDE) のデータコンピューティングノードに必要な機器と持ち上げ時の注意事項を確認してください。

### 前提条件

AIDE のデータ計算ノードをインストールする前に、以下の点を確認してください：

- AFX 1K ストレージ システム。



AFX 1K ストレージ システムのインストールについては、"[AFX 1K ストレージシステムのインストールドキュメント](#)"を参照してください。

### インストールに必要な機器

AIDE のデータ計算ノードをインストールするには、以下の機材とツールが必要です。

- データコンピューティングノードを構成するための Web ブラウザへのアクセス
- 静電気放電 (ESD) ストラップ
- 懐中電灯

- USB/シリアル接続を備えたラップトップまたはコンソール
- No.2プラス ドライバ

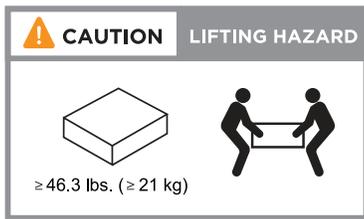
#### 持ち上げ時の注意事項

データ計算ノードは重いです。これらの品物を持ち上げたり移動したりするときには注意してください。

#### データ計算ノードの重み

データコンピューティングノードを移動または持ち上げる場合は、必要な予防措置を講じてください。

データ コンピューティング ノードの重量は最大 46.3 lbs (21 kg) になります。データ コンピューティング ノードを持ち上げるには、2 人または油圧リフトを使用します。



#### 関連情報

- ["安全情報と規制に関する通知"](#)

#### 次の手順

ハードウェア要件を確認した後、["データコンピューティングノードのインストールの準備"](#)。

### AI Data Engineのデータコンピューティングノードのインストール準備

AI Data Engine (AIDE) のデータ コンピューティング ノードをインストールする準備として、サイトの準備、箱の開梱、箱の中身と梱包明細書の照合、サポート特典を受けるためのシステム登録を行います。

#### ステップ1：サイトの準備

データコンピューティングノードをインストールするには、使用する予定のサイトとキャビネットまたはラックが構成の仕様を満たしていることを確認します。

#### 手順

1. ["NetApp Hardware Universe"](#)を使用して、サイトがデータコンピューティングノードの環境要件と電気要件を満たしていることを確認します。
2. 既存の AFX 1K ストレージ システムのインストール内に、データ コンピューティング ノード用の十分なキャビネット スペースがあることを確認します。
  - データ計算ノードごとに1U
  - 各 AFX 1K コントローラーノードに 2U
  - 各NX224シェルフに2U
  - スイッチモデルに応じて、スイッチあたり1Uまたは2U。

## ステップ2：箱を開ける

データコンピューティングノードに使用する予定のサイトとキャビネットまたはラックが必要な仕様を満たしていることを確認したら、すべてのボックスを開梱し、内容物を梱包明細書の項目と比較します。

### 手順

1. すべての箱を慎重に開けて、中身を整理して並べます。
2. 開梱した品目を梱包明細書と比較し、相違点があれば記録します。

配送用カートンの側面にある QR コードをスキャンすると、梱包リストを取得できます。

以下の項目は、箱の中に入っている可能性のある内容物の一部です。

ハードウェア	ケーブル	
<ul style="list-style-type: none"><li>• ベゼル</li><li>• 説明書付きレールキット</li><li>• データコンピューティングノード</li></ul>	<ul style="list-style-type: none"><li>• 電源コード</li></ul>	

## ステップ3：データコンピューティングノードを登録する

サイトがデータコンピューティングノードの仕様要件を満たしていることを確認し、注文した部品がすべて揃っていることを確認したら、システムを登録する必要があります。

### 手順

1. データコンピューティングノードのシリアル番号を確認します。

シリアル番号は次の場所にあります：

- 梱包明細書に記載
- 確認メールに記載
- 各データ計算ノード、または一部のシステムでは、各データ計算ノードのシステム管理モジュール。

SSN: XXYYYYYYYYYY



2. ["NetAppサポート サイト"](#)に移動します。
3. ストレージシステムを登録する必要があるかどうかを判断します：

もしあなたが...	次の手順に従ってください。
NetAppの既存のお客様	<ol style="list-style-type: none"> <li>自身のユーザ名とパスワードでサインインします。</li> <li>システム &gt; *マイシステム*を選択します。</li> <li>新しいシリアル番号がリストされていることを確認します。</li> <li>シリアル番号が記載されていない場合は、新しいNetAppのお客様向けの手順に従ってください。</li> </ol>
NetAppの新規のお客様	<ol style="list-style-type: none"> <li>今すぐ登録 をクリックしてアカウントを作成します。</li> <li><b>Systems</b> &gt; *Register Systems*を選択します。</li> <li>ストレージシステムのシリアル番号と要求された詳細を入力します。</li> </ol> <p>NetApp が登録を承認すると、必要なソフトウェアをダウンロードできます。承認には最大 24 時間かかります。</p>

#### 次の手順

データコンピューティングノードのインストール準備が完了したら、"[データ計算ノードをインストールする](#)"。

#### AI Data Engineのデータ計算ノードをインストールする

データ コンピューティング ノードをキャビネット内に設置して固定します。

#### 開始する前に

- レール キットに同梱されている手順書があることを確認してください。
- データ コンピューティング ノード、ストレージ システム、およびストレージ シェルフの重量に関連する安全上の懸念事項を理解します。
- ストレージ システムを通る空気の流れは、ベゼルまたはエンド キャップが取り付けられている前面から入り、ポートがある背面から排気されることを理解してください。



一般的に、スイッチはキャビネットの中央に設置する必要があります。ストレージ シェルフは、スイッチの下、2 番目に設置されたスイッチの上に設置する必要があります。コントローラ ノードは、キャビネット内のスイッチの上または下にインストールできます。データ コンピューティング ノードは、キャビネット内のコントローラ ノードの上または下にインストールできます。

#### 手順

1. 必要に応じて、キットに付属の説明書に従って、データ コンピューティング ノード用のレール キットをインストールします。
2. データ コンピューティング ノードをキャビネット内に設置して固定します。
  - a. データ コンピューティング ノードをキャビネットの中央のレールに配置し、アプライアンスを下から支えて所定の位置にスライドさせます。
  - b. 付属の取り付けネジを使用して、data compute nodeをキャビネットに固定します。

3. ベゼルをデータ コンピューティング ノードの前面に取り付けます。

次の手順

データ計算ノードをインストールしたら、"[AIDE のデータ計算ノードをケーブルで接続する](#)"。

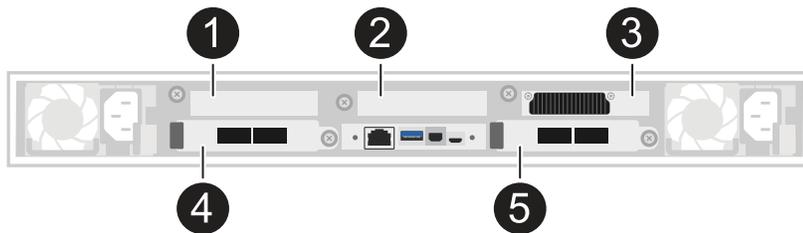
### AI Data Engineのデータ計算ノードのケーブル配線要件

Data Compute Node は、ホスト ネットワークおよびクラスター ネットワーク接続を通じて AFX 1K ストレージ システムと統合されます。展開の I/O スロット構成、ケーブルタイプ、および接続要件を確認します。

#### ケーブル構成

Data Compute Nodeは、AFX 1Kコントローラー ノードと同じクラスター スイッチに接続し、AIおよび機械学習のワークロードに最適化されたコンピューティング リソースでストレージ システムを拡張します。

初期のAI Data Engine (AIDE) 構成では、少なくとも3つのData Compute Nodeがサポートされます。包括的な構成の詳細とスロットの優先順位については、"[NetApp Hardware Universe](#)"を参照してください。



①	Data Compute Node上の未使用のスロット。
②	Data Compute Node上の未使用のスロット。
③	Data Compute Node 上の GPU スロット。
④	Data Compute Node 上の I/O スロット。
⑤	Data Compute Node 上の I/O スロット。

#### I/O スロット構成

Data Compute Node は、標準のサーバー構成とは異なる特定のスロット番号付けスキームを使用します。適切なケーブル配線を行うには、スロットのレイアウトを理解することが重要です。

- スロット 3：GPU 用に予約済み (I/O ケーブルではアクセスできません)
- スロット 4 と 5：ネットワーク接続に使用される I/O スロット
  - ポート a：クラスター ネットワーク接続
  - ポート b：ホスト ネットワーク接続

- スロット 1 と 2：未使用でアクセス不可

## ネットワーク接続

データ コンピューティング ノードを AFX 1K ストレージ システムと統合するには、2 種類のネットワーク接続が必要です。

### • ホスト ネットワーク接続

ホスト ネットワーク接続は、クライアント データへのアクセスを提供し、Data Compute Nodeがワークロードを処理できるようにします。各Data Compute Nodeは、個別のホスト ネットワーク スイッチへの冗長接続にポート e4b と e5b を使用します。

ポートの割り当て：

- e4b：ホスト ネットワーク スイッチ A に接続
- e5b：ホスト ネットワーク スイッチ B に接続

### • クラスタ ネットワーク接続

クラスタ ネットワーク接続により、ストレージ クラスタ内の Data Compute Node と AFX 1K コントローラー ノード間の通信が可能になります。各 Data Compute Node は、個別のクラスタ ネットワーク スイッチへの冗長接続にポート e4a と e5a を使用します。

ポートの割り当て：

- e4a：クラスタ ネットワーク スイッチ A に接続
- e5a: クラスタ ネットワーク スイッチBに接続

## サポートされているハードウェア コンポーネント

データ コンピューティング ノードには、AFX 1K ストレージ システムとの適切な接続とパフォーマンスを確保するために、特定のケーブルとネットワーク スイッチが必要です。

Data Compute Node	サポートされているスイッチ	対応ケーブル
Data Compute Node (最低3つ必要)	<ul style="list-style-type: none"> <li>• Cisco Nexus 9332D-GX2B (400GbE)</li> <li>• Cisco Nexus 9364D-GX2A (400GbE)</li> </ul>	<ul style="list-style-type: none"> <li>• Data Compute Node への接続用の 4x100GbE QSFP56 ケーブルへの 400GbE QSFP-DD ブレークアウト： <ul style="list-style-type: none"> <li>◦ Data Compute Node クラスタ ネットワーク ポートへの 100GbE (e4a、e5a)</li> <li>◦ 100GbE から Data Compute Node ホスト ネットワーク ポート (e4b、e5b)</li> </ul> </li> <li>• 管理接続用のRJ-45ケーブル</li> </ul>

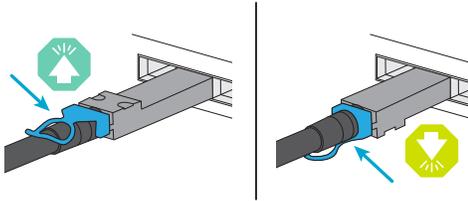


ブレークアウトケーブルは、各400GbEスイッチ ポートから4つの100GbE接続を提供しません。400GbE端をスイッチに接続し、100GbE端をData Compute NodeのI/Oポートに接続しません。

## ケーブルの向き

ケーブルをData Compute Nodeに接続する場合、ケーブルの向きを正しくすることで信頼性の高い接続が保証されます。

インストール手順の配線図には、コネクタをポートに挿入するときに、ケーブルコネクタのプルタブの正しい方向（上または下）を示す矢印アイコンが表示されます。コネクタを挿入すると、カチッと音がして所定の位置に収まるはずですが、カチッという音がしない場合は、取り外して裏返し、もう一度試してください。



コネクタ部品をカチッとはめるときは、繊細な部品を慎重に扱ってください。

## 次の手順

ケーブル構成を確認した後、"[ハードウェアのケーブル接続](#)" Data Compute Node用。

## AI Data EngineのData Compute Nodeをケーブル接続する

Data Compute Nodeをホスト ネットワークとクラスター ネットワーク スイッチに接続して、AIワークロード処理とAFX 1Kストレージ システムとの統合を可能にします。この手順では、ホスト ネットワーク アクセスとクラスター通信の両方に100GbE接続を使用し、AFXシステムの電源を切ることなく、ノードが既存のクラスター インフラストラクチャを活用できるようにします。

## タスク概要

これらの手順は一般的な構成を示しています。具体的なケーブル接続は、ストレージ システム用に注文したコンポーネントによって異なります。包括的な構成の詳細とスロットの優先順位については、"[NetApp Hardware Universe](#)"を参照してください。



データ コンピューティング ノードをケーブル接続するときに、AFX 1K ストレージ システムの電源をオフにする必要はありません。すでに電源がオンになっていて構成されている既存のAFX 1K ストレージ システムにデータ コンピューティング ノードを追加できます。

## 開始する前に

- 既存のAFX 1K ストレージ システムがインストールされています。AFX 1K ストレージ システムのインストールについては、"[AFX 1K ストレージシステムのインストールドキュメント](#)"を参照してください。
- 必要なネットワーク スイッチがインストールされ、構成されています。システムをネットワーク スイッチに接続する方法については、ネットワーク管理者にお問い合わせください。
- "[Data Compute Node のケーブル配線要件](#)"をレビューしました。



AI Data Engineを展開するには、最低3つのデータ計算ノードが必要です。

## ステップ1：Data Compute Nodeをホストネットワークに接続する

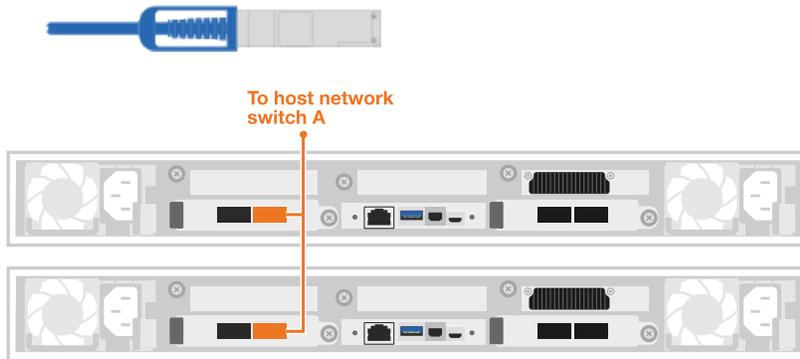
Data Compute Node のポートをホスト ネットワークに接続できます。

