NetApp AIPod Mini - NetAppとIntelによるエンタープライズRAG推論

NetApp artificial intelligence solutions

NetApp
February 12, 2026

目次

NetApp AIPod Mini - NetAppとIntelによるエンタープライズRAG推論 . . . 1
- エグゼクティブサマリー . . . 1
- Intel ストレージ パートナーの検証 . . . 1
- NetAppでRAGシステムを実行する利点 . . . 1
- 対象 . . . 2
- テクノロジ要件 . . . 2
  - ハードウェア . . . 2
  - ソフトウェア . . . 4
- ソリューションの展開 . . . 6
  - ソフトウェアスタック . . . 6
  - 導入手順 . . . 6
- サイズガイド . . . 12
- まとめ . . . 13
- 了承 . . . 13
- 部品表 . . . 13
- インフラ準備チェックリスト . . . 14
- 詳細情報の入手方法 . . . 15

NetApp AIPod Mini - NetAppとIntelによるエンタープライズRAG推論

このホワイト ペーパーでは、Intel Xeon 6 プロセッサとNetAppデータ管理ソリューションのテクノロジと組み合わせた機能を備えた、エンタープライズ RAG 向けNetApp AIPodの検証済みリファレンス デザインについて説明します。このソリューションは、大規模な言語モデルを活用したダウンストリーム ChatQnA アプリケーションを示し、同時ユーザーに正確でコンテキストに適した応答を提供します。応答は、エアギャップ RAG 推論パイプラインを通じて組織の内部知識リポジトリから取得されます。

Sathish Thyagarajan、Michael Oglesby、Arpita Mahajan、NetApp

エグゼクティブサマリー

生産性とビジネス価値を高めるために、検索拡張生成（RAG）アプリケーションと大規模言語モデル（LLM）を活用して、ユーザーのプロンプトを解釈し、応答を生成する組織が増えています。これらのプロンプトと応答には、組織の内部ナレッジベース、データレイク、コードリポジトリ、ドキュメントリポジトリから取得されたテキスト、コード、画像、さらには治療用タンパク質構造などが含まれます。本書では、NetApp AFF ストレージと Intel Xeon 6 プロセッサを搭載したサーバーで構成される NetApp AIPod Mini ソリューションのリファレンス設計について説明します。NetApp ONTAP データ管理ソフトウェアと Intel Advanced Matrix Extensions（Intel AMX）、および Open Platform for Enterprise AI（OPEA）上に構築された Intel® AI for Enterprise RAG ソフトウェアが含まれます。NetApp AIPod Mini for enterprise RAG を使用すると、組織はパブリック LLM をプライベート生成 AI（GenAI）推論ソリューションに拡張できます。このソリューションは、信頼性を高め、独自の情報をより適切に制御できるように設計された、効率的で対費用効果の高いエンタープライズ規模の RAG 推論を実現します。

Intel ストレージ パートナーの検証

Intel Xeon 6 プロセッサーを搭載したサーバーは、Intel AMX を使用して最高のパフォーマンスを実現し、要求の厳しい AI 推論ワークロードを処理できるように構築されています。最適なストレージ パフォーマンスとスケーラビリティを実現するために、このソリューションはNetApp ONTAPを使用して検証されており、企業は RAG アプリケーションのニーズを満たすことができます。この検証は、Intel Xeon 6 プロセッサを搭載したサーバーで実施されました。 Intel とNetApp は、最適化され、拡張可能で、顧客のビジネス要件に適合した AI ソリューションの提供に重点を置いた強力なパートナーシップを築いています。

NetAppでRAGシステムを実行する利点

RAG アプリケーションでは、PDF、テキスト、CSV、Excel などのさまざまな形式で企業のドキュメント リポジトリから知識を取得します。このデータは通常、S3 オブジェクト ストレージやオンプレミスの NFS などのソリューションにデータのソースとして保存されます。NetApp は、エッジ、データ センター、クラウドのエコシステム全体にわたるデータ管理、データ モビリティ、データ ガバナンス、データ セキュリティ テクノロジーのリーダーです。NetApp ONTAP データ管理は、バッチやリアルタイム推論を含むさまざまなタイプの AI ワークロードをサポートするエンタープライズクラスのストレージを提供し、次のような利点があります：

- 速度とスケーラビリティ。パフォーマンスと容量を個別に拡張できるため、大規模なデータセットを高速に処理してバージョン管理を行うことができます。
- データアクセス。マルチプロトコル サポートにより、クライアント アプリケーションは S3、NFS、SMB ファイル共有プロトコルを使用してデータを読み取ることができます。 ONTAP S3 NAS バケットは、マルチモーダル LLM 推論シナリオでのデータ アクセスを容易にします。
- 信頼性と機密性。 ONTAP は、データ保護、組み込みのNetApp Autonomous Ransomware Protection (ARP)、ストレージの動的プロビジョニングを提供し、機密性とセキュリティを強化するためにソフトウェアベースとハードウェアベースの両方の暗号化を提供します。 ONTAP は、すべての SSL 接続に関して FIPS 140-2 に準拠しています。

対象

このドキュメントは、エンタープライズ RAG および GenAI ソリューションを提供するために構築されたインフラストラクチャを活用したい AI 意思決定者、データ エンジニア、ビジネス リーダー、部門幹部を対象としています。 AI 推論、LLM、Kubernetes、ネットワークとそのコンポーネントに関する事前の知識は、実装フェーズで役立ちます。

テクノロジ要件

ハードウェア

Intel® AI テクノロジー

Xeon 6 をホスト CPU として使用することで、高速化されたシステムは、高いシングルスレッド パフォーマンス、メモリ帯域幅の拡大、信頼性、可用性、保守性 (RAS) の向上、および I/O レーンの増加といったメリットを享受できます。 Intel AMX は、INT8 および BF16 の推論を高速化し、FP16 トレーニング済みモデルのサポートを提供します。INT8 の場合はコアあたりサイクルあたり最大 2,048 回の浮動小数点演算、BF16/FP16 の場合はコアあたりサイクルあたり最大 1,024 回の浮動小数点演算が可能です。 Xeon 6 プロセッサを使用して RAG ソリューションを展開するには、通常、最低 250 GB の RAM と 500 GB のディスク容量が推奨されます。ただし、これは LLM モデルのサイズに大きく依存します。詳細については、Intel "Xeon 6プロセッサ"製品概要。

図1 - Intel Xeon 6プロセッサを搭載したコンピューティングサーバ

—

NetApp AFFストレージ

エントリーレベルおよびミッドレベルのNetApp AFF A シリーズ システムは、より強力なパフォーマンス、密度、および優れた効率性を提供します。 NetApp AFF A20、 AFF A30、 AFF A50 システムは、ハイブリッドクラウド全体で最低コストで RAG アプリケーションのデータをシームレスに管理、保護、および移動できる単一の OS に基づいて、ブロック、ファイル、およびオブジェクトをサポートする真の統合ストレージを提供します。