### 手順

1. 次の Data Compute Node のポート e4b をイーサネット データ ネットワーク スイッチ A に接続します：

- Data Compute Node 1、ポート e4b
- Data Compute Node 2、ポート e4b

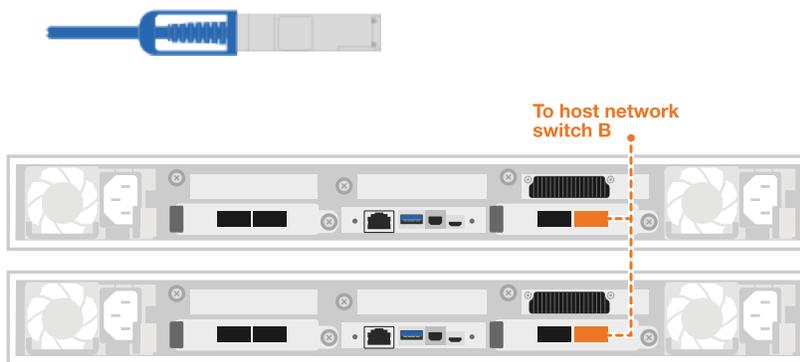
#### 100GbE ケーブル



2. 次の Data Compute Node のポート e5b をイーサネット データ ネットワーク スイッチ B に接続します：

- Data Compute Node 1、ポート e5b
- Data Compute Node 2、ポート e5b

#### 100GbE ケーブル



## ステップ2：Data Compute Nodeクラスタ接続をケーブル接続する

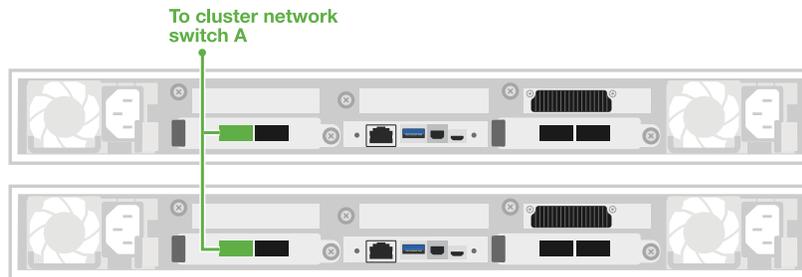
データコンピューティングノードの場合は、4x100GbEブレイクアウトケーブルを使用して、クラスタ接続用のe4a/e5aポートを接続します。

### 手順

1. 次の Data Compute Node のポート e4a をクラスタ ネットワーク スイッチ A の非 ISL ポートに接続します：

- Data Compute Node 1、ポート e4a
- Data Compute Node 2、ポート e4a

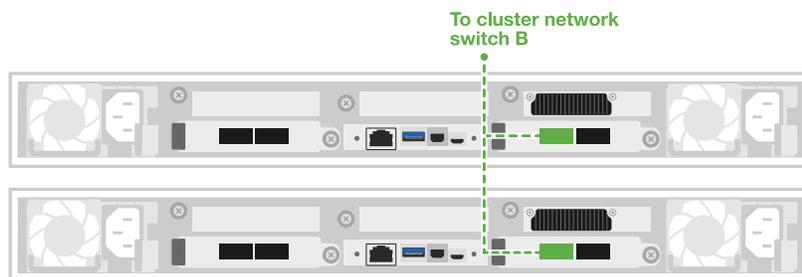
#### 4x100GbE ブレークアウトケーブル



2. 次のデータ コンピューティング ノードのポート e5a をクラスタ ネットワーク スイッチ B の非 ISL ポートに接続します：

- Data Compute Node 1、ポート e5a
- Data Compute Node 2、ポート e5a

#### 4x100GbE ブレークアウトケーブル



#### 次の手順

ハードウェアの配線が終わったら、"[Data Compute Node の電源をオンにする](#)".

#### AI Data Engineのデータ計算ノードの電源をオンにします

ラック ハードウェアをインストールし、データ コンピューティング ノードをケーブル 接続した後、まだ電源がオンになっていない場合は、DCN と AFX ストレージ システム のコントローラー ノードの電源をオンにする必要があります。

#### 開始する前に

- 棚の電源がオンになっており、各棚に固有の棚 ID が割り当てられていることを確認します。AFX ストレージシステムのシェルフ ID の割り当てについては、"[固有の棚IDの割り当てに関するドキュメント](#)"を参照してください。

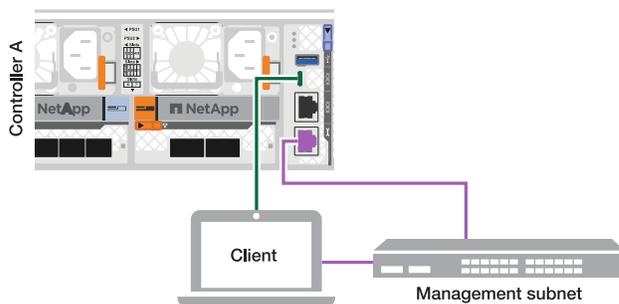
## 手順

ストレージシェルフの電源をオンにして一意の ID を割り当てた後、DCN の電源をオンにし、ストレージコントローラノードの電源をオンにします（まだオンになっていない場合）。

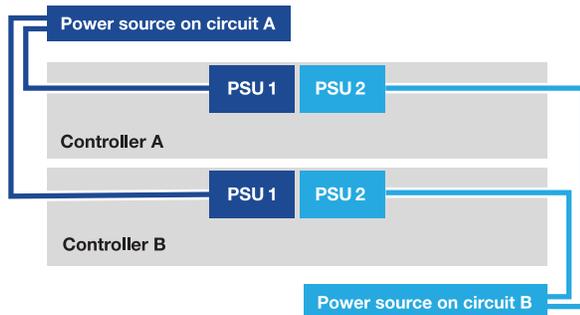
1. ラップトップをシリアルコンソールポートに接続します。これにより、コントローラの電源がオンになったときのブートシーケンスを監視できます。
  - a. ラップトップのシリアルコンソールポートを115,200ボー、N-8-1に設定します。

シリアル コンソール ポートの設定方法については、ラップトップのオンライン ヘルプを参照してください。

- b. コンソール ケーブルをラップトップに接続し、ストレージ システムに付属のコンソール ケーブルを使用してコントローラのシリアル コンソール ポートに接続します。
- c. ラップトップを管理サブネット上のスイッチに接続します。



2. 管理サブネット上のアドレスを使用して、ラップトップに TCP/IP アドレスを割り当てます。
3. 電源コードをコントローラーの電源に差し込み、別の回路の電源に接続します。



- システムの起動が始まります。最初の起動には最大8分かかる場合があります。
  - LED が点滅し、ファンが始動します。これは、コントローラーの電源がオンになっていることを示しています。
  - 起動時にファンの騒音が大きくなる場合がありますが、これは正常です。
4. 電源コードをデータコンピューティングノードの電源に差し込み、別の回路の電源に接続します。
  5. 各電源装置の固定装置を使用して電源コードを固定します。
  6. データコンピューティングノードの電源をオンにします。

電源スイッチにアクセスするには、ベゼルを取り外す必要がある場合があります。その場合は、後で必ず再度取り付けてください。

次の手順

データコンピューティングノードをオンにしたら、"[ONTAP AIDEクラスタをセットアップする](#)".

## AIDEシステムをセットアップする

### AI Data Engineのセットアップ

ONTAP ストレージ管理者は、AI Data Engine (AIDE) を AFX システムの展開に統合できます。初期 AFX ストレージ システムの導入の一部として AIDE をセットアップするか、すでに導入されてデータを提供している AFX クラスタに AIDE を追加することができます。どちらの場合も、基本セットアップの完了後に AIDE 設定を完了する必要があります。

#### 1. AIDEを設定し、AFXストレージシステムと統合する

AIDE を設定して AFX ストレージ システムと統合する方法は 2 つあります。ご使用の環境に適したオプションを選択してください。



AIDE をセットアップする前に、データ コンピューティング ノードがインストールされ、クラスタ スイッチに接続されていることを確認してください。詳細については、"[AIDEをインストール](#)"を参照してください。

#### 新しいAFXクラスタでAIDEを設定

初期 AFX システム導入の一環として AIDE をセットアップする場合、AIDE セットアップ プロセスは標準の AFX クラスタ セットアップ ワークフローの一部です。ONTAP クラスタのセットアップ中に、System Manager は接続された DCN を自動的に検出し、クラスタ構成の一部として含めます。その後、必要なライセンス構成を含む AIDE 構成を完了する準備が整います。

#### 既存のAFXクラスタでAIDEを設定

既存のAFXストレージシステムにデータ計算ノードのクラスタを追加する場合は、DCNを設定してAIDEクラスタのセットアップを完了する必要があります。

#### データコンピューティングノードをクラスタに追加する

ONTAP は、クラスタ ネットワークに接続された新しいデータ コンピューティング ノードを動的に検出し、System Manager に表示します。"[データ計算ノードをAIDEクラスタに追加する](#)"できます。また、以下の点にも注意してください：

- データ コンピューティング ノードは、クラスタ スイッチに物理的に接続されている必要があります。AIDE クラスタの作成時には、正確に 3 つの DCN が必要です。
- DCN ネットワークおよびその他の必要なネットワーク構成情報に対して十分な IP アドレスが必要です。

#### AIDE クラスタのセットアップを完了する

データコンピューティングノードを統合し、ネットワークの初期構成を完了したら、新しいAIDEクラスタのセットアップを完了できます。

## タスク概要

ページの右側に、必要な構成手順を示すようこそセットアップ ガイダンス ウィンドウが表示されます。すでに完了している各ステップまたはアクション項目の左側にチェックマークが表示されます。まだ完了していない項目には矢印 (→) が表示されます。未完了のアクションごとに次の操作を実行します：

- 有効なタイトル リンクを選択し、環境のセットアップ アクションを完了します。
- リンクが無効になっている場合は、タイトルの上にマウスを置くと、リンクを有効にするために必要なアクションが表示され、セットアップ アクションを続行できます。

各ステップには、可能な場合は追加のドキュメントへのリンクがあります。一部の構成は、AFX クラスターのセットアップの一部として含まれる場合があります。詳細については、"[AFXのドキュメント](#)"を参照してください。

## 手順

1. **\*Configure link aggregation groups and VLANs\***を選択します。

環境に必要な場合は、クラスタレベルの LAG と VLAN を設定できます。このオプションを選択すると、イーサネット ポート ページが起動します。

2. **\*ネットワーク プロトコルの設定\***を選択します。

ストレージ VM (SVM) ネットワーク プロトコルを設定できます。

3. **Update data compute software** を選択します。

データ コンピューティング ソフトウェア イメージが更新可能な場合は、このオプションが有効になります。これを選択すると、データ コンピューティング ソフトウェアの更新ページが起動します。詳細については、"[AI Data Engine softwareを更新](#)" を参照してください。

4. **\*データ エンジン ネットワークの構成\***を選択します。

このオプションを選択すると、データ エンジンのネットワーク アドレスを設定するページが表示されません。

5. **\* クラスタ間ネットワークインターフェイスの作成 \***を選択します

このオプションを選択すると、リモート データをカタログ化する場合、2 段階の手順の一部としてクラスタ間ネットワーク インターフェイスを設定するページが起動します。

6. **\*他の ONTAP ストレージ システムとのピア関係\***を選択します。

クラスタ間ネットワークインターフェイスを作成したら、他の ONTAP ストレージシステムとのピアリング関係を確立できます。

7. **\*データ コンテナの追加\***を選択します。

ボリュームを追加して AIDE で使用し、特定の SVM に関連付けることができます。

8. **\*Add Workspaces\***を選択します。

これにより、Data Engineワークスペースページが起動し、"[ワークスペースを作成する](#)"。

9. \*OpenID Connect の設定\*を選択します。

"[OpenID認証を設定する](#)"して、AIDE Consoleへのアクセスを有効にする必要があります。

#### 終了後の操作

すべてのアクション項目が完了すると、ガイダンス ウィンドウは非表示になります。必要に応じて手動で閉じて再度開くことができます。

## 2.AIDE 構成を完了します

AIDE クラスターのセットアップが完了し、AFX ストレージシステムと統合されたら、AIDE の設定を完了する必要があります。

#### 手順

1. "[必要なAIDEライセンスを追加](#)".

ベクトル化機能やData Guardrails機能を含む、完全なAIDE機能を使用するためにインストールする必要があるAIDEライセンスがあります。

2. "[OIDC 認証の設定](#)".