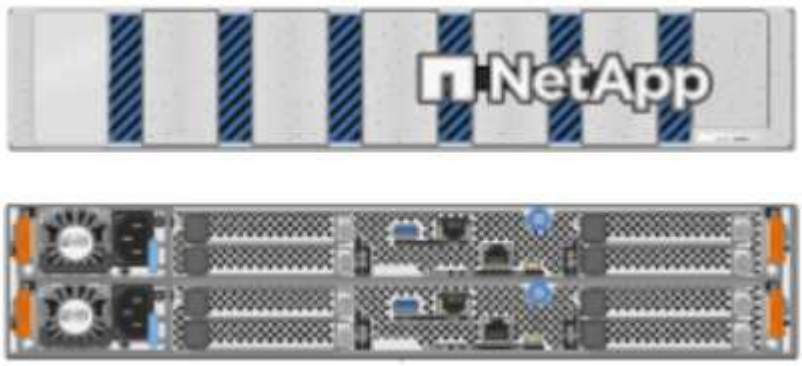


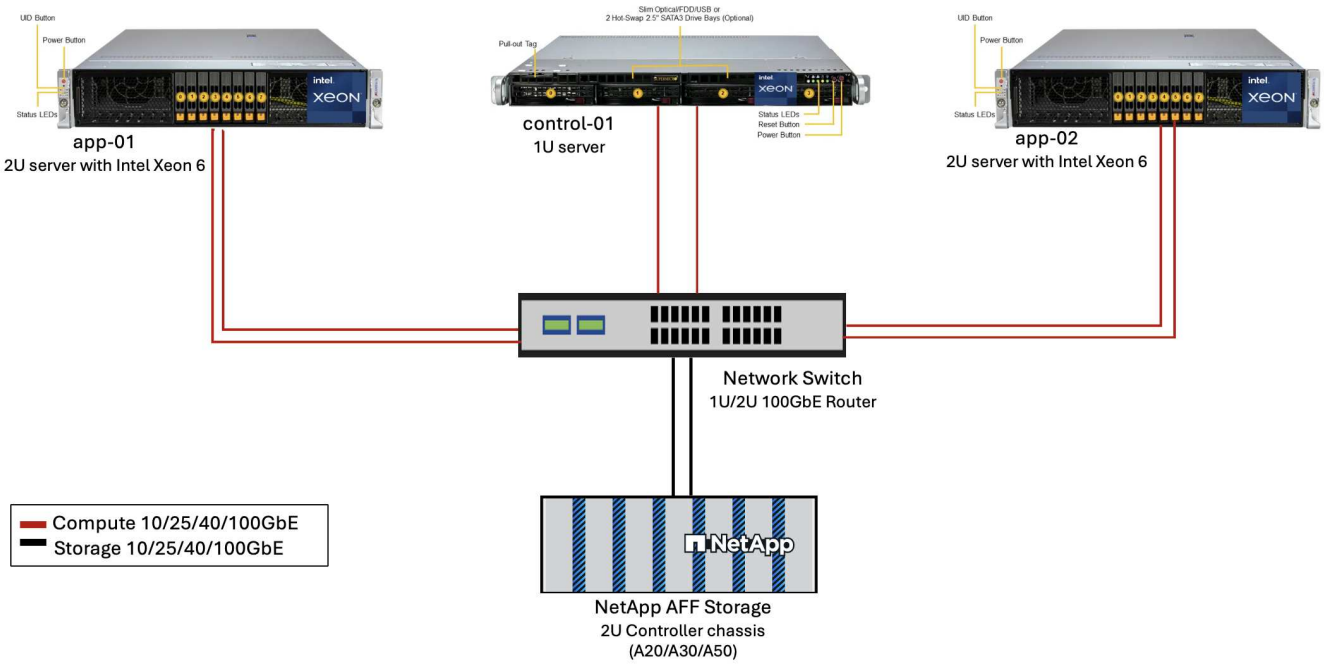
図 2 - NetApp AFF A シリーズ システム。

ハードウェア	量	コメント
Intel Xeon 第6世代（Granite Rapids）	2	RAG 推論ノード - デュアル ソケット Intel Xeon 6900 シリーズ（96 コア）または Intel Xeon 6700 シリーズ（64 コア）プロセッサと、DDR5（6400MHz）または MRDIMM（8800MHz）の 250GB ～ 3TB の RAM を搭載。2U サーバー。

ハードウェア	量	コメント
Intel プロセッサを搭載したコントロール プレーン サーバー	1	Kubernetes コントロール プレーン/1U サーバー。
100Gbイーサネットスイッチの選択	1	データセンタースイッチ。
NetApp AFF A20（またはAFF A30、 AFF A50）	1	最大ストレージ容量: 9.3PB。注: ネットワーク: 10/25/100 GbE ポート。