すべてのデプロイメントに対してOIDCを設定してください。AIDE Consoleにアクセスするには、OIDCを設定する必要があります。

## AFXシステムにAIDEライセンスをインストールする

NetAppライセンスをインストールして、AI Data Engine (AIDE) のすべての機能にアクセスする必要があります。ONTAPクラスター管理者は、ONTAP System Managerを使用してライセンス管理を実行できます。

### AI Data Engineのライセンス取得の準備

ライセンスは、1つ以上のソフトウェア権限の記録です。すべてのAFXライセンスは、NetApp ライセンスファイル (NLF) として提供されます。これは、複数の機能を有効にする単一のファイルです。AIDE をサポートするためにAFX ストレージシステムにライセンスをインストールする前に、考慮すべき点があります。

#### ライセンスの種類

AFX ストレージ システムを通じて AIDE の使用を開始するために必要なライセンスは主に 2 種類あります。

### ONTAP Oneライセンス

Metadata Engine Basicライセンスは通常、ONTAP Oneライセンスの一部として工場出荷時にインストールされます。これにより、AIDE操作の基本的な基盤を提供するMetadata Engine機能へのアクセスが可能になります。ONTAPシステムを管理するために必要なすべてのコア機能も含まれています。

### AI Data Engine

AIDEの全機能を有効化するために必要なプレミアムサービスにアクセスするには、AIDEライセンスを購入してインストールする必要があります。このライセンスにより、データ処理ノードのロックが解除され、ベクトル化、ガバナンスガードレール、推論、統合されたUIエクスペリエンスなどのAI機能が利用可能になります。

ライセンスには、GPU数と有効期限が含まれています。

#### ライセンスのインストール要件

AIDE ライセンスを購入し、関連する NLF ファイルをローカル システムにダウンロードする必要があります。その後、System Manager を通じてファイルを AFX ストレージ システムにアップロードできます。また、以下のものも確認してください：

- ONTAP System Managerにサインインするための管理者の資格情報
- ONTAP 9.18.1以降を実行しているAFXクラスター

#### AFXシステムにライセンスをインストールする

AFX ストレージ システムに必要な追加の AIDE 機能を有効にするために、AIDE ライセンスをインストールできます。

#### 手順

1. System Manager で、\* Cluster \* を選択し、\* Settings \* を選択します。
2. \*ライセンス\*の横にある → を選択します。
3. \*機能\*タブを選択すると、利用可能なONTAP機能が表示されます。
4. ライセンスをインストールするには、\*インストールされたライセンス\*タブを選択します。
5. **+ Add** を選択します。
6. ローカル ライセンス ファイルを選択し、追加 を選択します。

#### 関連情報

- ["ONTAPライセンスの概要"](#)
- ["NetApp Support SiteからNLFライセンスをダウンロードする方法"](#)
- ["ONTAP CLI : system license add"](#)

## ONTAP の AIDE に OpenID Connect を設定する

ONTAP クラスター管理者は、ONTAP System Manager を使用して、AI Data Engine (AIDE) クラスターの OpenID Connect (OIDC) 認証を設定できます。これにより、外部 ID プロバイダー (IdP) を介した安全で集中化されたログインが実現します。



AIDE Console にアクセスするには、OIDC を設定する必要があります。設定されている場合、すべての認証は OIDC を経由して行われます。OIDC が設定されていない場合、管理者だけでなくデータ エンジニアやデータ サイエンティストもコンソールにアクセスできなくなります。この場合、System Manager へのサインインはローカル認証に切り替わります。

AIDEアクセスのOIDC設定については、以下の点に注意してください：

- 大まかに言うと、OIDCの設定は2つの主要なステップから構成されます。まず、システムマネージャを使用して基本的なOIDC構成を実行する必要があります。その後、ONTAP CLIを使用して外部ロールマッピングを構成することでセットアップを完了できます。

- 既存の OIDC 構成を変更することはできません。変更が必要な場合は、まず構成を削除し、希望する設定で新しい構成を作成します。
- OIDCを無効化または削除すると、System ManagerはローカルのONTAPユーザ認証に戻ります。

## OIDCの概要

OpenID Connect (OIDC) は、OAuth 2.0 フレームワーク上に構築された認証プロトコルです。これは、主に認可に使用される OAuth 2.0 を拡張し、アイデンティティ レイヤーを追加します。OIDC では、認証イベントとユーザーの ID に関するクレームを含む JSON Web Token (JWT) である ID トークンの概念が導入されています。

AFXでAIDEとともにサポートされている外部IDプロバイダー (IdP) を選択して設定する必要があります。IdPはユーザーを認証し、AFXがSystem Managerを介してAIDE Consoleへのアクセスを許可するために使用できるトークンを発行します。

### サードパーティの ID プロバイダを設定する

OIDC を使用して認証するには、まず外部 IdP を構成する必要があります。ONTAP の OIDC 実装では、トークン内のロール要求を使用して RBAC を適用します。IdP を設定するときは、ID トークンとアクセス トークンでロール要求を返すように構成されていることを確認してください。ONTAP は OIDC 認証用に 2 つの IdPs をサポートしています (Entra ID および Active Directory フェデレーション サービス (AD FS) ) 。

### Entra ID

Entra IDは以下の概要手順を使用して設定できます：

1. Entra ID 設定ページで新しいアプリ登録を作成します。
2. リダイレクトURI (Web) の値を `https://\$CLUSTER\_MGMT\_IP/oidc/callback` に設定します。適切なクラス管理IPアドレスまたはFQDNに置き換えてください。
3. 「アプリ ロール」 で必要なロールを作成し、ユーザに割り当てます。
4. Token Configuration でトークン要求を更新して、id-token と access-token のロールを返します。

`https://learn.microsoft.com/en-us/power-pages/security/authentication/openid-settings["Entra IDでOpenID Connectプロバイダを設定する"]` 詳細については、こちらを参照してください。

### Active Directory フェデレーション サービス

AD FS は次の概要手順を使用して設定できます。

1. 新しい Application Group を作成し、\* Server application accessing a web API \* を選択します。
2. リダイレクトURI (Web) の値を `https://\$CLUSTER\_MGMT\_IP/oidc/callback` に設定します。適切なクラス管理IPアドレスまたはFQDNに置き換えてください。
3. トークン内のロールを返すようにクレームを設定します。

<https://learn.microsoft.com/en-us/azure/active-directory-b2c/identity-provider-adfs?pivots=b2c-user-flow>["AD FS を OpenID Connect ID プロバイダとして追加する"]の詳細については、こちらを参照してください。

## System ManagerでOIDCを設定する

IdP を設定した後、System Manager で OIDC 認証を設定して、AIDE Console への安全なアクセスを有効にできます。



メタデータURIを指定することで、System ManagerのOIDC構成フィールドを自動的に入力できます。正確なURI形式については、ご利用のIdPのドキュメントをご確認ください。

### 開始する前に

- System Manager への管理者アクセス権が必要です。
- OIDC ID プロバイダを設定し、アクセスできるようにする必要があります。

### 手順

1. System Manager で、\* Cluster \* を選択し、次に \* Settings \* を選択して、OpenID Connect カードを見つけます。
2. OIDC がすでに構成されている場合は、構成を編集または無効にすることができます。OIDC が設定されていない場合は、を選択してセットアッププロセスを開始します。
3. OpenID Connect の設定で、次のフィールドに値を入力します。
  - プロバイダ
  - Issuer
  - JSON Web キーセット URI
  - 認可エンドポイント
  - トークンエンドポイント
  - セッション終了エンドポイント
  - アクセストークン発行者 (オプション)
4. [クライアント構成] で、次のフィールドに値を入力します：
  - クライアント ID
  - Remote user claim
  - 更新間隔
5. 「接続の詳細」で、次のフィールドに値を入力します：
  - クラスタ IP アドレスまたは FQDN
  - 送信プロキシ (オプション)
6. 外部ロールマッピングで、既存のロールマッピングを選択するか、ONTAP `admin` ユーザの新しいロールを定義します。
7. \*今すぐ有効にする\*を選択し、\*保存\*を選択します。System Managerが更新され、新しい認証設定が適用

されます。

8. IdP クレデンシャルを使用してログインします。認証が成功すると、System Manager に戻ります。

#### 終了後の操作

外部ロールマッピングを設定して、OIDC のセットアップを完了します。

#### CLIを使用して外部ロールマッピングを設定する

外部ロールマッピングは、外部 IdP ロールを対応する ONTAP ロールにマッピングできる ONTAP 機能です。OIDC を介して認証するユーザーに ONTAP で適切な RBAC 権限が付与されるように、このマッピングを構成する必要があります。

#### タスク概要

このタスクは、データエンジニアとデータサイエンティストを対応する ONTAP ロールにマッピングします。ご使用の環境に合わせて、コマンド例を適切なロール名に更新する必要があります。System Manager での基本的な OIDC 設定時に、ストレージ管理者ロールを ONTAP `admin` ロールに既にマッピングしている必要があることに注意してください。

#### 手順

1. SSH を使用して、管理者権限を持つアカウントで ONTAP CLI にサインインします。
2. データエンジニアロールのロールマッピングを設定します。例：

```
security login external-role-mapping create -external-role dataEngineer  
-provider entra -ontap-role data-engineer
```

3. データサイエンティストロールのロールマッピングを設定します。例：

```
security login external-role-mapping create -external-role dataScientist  
-provider entra -ontap-role data-scientist
```

#### 関連情報

- ["ONTAP での OAuth 2.0 外部ロールマッピング"](#)
- ["OpenID Foundation"](#)

# ワークスペースを設定する

## AI Data Engineのデータを準備する

クラスタを作成したら、AI Data Engine (AIDE) で使用する予定のデータが含まれているデータ コンテナを確立します。このデータ コンテナは、ONTAPボリューム（ローカルボリュームまたはONTAP 9を実行しているピアリングされたONTAPクラスタのボリューム）である必要があります。

ONTAPクラスタデータをAIDEに手動でアップロードするのではなく、対象のクラスタとSVMをAIDEクラスタとピアリングし、AIDEメタデータ カタログで使用するNFSボリュームを決定する必要があります。データ コンテナを作成したら、ワークスペースを作成し、そのワークスペースにデータ コンテナを関連付けることができます。そのワークスペースのユーザーは、AIワークロードのワークスペース関連のデータ コレクションとリソースにアクセスして操作できるようになります。

### タスク概要

AIDE で使用するデータを含む各 SVM をピアリングする必要があります。クラスタのみをピアリングするだけでは不十分です。これにより、AIDE が意図したとおりにデータにアクセスし、オンボードし、インデックスを作成できるようになります。

SnapMirrorデスティネーションとして機能するAIDEクラスター内のSVMをピアリングする必要があります。

ONTAPクラスタとAIDEクラスタの間にSnapMirror関係を作成する必要はありません。これらの関係は、ワークスペースの作成時に自動的に作成されます。

### 開始する前に

- クラスタと SVM をピアリングし、データ コンテナを選択するには、ストレージ管理者 権限が必要です。
- AIDEで使用するデータが含まれているONTAPクラスタとSVMを特定しました。
- データ ソース ボリュームが次の要件を満たしていることを確認しました：
  - ボリュームはオンラインでアクセス可能です。
  - NFSプロトコルが有効になっています。NFS対応ボリュームのみがAIDEのデータコンテナとしてサポートされます。SMBおよびCIFSボリュームはサポートされていません。
  - ボリュームはFlexCacheボリュームではありません。
  - データ ソースは読み取り/書き込みボリュームです。データ保護ボリュームはサポートされていません。

### 手順

1. ["各ONTAPクラスタとSVMをピア接続"](#)AIDEで使用したいデータが含まれています。
2. ["AIDE で使用するボリュームを選択してください"](#)。

各ボリュームについて、次の情報をメモします：

- ボリューム名
- UUID
- SVM名とUUID

- クラスタ名とUUID

次の手順

"ワークスペースを作成する"作成したデータ コンテナをワークスペースに関連付けます。

関連情報

- "データ移行オプション"

## AI Data Engineでワークスペースを作成する

クラスターをセットアップしたら、ワークスペースを作成できます。ワークスペースを使用すると、クラスター上のデータをセグメント化し、個人のデータアクセスを制御し、AI Data Engine (AIDE) がアクセスすべきでないデータを除外できます。

ストレージを管理する場合は、ONTAP System Managerを使用してワークスペースを作成および管理します。

組織は、チーム、プロジェクト、データの機密性レベル、またはその他の関連基準に基づいてワークスペースを作成します。例えば、医療関係の仕事をしているなら、医療データをワークスペースに分類しますが、IT、法務、その他の部門に関するデータは除外します。

タスク概要

システム処理の制限はワークスペースの作成に影響します（通常、クラスターあたり 1 日あたり最大 15 GB）。複数のワークスペースを並行して、または立て続けに作成すると、各ワークスペースの処理に時間がかかり、大幅な遅延が発生する可能性があります。

ワークスペース インベントリ ページからワークスペース作成のステータスを監視します。最良の結果を得るには、これらの機能にすぐにアクセスする必要がある場合は、一度に多くのワークスペースを作成しないようにしてください。

開始する前に

- ワークスペースを作成し、データ コレクションに関連付けるには、*storage administrator* 権限が必要です。
- ワークスペースおよびAIDEで使用するリモート（ピアリング）データ ソースとローカル データ ソースを決定しました。
- ローカル ボリュームやピア クラスターのボリュームなど、ワークスペースが使用できる"[少なくとも1つのデータ コンテナを作成済み](#)"。



ワークスペースの想定される有効期間中に削除されないボリュームをワークスペースに追加します。ボリュームをワークスペースに追加した後に削除すると、ワークスペースは失敗状態になります。ワークスペースを確立する前に、ボリュームの長期的な実行可能性を確認してください。

- ボリューム上でNFSが有効になっているが、CIFSが有効になっていないことを確認します。ワークスペースはNFSを使用したボリュームのみをサポートします。CIFS (SMB) のボリュームはサポートされていません。

## ワークスペースを作成する

ワークスペースを作成し、AIDE で使用するデータを含むデータ コンテナを関連付けます。

### 手順

1. ONTAP System Managerで、\*Data Engine > ワークスペース\*に移動します。
2. \*追加\*を選択します。
3. \*ワークスペースの追加\*ダイアログで、ワークスペースに関連付ける利用可能なデータ コンテナを少なくとも1つ選択します。
4. "ピア クラスタ"を設定して、これらのクラスタのデータにワークスペース内からアクセスできるようにします
5. ワークスペースへのユーザーアクセスを設定する場合は、今すぐ設定するか、"[ワークスペースが作成されるまで待つ](#)"。
6. ワークスペースが関連付けられているデータ コンテナと同期して新しいデータまたは更新されたデータを取得する頻度の更新間隔を設定します（たとえば、6時間）。



データの鮮度とシステム パフォーマンスのバランスが取れる間隔を選択します。データ コンテナを複数のワークスペースに追加すると、システムは最も積極的な（最短の）間隔を自動的に使用します。詳細については、[ワークスペースの更新とバージョン管理](#)に関するドキュメントを参照してください。

7. \*Continue\*を選択します。
8. \*ワークスペースの確定\*ダイアログで、ワークスペース名と説明を入力します。
9. ワークスペースを作成するには、\*追加\*を選択します。

### 結果

ワークスペースの作成プロセスは、関連付けられているデータセットとそのファイル数、ファイル サイズ、その他の要因に応じて、完了するまでに数分から数時間かかります。

システムは、すべてのデータソースからメタデータを自動的に抽出し、メタデータカタログに保存します。ユーザーはこのカタログを使用して、プロジェクトに必要なファイルを検索できます。ユーザーをワークスペースに割り当てると、データ エンジニア ユーザーはAI Data Engine Consoleからワークスペースに関連付けられたコンポーネントを設定および操作できるようになります。

新しいワークスペースは、プロセスが完了し状態が `ready` に変わるまで、ワークスペースページに `Creating` の状態で表示されます。

## ワークスペースの詳細を確認する

ワークスペースを作成したら、ワークスペースの詳細を確認します。

### 手順

1. ワークスペースの詳細（合計サイズ、使用されているクラスタ容量の割合、最新のワークスペース更新の日付など）を確認します。
2. ワークスペース名を選択して詳細ページを開きます。
3. [概要] タブでは、関連付けられているデータ コンテナ、ユーザー、アクティビティを含むワークスペース

の詳細を表示します。

## ワークスペースの更新とバージョン管理

ワークスペースを更新するたびに、ワークスペース内のすべてのファイルとオブジェクトの現在の状態をキャプチャする不変のバージョンが作成されます。バージョンには、完全なメタデータ、抽出中に使用されるスナップショットへの参照、および追跡可能性のためのジョブIDが含まれます。これにより、データの系統、再現性、監査がサポートされます。

更新は、設定したスケジュール（6時間ごとなど）に従って実行されるか、手動でトリガーされたときに実行されます。サポートされる最小の更新間隔は1時間、最大は1年です。データ コンテナが複数のワークスペースに含まれている場合、システムはメタデータ抽出のスケジュール設定に最も頻繁で最短期間の更新間隔を使用します。

デフォルトでは、システムは以前のバージョン、現在のバージョン、および次の（進行中の）バージョンを保持します。システムは組織のポリシーに従って古いバージョンを保持し、必要に応じて削除できます。

ワークスペースのすべてのバージョンを一覧表示し、バージョン間の相違点を表示して、追加、変更、または削除されたファイルまたはオブジェクトを識別できます。これにより、時間の経過に伴う変更を追跡し、ワークスペース データの進化を把握できます。

## AI Data Engine ワークスペースへのユーザーアクセスの割り当て

ストレージ管理者として、組織の構造やニーズに応じて、データエンジニア、データサイエンティストなど、ユーザーの役割に基づいてワークスペースにユーザーを割り当てます。ユーザーは認証情報を使用して AI Data Engine (AIDE) Console にログインし、割り当てられたワークスペース内のデータ コンテナ リソースにアクセスします。

ONTAP System Manager を使用すると、AIDE ワークスペースにアクセスできるユーザーを管理できます。ユーザーを追加または削除することで、ワークスペースのデータやアクティビティを閲覧、変更、操作できるユーザーを制御できます。

開始する前に

- ワークスペースへのユーザー アクセスを管理するには、*storage administrator* 権限が必要です。
- ワークスペースが作成され、Workspaces インベントリ内でアクティブになっていることを確認します。
- 関連するすべてのデータ コンテナがワークスペースに追加され、アクセス可能であることを確認します。
- **"クラスタで OIDC が有効化され、設定されていることを確認します"**。IdP から ONTAP ロールへのロール マッピングは、関連する各データ エンジニアおよびデータ サイエンティストの IdP ユーザーまたはグループに対して完了する必要があります。

手順

- ワークスペースにユーザーを追加します：
  - a. ONTAP System Manager で、\*Data Engine > ワークスペース\* に移動します。
  - b. ワークスペース名を選択して詳細ページを開きます。
  - c. 「ユーザー」 タブに移動します。

- d. 追加 ボタンを選択して、ユーザー追加ダイアログを開きます。
  - e. 1人以上のユーザーの詳細を入力します。OIDC ユーザーのコンマ区切りリストとして詳細を入力します。
  - f. ユーザーにワークスペースへのアクセスを許可するには、\*追加\*を選択します。
- ワークスペースからユーザーを削除します：
    - a. ワークスペースの詳細ページの「ユーザー」タブで、削除するユーザーを見つけて選択します。
    - b. 削除 ボタンを選択します。
    - c. ダイアログで削除を確認します。
    - d. システムは直ちにユーザーを削除し、ワークスペースへのアクセスを取り消します。

## 結果

ワークスペースの [ユーザー] タブにリストされているユーザーのみが、ワークスペースのデータとアクティビティにアクセスして操作できます。

# 管理と監視

## クラスタプロセスを監視する

### AI Data Engineシステムとクラスタのステータスの表示

ストレージ管理者は、ONTAP System Managerを使用してダッシュボードにアクセスし、クラスタのステータスを表示できます。これは、AI Data Engine (AIDE) の管理タスクを開始する前、または運用上の問題が疑われる場合に実行すると良い最初のステップです。

開始する前に

- AIDE ONTAP 関連の管理タスクを実行するには、\_ストレージ管理者\_権限が必要です。

ダッシュボードからAIDEの健全性と容量を監視

1. クラスタ管理アドレスを使用して ONTAP System Manager に接続します：

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

2. 管理者アカウントでSign inします。
3. 左側のナビゲーションペインで **Dashboard** を選択します。
4. \*Health\*タイルを確認します。
  - クラスタの全体的な健全性を確認します。
  - **Data compute nodes** の数とステータスを確認します。
  - アラートを確認してください：
    - DCN ノードの問題または接続の問題
    - エラーのあるワークスペースまたはデータコレクション（コレクションの公開エラーなど）
5. \*容量\*タイルを確認します：
  - クラスタの合計容量と使用容量を確認します。
  - AIDEクラスタの場合は、次の点を確認します：
    - AIDE メタデータとアプリケーションボリュームで使用される容量（メタデータStorage VM）
    - ワークスペースとデータコレクションで使用される容量（利用可能な場合）
6. 必要に応じて、\*ネットワーク\*タイルと\*パフォーマンス\*タイルを確認し、AIDEワークロードに影響を与える可能性のあるクラスタ全体の動作（ネットワークの輻輳や保護ラグなど）を把握します。

データDCNの健全性と利用状況を表示

1. ナビゲーション ペインで、**Cluster** を選択し、次に **Overview** を選択します。
2. **Data compute** タブを選択します。

このタブには、クラスタ内のすべてのDCNノードが次の情報とともに表示されます：

- ノード名、モデル、シリアル番号、ソフトウェアバージョン
  - ノード全体の状態
  - CPUとメモリの使用率
  - GPU 使用率（GPU が存在する場合）
  - ノードレベルのエラーインジケーター
3. DCNノードを展開して詳細ビューを開き、次の項目を確認します：
- システムのCPUとメモリの使用状況
  - GPUメモリ使用量
  - 報告されたハードウェアまたはサービスの問題
4. \*クラスタ > 概要\* ページで\*配線\*を選択して、DCNノードがクラスタ スイッチに正しくケーブル接続されているかどうかを確認し、ポートまたはリンクの問題を特定します。

#### ワークスペースとメタデータのフットプリントを監視する

1. ナビゲーション ペインで、**Data engine** を選択し、次に **Workspaces** を選択します。
2. ページ上部のワークスペースの概要を確認します：
  - ワークスペースの数とその状態（例：Processing、Healthy、Error）。
  - ワークスペースの合計サイズ
  - すべてのワークスペースで消費されるクラスタ容量の割合。
3. ワークスペース グリッドを確認します：
  - 重要なワークスペースが\*正常\*な状態を示していることを確認します。
  - ワークスペースのサイズと容量の消費量を確認します。
  - `Error` または長時間実行されている `Processing` 状態のワークスペースを探します。
4. 特定のワークスペースの詳細を確認するには、その名前を選択します：
  - \*概要\* タブで、次のことを確認します：
    - ワークスペースの状態とサイズ
    - 含まれるデータコンテナ（ボリューム）とそのアイテム数
    - 各データソースの最終更新時刻
  - \*Data collections\* タブで、次のことを確認します。
    - そのワークスペースに存在するデータコレクション（データコレクションは System Manager では読み取り専用です）
    - 状態、サイズ、最終更新時刻
  - \*ユーザー\* タブで、どのAIDE Consoleユーザーがアクセス権を持っているかを確認します。

#### メタデータStorage VMとAIDE管理保護の監視

1. ナビゲーション ペインで、**Cluster** を選択し、次に **Storage VMs** を選択します。
2. サブタイプ `data-engine` を持つStorage VMを特定します（メタデータSVM）：

- メタデータ SVM がオンラインであることを確認します。
  - 必要に応じて詳細を開いて、次の数を確認します：
    - ボリューム
    - タイプが `Data compute network` の LIF (DCN-ONTAP通信に使用)
3. ワークスペースで使用されるリモートデータソースの保護を表示するには、\* Protection を選択してから Relationships \*を選択します：
- 命名パターンによってAIDEが作成したSnapMirror関係を識別：
    - 宛先ボリューム： <source\_volume\_name>\_dest\_<source\_volume\_UUID>
    - ポリシー： <source\_volume\_name>\_dest\_aide\_policy\_<source\_volume\_UUID>
  - このビューを使用して、関係が正常であること、および遅延時間がワークスペースの更新の期待値と一致していることを確認します。



メタデータStorage VM、AIDEで作成されたSnapMirror関係、またはAIDE管理スナップショット（またはそのスケジュール）をONTAPで直接変更しないでください。変更するとAIDEバージョン履歴が破損する可能性があります。更新動作を調整する必要がある場合は["ワークスペースの更新設定を調整する"](#)。

#### AIDE関連のアラートと通知を確認する

1. ナビゲーション ペインで、**Events & Jobs** を選択し、次に **System alerts** を選択します。
2. 以下に関連するアクティブなアラートを確認します：
  - DCNノードの健全性または接続性
  - データエンジンのネットワークの問題
  - ワークスペースまたはデータ収集エラー
  - ONTAP と DCN クラスタ間のソフトウェアバージョンの不一致
3. 必要に応じて、**Cluster > Settings > Notification management** で通知先（メール、syslog など）を設定し、AIDE 関連のアラートが運用ツールに転送されるようにします。

#### 関連情報

["AFXストレージシステムの管理を準備する"](#)

#### インサイトを表示してAI Data Engineシステムを最適化します

ストレージ管理者は、ONTAP System Managerの `_Insights_` 機能を使用して、NetAppのベストプラクティスに沿った推奨される設定の更新を表示できます。これらの変更により、AI Data Engine (AIDE) クラスタのセキュリティとパフォーマンスを最適化できます。

#### タスク概要

各分析情報はページ上で個別のタイルまたはカードとして表示され、実装するか無視するかを選択できます。特定のテクノロジーについて詳しく知るには、関連するドキュメント リンクを選択することもできます。

#### 手順

1. System Manager で、\* 分析 \* を選択し、次に \* Insights \* を選択します。
2. 利用可能な推奨事項を確認します。

次の手順

推奨されるアクションのいずれかを実行して、構成のベスト プラクティスを実装します。

## AI Data Engine システムイベント、ジョブ、および監査ログの表示

ストレージ管理者は、AI Data Engine (AIDE) によって生成されたイベント、ジョブ、監査ログ メッセージを確認して、内部処理を追跡し、潜在的な問題を診断できます。AIDE システムは、この情報を他の関連データとともに転送して、追加処理とアーカイブを行うように設定できます。

開始する前に

- AIDE ONTAP 関連の管理タスクを実行するには、\_ストレージ管理者\_権限が必要です。

### AIDE のアクティビティ、イベント、ジョブを監視する

集中化された アクティビティ ビューを使用して、すべてのワークスペースにわたる AIDE 固有のイベントとジョブを監視したり、個々のワークスペースに限定されたアクティビティを確認したりできます。

クラスター全体のAIDE活動を表示

ワークスペースの操作を監視し、メタデータ抽出の問題をトラブルシューティングし、AIDE 導入全体でデータ収集の公開を追跡します。

1. ONTAP System Managerで、ナビゲーション ペインで **Data engine** を選択し、次に **Activity** を選択します。
2. \*Events\* タブを選択します。
  - 次のような最近の AIDE 固有のイベントを確認します：
    - ワークスペースの作成、更新、または削除
    - データコンテナの追加/削除操作
    - データ収集の公開（存在する場合）
  - フィルター（重大度、オブジェクトタイプ、ワークスペース、または時間範囲別）を使用して、アクティビティまたは重要なイベントに焦点を絞ります。
3. 個々のイベントを選択して開き、確認します：
  - 概要およびタイムスタンプ
  - 影響を受けるワークスペース、データ収集、またはデータソース
  - 推奨されるアクション（提供されている場合）
4. \*ジョブ\* タブを選択します：
  - 次のような長時間実行ジョブを監視します：
    - ワークスペースの初期メタデータ抽出
    - ワークスペースの更新/カタログ更新ジョブ

- データ収集の公開または更新ジョブ
  - ジョブのステータスと進行状況を確認します。
- 5. ジョブを選択してピークビューを開き、確認します：
  - 開始時間と終了時間
  - 進捗率とフェーズ（例：**Scanning**、**Publishing**）
  - 影響を受けるワークスペース、データ収集、またはデータソース
  - 失敗したジョブのエラーメッセージ

ワークスペース固有のアクティビティを表示する

特定のワークスペースのトラブルシューティングを行うには、ワークスペースの詳細（**Data engine > Workspaces**、ワークスペースを選択）を開き、**Activity** タブを使用します：