このリファレンス デザインの検証には、Supermicro の Intel Xeon 6 プロセッサ (222HA-TN-OTO-37) を搭載したサーバーと Arista の 100GbE スイッチ (7280R3A) が使用されました。

図3 - AIPod Miniデプロイメントアーキテクチャ



ソフトウェア

エンタープライズ**AI**向けオープンプラットフォーム

Open Platform for Enterprise AI (OPEA) は、Intel がエコシステム パートナーと協力して主導するオープン ソース イニシアチブです。 RAG に重点を置いた最先端の生成 AI システムの開発を加速するように設計された、構成可能なビルディング ブロックのモジュール式プラットフォームを提供します。 OPEA には、LLM、データストア、プロンプト エンジン、RAG アーキテクチャ ブループリント、パフォーマンス、機能、信頼性、エンタープライズ準備に基づいて生成 AI システムを評価する 4 段階の評価方法を備えた包括的なフレームワークが含まれています。

OPEA は、主に次の 2 つの主要コンポーネントで構成されています。

- GenAIComps: マイクロサービスコンポーネントで構成されたサービスベースのツールキット

- GenAIExamples: ChatQnAのような、実用的なユースケースを示すすぐに導入できるソリューション

詳細については、 "OPEAプロジェクトドキュメント"

OPEA 搭載の Intel® AI for Enterprise RAG

OPEA for Intel® AI for Enterprise RAG は、エンタープライズデータを実用的な洞察に変換する作業を簡素化します。Intel Xeon プロセッサーを搭載し、業界パートナーのコンポーネントを統合して、エンタープライズソリューションの導入を合理化するアプローチを提供します。実績のあるオーケストレーションフレームワークを使用してシームレスに拡張し、企業が必要とする柔軟性と選択肢を提供します。

OPEA の基盤を基に、Intel® AI for Enterprise RAG は、スケーラビリティ、セキュリティ、ユーザー エクスペリエンスを強化する主要な機能によってこの基盤を拡張します。これらの機能には、最新のサービスベースのアーキテクチャとのシームレスな統合を実現するサービス メッシュ機能、パイプラインの信頼性を実現する本番環境対応の検証、ワークフローの管理と監視を容易にする RAG as a Service 用の機能豊富な UI などが含まれます。さらに、Intel とパートナーのサポートにより、安全でコンプライアンスに準拠した運用を実現する UI とアプリケーションを備えた統合 ID およびアクセス管理（IAM）を組み合わせた、幅広いソリューション エコシステムへのアクセスが提供されます。プログラム可能なガードレールにより、パイプラインの動作をきめ細かく制御でき、セキュリティとコンプライアンスの設定をカスタマイズできます。

NetApp ONTAP

NetApp ONTAP は、NetApp の重要なデータ ストレージ ソリューションを支える基盤テクノロジーです。ONTAP には、サイバー攻撃に対する自動ランサムウェア保護、組み込みのデータ転送機能、ストレージ効率機能など、さまざまなデータ管理およびデータ保護機能が含まれています。これらの利点は、オンプレミスから、NAS、SAN、オブジェクト、LLM 展開のソフトウェア定義ストレージのハイブリッド マルチクラウドまで、さまざまなアーキテクチャに適用されます。 ONTAPクラスタ内のONTAP S3 オブジェクト ストレージ サーバを使用して RAG アプリケーションを導入し、承認されたユーザーとクライアント アプリケーションを通じて提供されるONTAPのストレージ効率とセキュリティを活用できます。詳細については、 "ONTAP S3 の構成について学ぶ"

NetAppTrident

NetApp Tridentソフトウェアは、Red Hat OpenShift を含むコンテナおよび Kubernetes ディストリビューション向けのオープンソースで完全にサポートされているストレージ オーケストレーターです。 Trident は、NetApp ONTAPを含むNetAppストレージ ポートフォリオ全体と連携し、NFS および iSCSI 接続もサポートします。詳細については、 "Git 上のNetApp Trident"