- そのワークスペースのみを対象とするイベントとジョブを確認します。
- このビューを使用して、単一のワークスペースが `Processing` 状態で停止しているなどの問題を特定します。

クラスター全体のイベントを表示する

システム アクティビティの貴重な記録としてイベント メッセージを確認します。各イベントには、説明と一意の識別子、および推奨されるアクションが含まれます。

1. ONTAP System Managerで、\*イベントとジョブ\*を選択し、次に\*イベント\*を選択します。
2. 自動更新を有効にするなど、ページの上部にある推奨アクションを確認して対応します。
3. **Events log** タブを選択して、メッセージのリストを表示します。
4. イベントメッセージを選択すると、シーケンス番号、説明、イベント、推奨アクションなどの詳細を確認できます。
5. オプションで **Active IQ suggestions** タブを選択し、Active IQ に登録してクラスターの詳細なリスク情報を取得します。

クラスター全体のジョブを表示する

AIDEクラスターで実行中のすべてのジョブ（AIDE固有のジョブと一般的なONTAPジョブを含む）を表示します。

1. ONTAP System Managerで、\*イベントとジョブ\*を選択し、\*ジョブ\*を選択します。
2. 必要に応じて、表示をカスタマイズしたり、ジョブ情報を検索したりダウンロードしたりできます。

監査ログを表示

監査ログを使用して、HTTP などのアクセス プロトコルの使用に基づいてシステム アクティビティの記録を確認します。

1. ONTAP System Managerで、\*イベントとジョブ\*を選択し、次に\*監査ログ\*を選択します。
2. 追跡される操作を有効または無効にするには、\*設定\*を選択します。

## 通知の管理

AIDE イベントと監査ログを自動的に転送するように通知先を設定します。

### 手順

1. ONTAP System Managerで、\*Cluster\*を選択し、次に\*Settings\*を選択します。
2. \*通知管理\*に移動して  を選択します。
3. AIDE で使用される送信先を表示または設定するには、適切なアクションを選択します：
  - a. イベントの送信先：\*[View event destinations]\*を選択します
  - b. 監査ログの保存先：監査ログの保存先を表示 を選択します。
4. 必要に応じて\*追加\*を選択し、宛先情報を入力します。
5. \*保存\*を選択します。

### 関連情報

- ["ONTAP イベント、パフォーマンス、ヘルス監視"](#)

## AI Data Engineワークスペースを管理

ワークスペースとは、AI Data Engine (AIDE) が特定のプロジェクトまたはユースケースのメタデータカタログを構築および更新するために使用するデータソース（ボリューム）のセットです。ストレージ管理者は、ONTAP System Managerを使用して、ワークスペースの状態を監視し、構成を調整し、データソースを制御し、ユーザーを管理し、不要になったワークスペースを削除できます。

### 開始する前に

- ワークスペースを管理するには、*storage administrator* 権限が必要です。

### ワークスペースのステータスを確認する

ワークスペースの健全性、容量の使用状況、メタデータのステータスを確認して、Metadata Engineが期待どおりに動作し、予期しないリソースを消費していないことを確認します。

### 手順

1. ONTAP System Managerのナビゲーション ペインで、\*Data engine > Workspaces\*を選択します。
2. ページ上部にある概要で、ワークスペースの合計数、ワークスペース全体の健全性、容量の使用状況を確認します。
3. ワークスペース固有の情報については、ワークスペース名を選択してください。\*概要\*タブで次のことを確認します：
  - ワークスペースの状態とサイズ。
  - ワークスペースに含まれるデータコンテナ（ボリューム）。
  - データソースごとのアイテム数と最終更新時刻。
  - ワークスペースレベルの警告。



ワークスペースまたはデータコレクションにエラー状態が表示される場合は、すべてのソースボリュームがオンラインでアクセス可能であることを確認します。

4. \*データ収集\*タブを選択すると、次の情報が表示されます：

- このワークスペースに関連付けられているすべてのデータコレクション。
- 状態（`Published`または`Error`など）、サイズ、最終更新時刻。



System Managerは、データ コレクションに対して読み取り専用です。データ エンジニアは、AI Data Engine Consoleでデータ コレクションを作成および管理できます。

5. \*ユーザー\*タブを選択して表示します：

- このワークスペースにアクセスできるユーザーのリスト。

6. このワークスペースに関連するイベントとジョブのみを表示するには、アクティビティ タブを選択します。

## ワークスペースのプロパティと更新スケジュールを編集する

ワークスペースの名前、概要、更新間隔、および（ライセンスがある場合）ガードレール ポリシーを調整できます。

### 手順

1. \*Data engine > Workspaces\*から、ワークスペースの横にある  を選択し、\*Edit\*を選択します。
2. ワークスペースのプロパティを編集：
  - 必要に応じて、\*名前\*と\*概要\*を更新します。
  - 更新間隔（メタデータの更新頻度）を許容範囲（時間と日数）内で調整します。
  - AIDE ライセンスがインストールされている場合は、\*ガードレール ポリシー\*を選択できます。
3. \*保存\*を選択します。



更新間隔またはメタデータ処理の変更は、このワークスペースのリモートSnapMirror関係が更新される頻度に影響する可能性があります。

## 既存のワークスペースにデータ コンテナを追加する

追加のマウントされたボリューム（ローカルまたはピアリングされたリモート クラスタから）を追加して、そのメタデータがワークスペース カタログに含まれるようにすることができます。

1. **Data engine > Workspaces** から、次のいずれかを実行します：
  - ワークスペースの横にある  を選択し、\*データ コンテナの追加\*を選択します。
  - ワークスペースを開き、[概要] タブを選択して、データ コンテナ セクションで [追加] を選択します。
2. データ コンテナをワークスペースに追加 ダイアログで次の操作を行います。
  - AIDE クラスタ上のローカルボリュームを特定します。
  - ピアリングされたクラスターを展開してリモート ボリュームを選択します（リモート ボリュームには

クラスターと SVM のピアリングが必要です)。



選択できるのは、グローバルに除外されておらず、まだワークスペースの一部ではない、適格なオンライン ボリュームのみです。

3. リモート ボリューム マッピングのプロンプトが表示された場合：

- 選択したリモート ボリュームのSnapMirrorデスティネーションを受信するAIDEクラスター上のターゲットStorage VMを選択します。

4. \*追加\*を選択します。

5. ワークスペースの\*アクティビティ\*タブまたは\*データ エンジン > アクティビティ\*を使用して、メタデータ抽出と新しいデータ ソースのSnapMirror初期化を追跡します。

## ワークスペースからデータ コンテナを削除する

データ コンテナがワークスペースの目的に関連しなくなった場合、またはそのワークスペースのメタデータ管理の範囲を縮小する場合は、データ コンテナを削除することができます。データ コンテナを削除すると、そのボリュームのメタデータの更新が停止され、メタデータ カタログからそのメタデータが削除されます。



ワークスペースに追加されたONTAPのソースボリュームを削除しないでください。ボリュームを削除すると、ワークスペースは障害状態になります。基盤となるONTAPボリュームを削除する前に、必ずワークスペースからデータ コンテナを削除してください。

### 手順

1. **Data engine > Workspaces** に移動し、データ コンテナを含むワークスペースを選択します。
2. \*概要\*タブで、削除するデータ コンテナを見つけます。
3. データ コンテナの横にある\*削除\*を選択します。
4. 確認ダイアログを確認し、\*削除\*を選択します。



ワークスペースからデータ コンテナを削除しても、基盤となるONTAPボリュームまたはそのSnapMirror関係は削除されません。AIDE内のメタデータの使用にのみ影響します。

## ワークスペース ユーザーの管理

データ エンジニアおよびデータ サイエンティストのユーザーにワークスペースへのアクセスを許可または取り消すことができます。これらのユーザーはアイデンティティプロバイダ (OIDC) で定義され、ONTAP ロールにマッピングされます。ユーザー アクセスを管理する方法については、"[ユーザーをワークスペースに割り当てる](#)"のドキュメントを参照してください。

## ワークスペースを削除する

ワークスペースを削除すると、ワークスペースの定義と関連付けられたAIDEメタデータが削除されます。ワークスペースに関連付けられたデータ コレクションとベクトル埋め込みも削除されます。



基礎となるONTAPデータ (ボリューム、SnapMirror関係) は削除されません。

### 手順

1. **Data engine > Workspaces** で、次のいずれかを実行します。
  - 単一のワークスペースを削除するには、を選択し、\*削除\*を選択します。
  - 複数のワークスペースを削除するには、ワークスペースのチェックボックスをオンにして、削除 を選択します。
2. 確認ダイアログで、続行する前にアクションの影響を確認します：
  - ワークスペースのメタデータは完全に削除されます。
  - ワークスペースに関連付けられたデータ コレクションと埋め込みは完全に削除されます。

 ソフト削除または復元オプションはありません。
3. 理解したことを確認するチェックボックスを選択し、\*削除\*を選択します。

#### 関連情報

- ["ユーザーをワークスペースに割り当てる"](#)

## AIDEシステムのアップグレードとメンテナンス

### AI Data Engineシステムのアップデートと互換性

AI Data Engine (AIDE) システム コンポーネントを最新の状態に維持して、最適なパフォーマンスを維持し、新機能にアクセスできるようにします。デプロイメント後、新しいソフトウェアまたはファームウェアが利用可能になったとき、ノードを追加または交換したとき、または機能更新のために定期的に、コンポーネントを更新します。

#### AIDE システム コンポーネント

AIDE システムでアップデートが必要な主なコンポーネントは 2 つあります：ONTAP ソフトウェアと AIDE ソフトウェア (DCN ファームウェアを含む)。

#### ONTAPソフトウェア

ONTAPは、AIDEデプロイメントで使用されるものを含むNetAppストレージシステムで実行されるオペレーティングシステムです。システムの安定性、セキュリティ、およびAIDEコンポーネントとの互換性を維持するために、ONTAPを最新の状態に保ってください。AIDEコンポーネントは個別に更新されます。

#### AIDE ソフトウェア アップデート

AI Data Engine ConsoleソフトウェアとDCNファームウェアのアップデートは、単一のパッケージ (.tgz) として配布され、ONTAPイメージには組み込まれていません。アップデートにより、AI Data Engineシステム内のハードウェアコンポーネントが適切に機能することが保証され、新機能、パフォーマンスの向上、バグ修正が提供されます。

#### 更新プロセスを理解する

AIDE ソフトウェア アップデートは ONTAP System Manager を通じて管理できます。

AIDEはONTAP自動ソフトウェア更新機能をサポートしていません。"[NetApp Support Site ダウンロード](#)"から

通知を受け取るにはサインアップできますが、AIDEソフトウェアのすべての更新は管理者によって手動で実行されます。

リリースの種類と範囲：

- ONTAPメジャーリリース (9.x.x) とAIDEメジャーリリース (9.x.x U0) では、ONTAP統合に影響する新しい機能、API、または変更が導入されます。
- ONTAPパッチリリース (9.x.x Px) とAIDEアップデートリリース (9.x.x Ux) には、ONTAP統合に影響しない修正プログラムとアップデートが含まれています。

## 互換性マトリックス

アップデートを計画する際は、ONTAPとAI Data Engine softwareの互換性を確認してください。

AIDE ソフトウェアは「U」リリースとしてリリースされます。AIDE メジャー リリースは「U0」リリースであり、後続のマイナー リリースは「U1」以降になります。

### ONTAPとAI Data Engineの互換性

AIDEリリース	サポートされている ONTAP バージョン
9.18.1 U0	9.18.1 GA およびすべての 9.18.1 Px
9.18.1 U1 以降	9.18.1 GA およびすべての 9.18.1 Px



「Px」は、メジャー バージョンのすべてのONTAPパッチ リリースを表します（例：9.18.1 P1、9.18.2など）。

### AIDE アップグレード パス

以下の例では、仮想的な将来のバージョンを使用して、AIDE 9.18.1 U0および9.18.1 U1から許可されているアップグレードパスとアップデートパスを示しています。

現在の AIDE リリースが...	対象となるAIDEリリースは...	アップグレードまたは更新パスは...
9.18.1 U0	9.18.1 U1	直接
9.18.1 U0	9.18.1 U3	直接 (9.18.1 では、任意の Ux から任意の以降の Ux にアップデートできます)
9.18.1 U1	9.18.1 U3	直接 (9.18.1 では、任意の Ux から任意の以降の Ux にアップデートできます)

### 元に戻す際の制限事項

AIDEシステムは、DCNファームウェア、AIDEソフトウェアアップデート、またはAFXストレージシステム上のONTAPのリバート処理をサポートしていません。アップデートまたはアップグレードをインストールした後は、以前のバージョンに戻すことはできません。アップデートまたはアップグレードする前に[互換性要件](#)を確認してください。

### 関連情報

- ["AFXシステムのONTAPソフトウェアをアップグレードする"](#)

- ["AI Data Engine softwareを更新"](#)

## AI Data Engine softwareのアップデート

ストレージ管理者は、ONTAP System Managerを使用して、AI Data Engine (AIDE) ソフトウェア、data compute node (DCN) ファームウェア、およびその他のシステム ファイルをAIDEシステムで更新できます。

AIDEはONTAP自動ソフトウェア更新機能をサポートしていません。["NetApp Support Site ダウンロード"](#)から通知を受け取るにはサインアップできますが、AIDEソフトウェアのすべての更新は管理者によって手動で実行されます。

### タスク概要

組み合わせたAIDEソフトウェア パッケージは、一般的なONTAPアップデート パッケージよりもかなり大きくなります (AIDEパッケージは約40GB)。AIDEソフトウェアをアップデートする際は、アップロードとインストールに時間がかかることを考慮してください。

### 開始する前に

- DCNファームウェアとAIDEソフトウェアを更新するには、ストレージ管理者権限が必要です。
- アクティブなアカウントのNetApp Support Site のクレデンシャルが必要です。
- ["アップデートを計画する際は、ONTAP、DCNファームウェア、AIDEソフトウェア間の互換性を確保してください"](#)。



DCN ファームウェアまたはAIDE ソフトウェアのアップデートでは、元に戻す機能はサポートされていません。アップデートをインストールすると、以前のバージョンに戻すことはできません。

### 手順

1. ["統合されたDCNファームウェアとAIDEソフトウェア更新ファイルをローカルクライアントにダウンロードします"](#)。
2. System Manager で、**クラスター > 設定 > ソフトウェア更新** を選択します。
3. \*Software updates\*の横にある [→](#) を選択します。
4. AI Data Engine の更新で、**AI Data Engine software** ファイルの追加 を選択し、更新パッケージを選択します。
5. パッケージのアップロードが完了したら、\*更新\*を選択して、DCNノードへの更新のインストールを開始します。



DCN ノードの更新が進行中の場合、AI Data Engine Console は使用できないか、アクセスできません。

### 結果

DCNはAIDEソフトウェアで更新され、更新されたバージョンが各ノードに表示されます。

### 関連情報

- ["AFXシステムのONTAPソフトウェアをアップグレードする"](#)

## データコンピューティングノードをAI Data Engine クラスタに追加

新しいAI Data Engine (AIDE) クラスタの作成時または既存のクラスタの拡張時に、データコンピューティングノード (DCN) を追加できます。ワークフローは、ONTAP System Managerを使用したノードの検出と設定で構成されます。

### ノードを追加する準備

DCNノードを追加するにはいくつかの考慮事項があります。

### 新しいAIDEクラスタを作成する場合

新しいAIDEクラスタには、ちょうど3つのDCNノードが必要です。

### ハードウェアのインストールとアドレス指定

次の前提条件が満たされていることを確認してください：

- 新しいDCNハードウェアがラックに設置され、電源がオンになり、クラスタスイッチにケーブル接続されます。
- DCN から ONTAP バックエンドサブネットに使用可能な IP アドレス空間範囲があります。
- ONTAPクラスタは初期化されており、クラスタ管理LIFからアクセス可能です。

### System Managerのクレデンシャル

AIDE クラスタの作成または拡張タスクを実行するには、\_ストレージ管理者\_権限が必要です。

### ソフトウェアの互換性

以下の文書を確認して、DCNハードウェアとソフトウェアのバージョンがONTAPクラスタと互換性があることを確認してください：

- ["DCNソフトウェアは、ONTAPバージョンと互換性があります"](#)。

ノード追加操作中に、System Manager は新しいノードが次のソフトウェアバージョンと互換性があることを確認します：

- これが初めてのDCN参加の場合、ONTAPクラスタの有効バージョン (ECV) 。
- DCN がすでに存在する場合は、既存の DCN クラスタのバージョン。

ノードに互換性がない場合：

- 影響を受けるDCNの横に\*追加\*ダイアログでエラーが表示されます。
- まず、DCN ソフトウェア (または必要に応じて ONTAP) を互換性のあるバージョンに更新する必要があります。

### データコンピューティングノードを追加する

新しいAIDEクラスタを作成するか、既存のクラスタを拡張する際に、DCNノードを追加します。

### 手順

1. System Managerのナビゲーション ペインで\*Dashboard\*を選択し、次に\*Health\*カードを選択します。

2. 追加するノードがあることを確認し、詳細を表示 を選択してリストを表示します。

リストには、まだAIDEクラスタの一部ではない検出されたノードが含まれています

3. または、クラスタ\*と\*概要、\*データコンピューティング\*タブを選択してリストを表示することもできます。

4. データ計算ページの下部で、ノードのリストの上にある 追加 を選択します。

5. \*データ計算ノードの追加\*ダイアログで、追加したいDCNノードを選択します。

オプションで、個々のノードを追加する前に名前を変更できます。

6. 初めてノードを追加する場合で、バックエンドサブネットが存在しない場合は、\*サブネットの追加\*を選択し、以下を入力します：

- サブネット名（内部使用用）
- サブネットアドレスとマスク
- このバックエンドネットワーク上のDCNおよびONTAPノードのIPアドレス範囲

System Managerは、追加されるすべてのDCNとクラスタ内のすべてのONTAPノード、およびDCNとONTAPの通信に使用される追加のクラスタレベルのフローティングIPに対して、範囲に十分な空きIPアドレスが含まれていることを検証します。

7. バックエンド サブネットを追加したか、すでに存在しているかに関わらず：

a. 利用可能な IP アドレスを確認します。

b. 必要に応じて、サブネットの編集 を選択し、IP 範囲を拡張します。

- 範囲を拡大することしかできません。サブネットの縮小または変更はサポートされていません。
- サブネットまたは IP 範囲を変更するには、DCN 上の基盤となる Kubernetes クラスタの再作成が必要になる場合があります、数分かかることがあります。

8. 必要に応じて、次の情報を指定してData Engineサービスインターフェイスを構成します：

- サービスIPアドレス
- ネットワーク マスク
- ゲートウェイ（環境に必要な場合）

単一のIPアドレスはDCN間で負荷分散され、AIDE Consoleおよび関連APIのフロントエンドアドレスとして使用されます。

9. 選択したノード、バックエンド サブネット、およびData Engineサービス インターフェイスの設定を確認します。

10. \*追加\*を選択し、操作が完了するまで待ちます。System Managerは次のアクションを実行します：

- 選択したノードをDCNクラスタに追加します
- バックエンドネットワークをプロビジョニングし、KubernetesベースのDCNクラスタにノードを参加させます
- DCN検出用の内部メタデータを更新します

11. 完了したら、\*Cluster\*と\*Overview\*を選択して確認します：

- a. \*データ計算\*の下に新しいDCNがクラスターの一部として表示されます
- b. すべてのノードが\*正常\*です
- c. ダッシュボードの\*Health\*カードに更新されたノード数が表示されていることを確認します。

## AI Data Engine クラスタ内のノードを置き換える

AI Data Engine (AIDE) クラスタ内のdata compute node (DCN) が機能しなくなった場合、またはハードウェア障害、アップグレード、メンテナンスのために交換が必要な場合は、交換する必要があります。これにより、AIDEクラスタは正常な状態を維持し、稼働し続けます。この手順は、実行中のサービスを中断することなく実行できます。

### ノードの交換の準備

AI Data Engine クラスタ内のノードを交換する前に、考慮すべき点がいくつかあります。

#### 要件

次のものが必要になります：

- *storage administrator* 権限を持つユーザー認証情報
- 新しい交換ノードのシリアル番号

#### 制限事項

AIDEクラスタ内のノードを交換する際には、以下の制限事項に注意してください：

- ノードの置き換えは、CLIおよびオプションでREST APIを使用する場合にのみサポートされます。
- System Manager を使用してノードの置換を実行することはできません。
- 新しいノードはクラスタのソフトウェアバージョンと一致する必要があります。必要に応じて、ONTAPによって更新されます。
- IP アドレスの競合を避けるため、障害が発生したノードは、クラスタネットワークに接続されている間は電源をオンにしないでください。

## AIDE クラスタ内の DCN ノードを交換する

次の手順を使用して、AIDEクラスタ内のDCNノードを交換できます。

#### 手順

1. 障害が発生したノードを物理的に取り外す

ノードの電源をオフにし、クラスタネットワークから切断します。交換プロセス中は、ノードがネットワーク上で起動されていないことを確認してください。

2. 次のコマンドを使用して、障害が発生したノードをクラスタから削除します：

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

<node\_name>値の実際の名前を入力してください。

3. 新しいノードをクラスタに物理的に接続する

ノードがケーブル接続され、電源がオンになっており、検出可能であることを確認します。

4. 検出可能な未構成のノードを表示して、新しいノードがオンラインであることを確認します：

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

5. 次のコマンドを使用して、ノードをクラスターに追加します：

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

ONTAPは新しいノードにIPアドレスを割り当てます。ノードのソフトウェアバージョンがクラスタと一致しない場合は、ONTAPによってノードが自動的に更新されます。

6. 次のいずれかのコマンドを使用して、クラスタの健全性とノードの統合を確認します：

```
dcn cluster node show
```

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

#### 関連情報

- ["コンピューティングクラスターを拡張する"](#)

# ベクトル化とデータコレクションの管理

## AI Data EngineのData-to-RAGクイック スタート

新しく導入されたAI Data Engine (AIDE) システムを、このワークフローを使用して機能する検索拡張生成 (RAG) エンドポイントに接続します。ストレージ管理者、データエンジニア、データサイエンティストがONTAP System ManagerとAIDE Consoleを使用してどのように連携しているかを理解します。

開始する前に

- データコンピューティングノード (DCN) をインストールしてONTAPクラスタに追加しました。
- AIDE ソフトウェアをベクトル化およびガードレール用にインストールし、ライセンス認証を完了しました。
- "OpenID Connect (OIDC) "を設定し、管理者、データ エンジニア、データ サイエンティストの役割をマッピングしました。

1

データの範囲とガバナンスを定義する

ストレージ管理者またはセキュリティ管理者として、AIDE Console と ONTAP System Manager で環境を準備する必要があります：

- "1つ以上のワークスペースを作成する"ローカルおよびリモートのデータ ソースから。
- AIDE コンソールの"分類子とガードレールポリシーを設定する"。
- "データエンジニアとデータサイエンティストにワークスペースへのアクセスを割り当てる"。

2

ワークスペースのメタデータを確認する

データエンジニアやデータサイエンティストとして、AIDE Consoleを使用してワークスペースのメタデータを探索する場合：

- "ワークスペースのメタデータを確認する"利用可能なコンテンツを理解します。
- RAG に供給するデータの論理サブセットを 1 つ以上定義します (たとえば、サポート記事、製品マニュアル、匿名化された医療データなど) 。

3

データ収集を作成して公開する

データ エンジニアまたはデータ サイエンティストとして、選択したサブセットを RAG 対応のコレクションに変換したいと考えています：

- 選択したフィルターを使用してワークスペースから"データ収集を作成する"。
- "データ収集を公開する"インデックス作成を監視し、`Ready` 状態になるまで待ちます。
- 選択したコレクションの取得エンドポイント URI をコピーし、データ サイエンティストまたはアプリケーション開発者に提供します。

- "データ収集ステータスとベクターフットプリントを表示する"必要に応じて。

次の手順

- "AIDE でデータ資産とガードレール ポリシーを定義する"
- "AIDE Consoleでワークスペースのメタデータを探索する"
- "AIDE Console でデータコレクションを作成する"

## AI Data Engine Consoleでワークスペースのメタデータを探索する

データ エンジニアまたはデータ サイエンティストとして、AI Data Engine (AIDE) での最初のタスクは、ワークスペースで利用可能なデータを理解することです。AIDE Consoleを使用してメタデータ カタログを照会し、関連するファイルを検索し、データ コレクションに変換するデータのサブセットを識別します。

開始する前に

- AIDE Consoleで\_データ エンジニア\_または\_データ サイエンティスト\_の権限と、少なくとも1つのワークスペースへのアクセス権が必要です。
- ストレージ管理者には次の権限があります：
  - ONTAP System Managerで1つ以上のワークスペースを作成しました。
  - 関連するワークスペースへのアクセス権をユーザーまたはグループに割り当てました。
- ワークスペースのメタデータ抽出が完了し、ワークスペースは `Ready` 状態です。
- 分類子が有効になっているため、メタデータには分類タグ (PII インジケータなど) が含まれます。

データエンジニアまたはデータサイエンティストとして **AI Data Engine Console** に **Sign in** します

手順

1. ブラウザで、AIDE Console URL に移動します：

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

2. 組織の OIDC プロバイダーを通じて認証します。
3. 自分の役割がデータ エンジニアまたはデータ サイエンティストとして認識されていることを確認します (たとえば、利用可能なワークスペースやデータ収集アクションなど)。詳細については、"[AIDE ロールのドキュメント](#)"を参照して、データ エンジニアとデータ サイエンティストがAIDE コンポーネントをどのように使用するかを確認してください。

結果

AIDE Consoleにログインしており、アクセスを許可されたワークスペースのみが表示されます。

アクセス可能なワークスペースを表示する

手順

1. AIDE Consoleで、\*Data Curator > ワークスペース\*に移動します。
2. アクセスできるワークスペースのリストを確認します。
3. ワークスペースを選択して詳細を開きます。

#### 結果

これで、ストレージ管理者がプロジェクトで利用できるようにしたデータ資産のワークスペース スコープビューが表示されます。

#### 次の手順

- ["ワークスペースから RAG のデータ コレクションを作成する"](#)

## AI Data Engine Consoleでデータコレクションを作成する

データ コレクションは、AI Data Engine (AIDE) における RAG のコア構成要素です。データ エンジニアまたはデータ サイエンティストは、コレクションに属するファイルを定義し、埋め込みおよびインデックス作成のオプションを設定し、アプリケーションが取得エンドポイントを通じてクエリを実行できるようにコレクションを公開します。

すべてのデータ収集タスクは AIDE Console で実行します。

#### 開始する前に

- AIDE Console(`\https://<cluster\_management\_ip>/console`で`\_data engineer\_`または`\_data scientist\_`の権限が必要です。
- メタデータが抽出され、`Ready`状態になっている少なくとも 1 つのワークスペースにアクセスできます。
- ワークスペースのメタデータを調査し、意味のあるデータのサブセットを定義するクエリまたはフィルターを特定しました。
- AIDEソフトウェアライセンスがインストールされ、推論機能が有効になっています。

### ワークスペースのメタデータからデータコレクションを作成する

#### 手順

1. \*Data Curator > ワークスペース\*に移動し、ターゲット データが含まれているワークスペースを選択します。
2. データ収集の追加 を選択します。
3. 新しいデータ コレクションの作成 ページで、次の操作を行います：
  - a. コレクションの名前と説明を入力します (例: Support\_KB\_RAG\_EN)。
  - b. コレクションを次のいずれかに設定するかを選択します：
    - 動的: 定義したフィルタリング基準に基づいて、新しいファイルが自動的に識別され、データ コレクションに追加されます。これはワークスペースの更新中に発生します。
    - 静的: コレクションに含めるファイルを選択します。データ コレクションが`draft`状態の場合、ファイルを編集できます。データ コレクションが`Published`状態に移行すると、編集できなくなります。

#### 4. ソースサブセットを指定します：

- a. キーワードとフィルター（ファイルの種類、タイムスタンプ、その他の属性）を使用して、含める関連ファイルを検索します。



ファイル名を選択すると、コンテンツのプレビュー ウィンドウが開きます。

#### 5. これらのファイルをデータ コレクションに追加します。

#### 6. コレクションを確定するには、\*保存\*を選択します。

### 結果

データ収集の範囲を定義し、必要なファイルを追加しました。AIDE はコレクションを公開するときに埋め込みを生成し、ベクター インデックスを構築します。



単一の「すべて」のコレクションではなく、小さな焦点を絞ったコレクション（たとえば、ケースまたはドメインごと）を作成します。これにより、検索の関連性と管理性が向上します。

## データ収集を公開する

データ コレクションを公開して、RAG 取得エンドポイントを通じて AI アプリケーションでクエリできるようにします。公開すると、選択したファイルからベクター埋め込みが生成され、セマンティック検索用にインデックスが作成されます。コレクションが `Ready` 状態に達すると、そのエンドポイントは、データ サイエントリストがノートブック、パイプライン、および AI アプリケーションに統合して、検索拡張生成 (RAG) と検索を行うために利用できるようになります。



大規模なコレクションの場合、リソースの競合を最小限に抑えるために、最初の公開と主要な再公開をオフピーク時にスケジュールすることを検討してください。

### 手順

1. **Data Curator > Data collections** に移動し、データコレクションのオプションメニュー (**...**) を選択します。
2. **\*Publish\*** を選択します。
3. デフォルトまたはカスタムの最適化設定を選択します。
4. データ変換を開始するには、**Publish** を選択します。
5. AIDE Consoleで、コレクションの詳細ビュー (**Data Curator > データコレクション**) を開いてステータスの更新を確認します。

### 結果

コレクションは `Ready` 状態になり、下流のアプリケーションやデータ サイエントリストが使用できるようになります。

**Data Curator > データ コレクション** から、**URI** をコピー を選択して、API を使用してデータ コレクションにアクセスするために必要な情報を取得できます。

## データコレクションを更新または削除する

時間が経つにつれて、データコレクションを改良したり廃止したりする必要が生じる場合があります。コレクションを調整するには、フィルタを調整してファイルを追加または削除したり、埋め込み設定を変更したり、コレクションの概要を更新したりします。コレクションを削除すると、完全に削除され、その取得エンドポイントは使用できなくなります。

### データ収集を更新する

データコレクションが `draft` 状態のときに更新できます。

#### 手順

1. \*Data Curator > データ収集\*に移動します。
2. 変更するコレクションを選択します。
3. \*Edit\*を選択します。
4. 次のいずれかを調整します：
  - 名前と説明
  - フィルター（パス、ファイルタイプ、分類タグ）。
  - 埋め込みとチャンクの設定。
5. 変更を保存します。
6. 新しい定義と埋め込みが有効になるように、コレクションを再度公開します。

#### 結果

更新された構成で新しいインデックス作成ジョブが実行され、完了するとコレクションは `Ready` 状態に戻ります。

### コレクションを削除する

コレクションの削除は永続的です。コレクションを削除する前に、実稼働アプリケーションがコレクションの取得エンドポイントに依存していないことを確認してください。

#### 手順

1. \*Data Curator > データ コレクション\*に移動し、コレクションのオプションメニュー（...）を選択します。
2. \*削除\*を選択します。
3. 削除を確認します。

#### 結果

コレクション定義とその埋め込みはAIDEから削除されます。コレクションが削除された後、以前の取得エンドポイントにクエリを実行しようとするアプリケーションは失敗します。

### 次の手順

- ["データコレクションを表示"](#)

# AI Data Engineでデータコレクションを表示

データエンジニアやデータサイエンティストがワークスペースからデータコレクションを作成して公開した後、そのステータス、サイズ、およびAI Data Engine (AIDE) クラスターへの影響を可視化する必要があります。

ストレージ管理者、データエンジニア、データサイエンティストであれば、ONTAP System ManagerとAI Data Engine Consoleでデータコレクションを表示できます。

開始する前に

- データ コレクションを表示するには、ONTAP System Managerの\_ストレージ管理者\_権限、またはAIDE Console(`\https://<cluster\_management\_ip>/console`の\_データ エンジニア\_または\_データ サイエンティスト\_権限が必要です。
- メタデータが正常に抽出されたワークスペースが少なくとも1つ存在します。
- データ エンジニアまたはデータ サイエンティストが、AIDE Consoleから少なくとも1つのデータ コレクションを作成して公開している。
- AIDEソフトウェアライセンスがインストールされ、推論機能が有効になっているため、ベクトル化エンドポイントと検索エンドポイントがアクティブになっています。

## クラスター全体のデータ コレクションを表示する

ストレージ管理者にとって、ONTAP System Manager は、データ コレクションとそのフットプリントのクラスター全体のビューを提供しますが、管理者がそれらを作成または変更することはできません。

手順

1. System Managerで、\*Data Engine > Data collections\*に移動します。
2. ページ上部の在庫概要を確認します：
  - ステータス別のデータ収集の総数
  - すべてのコレクションにわたるベクターデータベースで消費される合計容量
  - クラスター全体の容量に対するベクトル空間の割合
3. 個々のデータ コレクションを選択して確認します：
  - コレクション名と説明
  - UUID
  - 関連付けられたワークスペース
  - ステータス
  - コレクションのサイズ
  - 作成者
  - 最終更新時間

結果

これで、クラスター内のすべてのデータ コレクションとそのストレージへの影響の概要が表示されます。このビューを使用して、大きい、古い、または準備完了状態になっていないコレクションを識別します。

また、個々のデータ コレクションがアクティブに更新されているかどうか、および何らかの障害によって RAG の使用がブロックされているかどうかを確認できます。

## コレクション関連のジョブとイベントを監視する

ストレージ管理者は、クラスター全体の **Activity** ページとワークスペースの詳細から、コレクションを構築および更新するジョブを監視できます。

### 手順

1. System Manager で、**Data Engine** > アクティビティ に移動します。
2. \*イベント\*タブ：
  - a. タイプ（ワークスペース、データ収集など）または重大度でフィルタリングします。
  - b. データ収集に関連するイベント（「データ収集の公開に失敗しました」など）を展開して詳細を表示します。
3. \*ジョブ\*タブで：
  - a. データ収集のインデックス作成と公開ジョブに焦点を絞るためのフィルタ。
  - b. 各ジョブについて、ピークビューを開いて次の内容を確認します：
    - 進捗率。
    - 開始時間と終了時間。
    - 報告されたエラー メッセージまたは警告。
4. 必要に応じて、影響を受けるワークスペース（**Data Engine** > ワークスペース）に戻り、そのアクティビティ タブを開いて、そのワークスペースのみにスコープされたイベントとジョブを表示します。

### 結果

データ収集のライフサイクルを追跡し、停止または失敗したジョブを特定し、コンテキスト情報を収集してデータ エンジニア、データ サイエンティスト、またはサポートに渡すことができます。



データ収集が `Publishing` 状態のまま長時間続く場合は、障害と判断する前に、アクティビティ ページで対応する長時間実行ジョブを確認してください。

## AIDE Console からデータ コレクションを表示

データエンジニアやデータサイエンティストは通常、データコレクションの作成と公開が行われる AIDE Console から直接監視します。

### 手順

1. データ エンジニアまたはデータ サイエンティストとして AIDE Console にログインします。
2. **Data Collections** に移動し、目的のデータ コレクションを選択します。
3. 各コレクションについて：
  - a. 状態を確認する(Draft、Publishing、Ready、または Failed)。
  - b. データ収集名を選択して、定義の詳細（フィルター、含まれるファイルの種類、分類子のオプション、埋め込み設定）を確認します。

- c. 最後の公開または更新のタイムスタンプを検査します。
4. 必要に応じて、ジョブの詳細またはログ（利用可能な場合）を開いて、失敗または不完全な実行を確認します。

#### 結果

データ エンジニアとデータ サイエンティストは、ストレージ管理者を介さずに、ステータスと健全性を監視しながらコレクション定義を反復処理し、再度公開できます。

#### 次の手順

- ["AIDE ConsoleでRAG用のデータコレクションを作成する"](#)

# ガードレールを実装する

## データ資産のためにAI Data Engineでガードレールポリシーを定義する

データまたはプラットフォームの所有者として、AI Data Engine Console を使用して、どのデータがAIの適用範囲内にあるか、どのデータが常に禁止されているか、そのデータが分類および検索拡張生成（RAG）に使用されるときにどのような安全ルールが適用されるかを定義します。

これらの手順を使用して、AIDE Console でこれらのポリシーを定義し、ONTAP System Manager がワークスペース内のすべてのデータにそれらを適用できるようにします。

開始する前に

- AIDE Console(`\https://<cluster\_management\_ip>/console`でグローバル ポリシーを作成および管理するには、\_ストレージ管理者\_権限が必要です。
- デプロイされた正常なデータ コンピューティング ノードを含むAIDE クラスタがあります。
- "OpenID Connect (OIDC) "が設定され、IdP ロールがデータ ポリシー管理を許可する AIDE 管理者ロールにマッピングされている。
- AIDE のソフトウェア ライセンスがインストールされ、ガードレール機能と推論機能が有効になっています。
- 少なくとも 1 つのワークスペースが存在するか、管理者と調整してワークスペースで使用されるデータ ソース（ボリューム）を把握しています。

## ポリシーのタイプを理解する

AIDE Consoleは、データ資産を形成する次のポリシー タイプを公開します：

- 分類子：分類子を有効にして、すべてのワークスペースで PII、セキュリティの問題、またはその他のパターンを検出します。
- 分類子のカテゴリ：組織と管理のために、分類子をコンプライアンス カテゴリにグループ化します。
- ガードレール ポリシー：取得または推論時に適用される安全性と編集のルール。

ONTAP System Managerを使用して、これらのガードレール ポリシーを作成または管理することはできません。ストレージ管理者がワークスペースに適用した場合にのみ、それらを読み取って適用します。すべてのポリシーの定義とメンテナンスは、AIDE Consoleで行われます。

## 分類器を有効にする

分類子はメタデータとコンテンツの両方を分析して、ファイルとオブジェクトに注釈を付けます（たとえば、PII または機密カテゴリを検出します）。ワークスペースデータで分類器を実行する前に、AI Data Engine Consoleで分類器を有効にする必要があります。

タスク概要

分類器の動作は、AI Data Engine Consoleでグローバルに制御されます。有効になっているすべての分類器

は、すべてのワークスペースで実行されます。これらはグローバルに適用されるため、個々のワークスペースに対して有効または無効にすることはできません。これらはグローバルにのみ有効化または無効化できます。

#### 手順

1. AIDE Consoleで、\*Data Guardrails > Classifiers\*に移動します。
2. 分類子カテゴリを選択すると、そこに含まれる分類子が表示されます。
3. 有効にする分類子のチェックボックスを選択するか、すべての行を選択して分類子を一括で有効にします。
4. \*Enable\*を選択します。



一括選択オプションを使用して、複数の分類子を一度に有効にします。分類子を有効にするたびに、すべてのワークスペースでワークスペースの更新がトリガーされます。不要な更新を最小限に抑えるには、分類子を1つずつではなく、一度に複数有効にします。

#### 結果

新しく作成されたワークスペースと既存のワークスペースはすべて、メタデータ処理中に有効な分類子を実行します。

分類タグはメタデータ カタログに書き込まれ、データ エンジニアがデータ コレクションを作成するときにフィルタリングに使用できるようになります。

## 分類カテゴリを管理する

分類子は、カテゴリ（「PII」や「財務データ」など）に分類されます。カテゴリを使用すると、関連する分類子をグループ化して、管理とコンプライアンスの可視性を容易にすることができます。AIDEが提供するデフォルトのカテゴリを使用するか、コンプライアンス要件に合わせてカスタム カテゴリを作成できます。

#### 手順

1. AIDE Consoleで、\*Data Guardrails > Classifiers\*に移動します。
2. 既存の分類カテゴリを表示します。分類には主に2つのカテゴリがあります：
  - コンテンツまたはデータ：ファイル内の特定の種類のデータを検出します。
  - ドキュメント：コンテンツに基づいてドキュメントの種類を分類します。
3. デフォルトの分類子のサブカテゴリで十分かどうか、または独自のサブカテゴリを作成するかどうかを決定します。
  - デフォルトの分類サブカテゴリ（たとえば、**General Privacy**）を使用している場合：
    - i. 分類子カテゴリでカテゴリ名を選択すると、関連付けられている分類子が表示されます。
    - ii. 分類子のリストを調べます。
    - iii. 利用可能な分類子の完全なリストからリストされていない分類子を見つけて追加するには、追加を選択します。
  - カスタムカテゴリを作成する場合は、**+ Add** を選択します。
    - i. 一意の名前と説明を追加し、利用可能な分類子をカテゴリに割り当てます。
    - ii. **追加** を選択
4. カテゴリ内の分類子を無効にするには、分類子の **...** を選択し、\*無効にする\*を選択します。すべての行を

選択して、状態を一括で変更することもできます。

## 結果

カテゴリは、コンプライアンスの可視性を確保するために分類子を整理します。データ エンジニアは、データ コレクションをフィルタリングおよび作成するときに分類タグを使用できます。

## ガードレール ポリシーの作成と管理

ガードレール ポリシーは、分類器が機密コンテンツを検出した場合、またはプロンプトと検索結果がコンテンツ ルールに違反した場合に、AIDEがどのように応答するかを決定します。

典型的なガードレールの動作は次のとおりです：

- 取得したスニペットから PII をマスクまたは編集します。
- コンプライアンス ルールに違反する回答をブロックします。
- 監査のために違反をログに記録またはタグ付けします。

## タスク概要

ガードレールポリシーの作成と管理は、AI Data Engine Consoleでのみ行います。

ONTAP System Managerでは、ワークスペースを一度に1つのガードレールポリシーにのみ関連付けることができます。

## 手順

1. AIDE Consoleで、**Data Guardrails** > \*Guardrail policies\*に移動します。
2. \*追加\*を選択します。
3. スコープを明確に説明する名前と説明を入力します（例：Customer PII redaction for support KB）。
4. ガードレールのアクティブ化に必要なデータ分類器駆動条件を設定します：
  - a. ガードレールの作動条件を定義します：
    - i. 各条件の分類子カテゴリまたは分類子タイプを選択します。
    - ii. 必要に応じて追加の条件を追加して定義します。
    - iii. \*検索\*で特定の検索条件を定義し、\*Accept\*を選択します。
  - b. コンテンツの匿名化やデータ コレクションからのファイルのブロックと削除など、ガードレール ポリシーのアクションを定義します。
5. ガードレールを適用するワークスペースを選択します。
6. ポリシーの状態を設定します：
  - 有効：ポリシーを直ちにアクティブ化します。
  - テスト モード：ポリシーを有効にする前にその影響を検証できます。
  - 無効：ガードレールを強制せずに保存します。
7. 追加 を選択してポリシーを保存し、ワークスペースに適用します。



厳格な適用を有効にする前に、パイロット ワークスペースと非本番環境のデータ収集で **Test Mode** を使用して、影響を受ける応答の数を把握します。

## 結果

新しいガードレール ポリシーがアクティブになり、選択したワークスペースに範囲が指定されます。

## ポリシーとワークスペースの連携方法

ポリシーが定義された後：

- ストレージ管理者は ONTAP System Manager を使用してワークスペースを作成し、データ コンテナを選択し、ガードレール ポリシーを関連付けます。
- 分類子は、有効にした内容に基づいて、ワークスペースのコンテンツに対して自動的に実行されます。
- ワークスペースにアタッチされた Data Guardrails は、取得エンドポイントの動作に影響を与えます。

データ エンジニアとデータ サイエンティスト向け：

- 表示されるデータ資産（ワークスペースとデータ コレクション）は、すでにロールの割り当てによってフィルタ処理されています。
- クエリするメタデータ（PII タグなど）は、有効になっている分類子によって制御されます。
- RAG パイプラインが受信する応答は、ワークスペース レベルで構成された Data Guardrails によって制約されます。

## 関連情報

- ["AIDE でデータ コレクションを表示する"](#)
- ["AIDE のデータから RAG へのクイックスタート"](#)

# NetApp AI Data Engineに関するFAQ

このFAQでは、NetApp AI Data Engine (AIDE) に関するよくある質問について、アーキテクチャ、展開、ユーザータイプ、技術的機能、統合、ライセンスなどを含めて説明します。

## AIDEの基本

### NetApp AI Data Engineとは

NetApp AI Data Engine (AIDE) は、生データの検出と準備から、生成AI (GenAI)、Retrieval-Augmented Generation (RAG)、エージェントAI、AIファクトリーを強化するための検索エンドポイントの提供まで、AI ライフサイクル全体にわたるストレージ統合型AIデータサービスです。AIDEは同期と変更の検出を自動化し、データの検出とキュレーションのために、選択したデータの統合された最新のビューを提供します。

### AIDEはどのように機能しますか？

AIDEはNetApp ONTAPストレージシステムと直接統合し、自動化された変更検出と同期を備えたNetAppデータ資産全体のグローバルで構造化されたビューを作成します。AIDEは、圧縮と重複排除を使用したリアルタイムベクトル化、ポリシーベースのガードレール、AIツールとの統合を提供します。

## ユーザーとロール

### 誰がAI Data Engineを使用しますか？

AIDE の主なユーザーには次のものが含まれます：

1. **ONTAP** ストレージ管理者：インフラストラクチャ、AI 固有のストレージのニーズ、セキュリティ、コンプライアンスを管理します。
2. データ エンジニア：環境間でのデータの移動、準備、統合を管理します。
3. データ サイエンティスト：AI での使用のために関連データを準備し、変換します。

## 要件と導入

### どのようなハードウェアが必要ですか？

AIDE導入にはAFXシステム (AFXコントローラー、ディスク シェルフ、ネットワーク スイッチを含む) が必要ですが、SnapMirrorとクラスタピアリングを使用してONTAP 9を実行しているクラスタのクラスタデータを使用できます。高可用性とパフォーマンスを確保するために、AIDE導入には少なくとも4つのAFXコントローラーノードが必要です。

AIDEはNetAppデータ コンピューティング ノード (DCN) で実行されます。3つのDCNが必要です。DCNは、Metadata Engine、Data Sync、Data Curator、Data Guardrailsを含むAIDEソフトウェアをホストします。

### 自分のDCNを使用できますか？

いいえ。DCNはNetAppが提供するデータコンピューティングハードウェアノードであり、AIDEの唯一の導入メカニズムです。

### 必要なDCNの最小数はいくつですか？

正確に 3 つの DCN が必要です。

**DCN** で実行される **OS** は何ですか？

DCNは、NetAppが提供するソフトウェアスタックとAIDEを実行します。

**AIDE**は**AFX**なしで展開できますか？

いいえ。AIDE の導入には AFX が必要です。AIDE は Trident を使用して、AFX ボリュームを内部ストレージ（永続ボリューム）として消費します。AIDE にストレージを提供する AFX クラスタは、ONTAP 9 システムまたはクラスタとピアリングできます。クラスタ ピアリングと SnapMirror を使用して、リモート ONTAP クラスタから AFX システムにデータを同期します。

## 管理とインターフェイス

**AIDE Console**は**NetApp Console**の一部ですか、それとも別のインターフェイスですか？

AIDE Consoleは、DCN上で動作する別の管理インターフェイスです。AIDE Consoleを使用して、Data GuardrailsやData CuratorなどのAIDEサービスを管理します。ONTAP System Managerを使用してAIDEクラスタを監視することもできます。

## 機能

**AIDE**の主な特徴は何ですか？

AIDEには4つの主な機能があります：

### Metadata Engine

- 構造化された最新のインタラクティブなデータ ビューを自動的に生成します。
- ONTAP に保存されているデータで動作します。
- データ専門家がストレージ管理者と連携してデータを検索し、理解できるようにします。
- API はメタデータをクエリして機能を提供し、ストレージ システムの NFS トラフィック負荷を軽減します。
- メタデータ抽出とカタログ作成機能は、AIDE専用構築されており、継続的に動作し、スナップショットなどのONTAP機能を活用します。

### データ同期

- ソース データが変更されても、手動による介入なしにデータの最新性が自動的に維持されます。
- 管理者は、データの更新間隔を日数または時間数で定義します。
- データ全体の増分データモビリティと同期を提供し、AIデータの冗長コピーを排除します。

### Data Guardrails

- AIライフサイクル全体を通して、機密データを自動的に識別し保護します。AIDE Consoleからアクセスできます。
- データを継続的にスキャン、分類、カテゴリ化します。
- 機密データ（PII など）とリスクを識別します。
- 企業および規制基準に沿って機密データを自動的に処理するためのポリシーの作成を容易にします。
- データ保護のために機密情報の自動編集を提供します。

- 必要に応じて機密ファイルへのアクセスを制限します。

### Data Curator

- データサイエンティストがストレージ全体で関連データを検索できるようにします。
- AFX ボリューム上に存在するデータを使用して、キュレーションされたデータコレクションを作成します。
- ストレージレイヤでベクトル埋め込みを生成し、データの肥大化を抑え、パフォーマンスを向上させます。
- ベクトルセマンティック検索と再ランク付けを備えた AI アプリケーションの検索エンドポイントを提供します。

## 統合と相互運用性

**AIDE**は複数のONTAPクラスタにわたる統合メタデータをサポートしていますか？

AIDEは、SnapMirrorとクラスタピアリングを使用して複数のONTAPクラスタに接続でき、メタデータの可視性を一元化できます。

メタデータはどこに保存されますか？

AIDEは、AFXが提供する永続ボリュームを使用して、接続されたAFXクラスタにメタデータを保存します。DCNは内部操作にローカルストレージを使用します。

**AIDE Metadata Engine**はデータを分類しますか？

いいえ。AIDE Metadata Engineはファイルシステムのメタデータをカタログ化し、カタログ化されたメタデータを照会するためのAPIを提供します。

どのようなデータソースがサポートされていますか？

AIDEは、ONTAPボリューム（ローカルまたはリモート）をデータソースとしてサポートします。リモートONTAPクラスタは、ONTAP 9を実行し、クラスタピアリングとSnapMirrorを介して接続されている必要があります。

ONTAP S3バケットとStorageGRIDオブジェクトは、AIDE 9.18.1ではデータソースとしてサポートされていません。

**AIDE**で分類、ベクトル化、セマンティック検索を処理できるファイルの種類は何ですか？

AI Data Engineは、PDF、DOCX、PPTX、TXT、OCR機能付き画像ファイルなど、幅広いファイルタイプをサポートします。

**AIDE**は英語以外のデータの分類をサポートしていますか？

AIDEは英語のデータのみをサポートします。

**AIDE**はどのような統合をサポートしていますか？

AIDEは、直接API呼び出しまたはModel Context Protocol (MCP) サーバーを通じてアクセスできるRAG API エンドポイントを提供します。これにより、エージェントAIフレームワークおよびツールとの統合がサポートされます。

# 導入とライセンス

導入オプションには何がありますか？

AIDEは、DCNを使用したAFXインフラストラクチャ上のオンプレミスに展開されます。NetApp ONTAP AFX インストールと直接統合します。

**AIDE** のライセンスはどのようになっていますか？

AIDEでData GuardrailsとData Curatorを実行するには、ソフトウェアライセンスが必要です。

Metadata Engineのみが必要な場合は、すべてのAFXシステムに含まれるONTAP Oneライセンスで、Metadata Engineのみの機能の使用権限が提供されます。

関連情報

- ["ONTAP System ManagerでAIDEライセンスをインストールする"](#)
- ["AIDE のアーキテクチャとコンポーネントについて"](#)

# 法律上の表示

法的通知から、著作権情報、商標、特許などを確認できます。

## 著作権

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

## 商標

NetApp、NetAppのロゴ、NetAppの商標一覧のページに掲載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。

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

## 特許

現在NetAppが所有する特許の一覧は以下のページから閲覧できます。

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

## プライバシー ポリシー

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

## オープンソース

通知ファイルは、NetAppソフトウェアで使用されるサードパーティの著作権とライセンスに関する情報を提供します。

## AI Data Engine

["AIDE 9.18.1に関するお知らせ"](#)

## 著作権に関する情報

Copyright © 2026 NetApp, Inc. All Rights Reserved. Printed in the U.S.このドキュメントは著作権によって保護されています。著作権所有者の書面による事前承諾がある場合を除き、画像媒体、電子媒体、および写真複写、記録媒体、テープ媒体、電子検索システムへの組み込みを含む機械媒体など、いかなる形式および方法による複製も禁止します。

ネットアップの著作物から派生したソフトウェアは、次に示す使用許諾条項および免責条項の対象となります。

このソフトウェアは、ネットアップによって「現状のまま」提供されています。ネットアップは明示的な保証、または商品性および特定目的に対する適合性の暗示的保証を含み、かつこれに限定されないいかなる暗示的な保証も行いません。ネットアップは、代替品または代替サービスの調達、使用不能、データ損失、利益損失、業務中断を含み、かつこれに限定されない、このソフトウェアの使用により生じたすべての直接的損害、間接的損害、偶発的損害、特別損害、懲罰的損害、必然的損害の発生に対して、損失の発生の可能性が通知されていたとしても、その発生理由、根拠とする責任論、契約の有無、厳格責任、不法行為（過失またはそうでない場合を含む）にかかわらず、一切の責任を負いません。

ネットアップは、ここに記載されているすべての製品に対する変更を随時、予告なく行う権利を保有します。ネットアップによる明示的な書面による合意がある場合を除き、ここに記載されている製品の使用により生じる責任および義務に対して、ネットアップは責任を負いません。この製品の使用または購入は、ネットアップの特許権、商標権、または他の知的所有権に基づくライセンスの供与とはみなされません。

このマニュアルに記載されている製品は、1つ以上の米国特許、その他の国の特許、および出願中の特許によって保護されている場合があります。

権利の制限について：政府による使用、複製、開示は、DFARS 252.227-7013（2014年2月）およびFAR 5252.227-19（2007年12月）のRights in Technical Data -Noncommercial Items（技術データ - 非商用品目に関する諸権利）条項の(b)(3)項、に規定された制限が適用されます。

本書に含まれるデータは商用製品および/または商用サービス（FAR 2.101の定義に基づく）に関係し、データの所有権はNetApp, Inc.にあります。本契約に基づき提供されるすべてのネットアップの技術データおよびコンピュータソフトウェアは、商用目的であり、私費のみで開発されたものです。米国政府は本データに対し、非独占的かつ移転およびサブライセンス不可で、全世界を対象とする取り消し不能の制限付き使用权を有し、本データの提供の根拠となった米国政府契約に関連し、当該契約の裏付けとする場合にのみ本データを使用できます。前述の場合を除き、NetApp, Inc.の書面による許可を事前に得ることなく、本データを使用、開示、転載、改変するほか、上演または展示することはできません。国防総省にかかる米国政府のデータ使用权については、DFARS 252.227-7015(b)項（2014年2月）で定められた権利のみが認められます。

## 商標に関する情報

NetApp、NetAppのロゴ、<http://www.netapp.com/TM>に記載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。