ソフトウェア	バージョン	コメント
OPEA - Intel® AI for Enterprise RAG	2.0	OPEAマイクロサービスに基づくエンタープライズRAGプラットフォーム
コンテナ ストレージ インターフェース (CSI ドライバー)	NetApp Trident 25.10	動的プロビジョニング、 NetAppスナップショット コピー、ボリュームを有効にします。
Ubuntu	22.04.5	2ノードクラスタのOS。
コンテナオーケストレーション	Kubernetes 1.31.9（Enterprise RAG インフラストラクチャ プレイブックによってインストール）	RAGフレームワークを実行する環境
ONTAP	ONTAP 9.16.1P4以上	AFF A20上のストレージOS。

ソリューションの展開

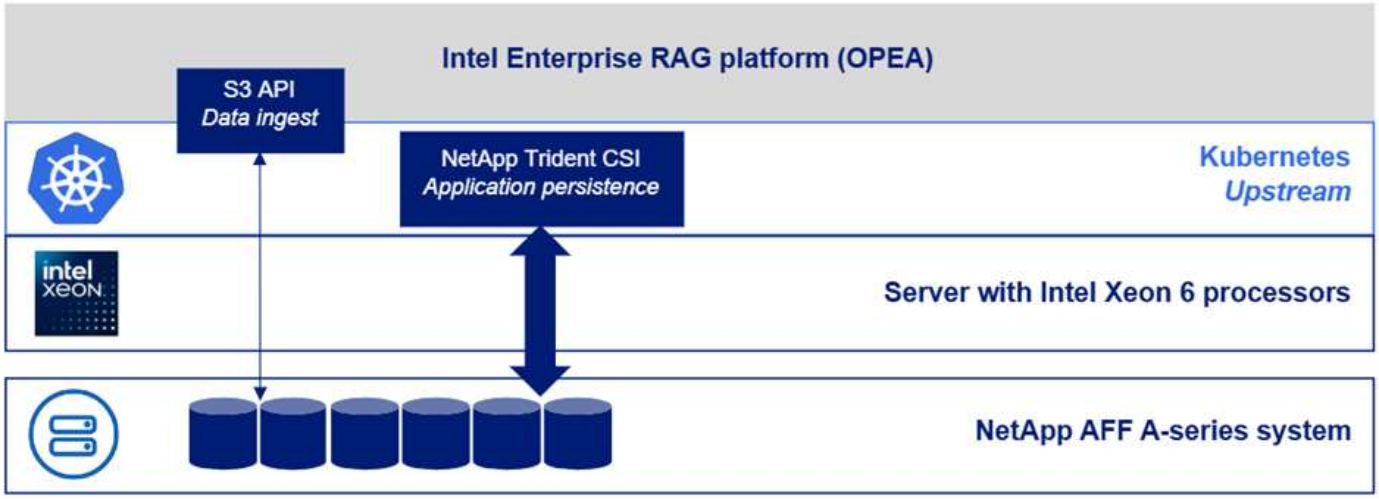
ソフトウェアスタック

このソリューションは、Intel Xeon ベースのアプリケーション ノードで構成される Kubernetes クラスターに展開されます。 Kubernetes コントロール プレーンの基本的な高可用性を実装するには、少なくとも 3 つのノードが必要です。次のクラスター レイアウトを使用してソリューションを検証しました。

表3 - Kubernetesクラスタレイアウト

ノード	ロール	数量
Intel Xeon 6プロセッサと1TB RAMを搭載したサーバー	アプリノード、コントロールプレーンノード	2
汎用サーバー	コントロールプレーンノード	1

次の図は、ソリューションの「ソフトウェア スタック ビュー」を示しています。



導入手順

ONTAPストレージアプライアンスを導入する

NetApp ONTAPストレージ アプライアンスを導入およびプロビジョニングします。参照 "ONTAPハードウェア システムのドキュメント"詳細については。

NFSおよび**S3**アクセス用に**ONTAP SVM**を構成する

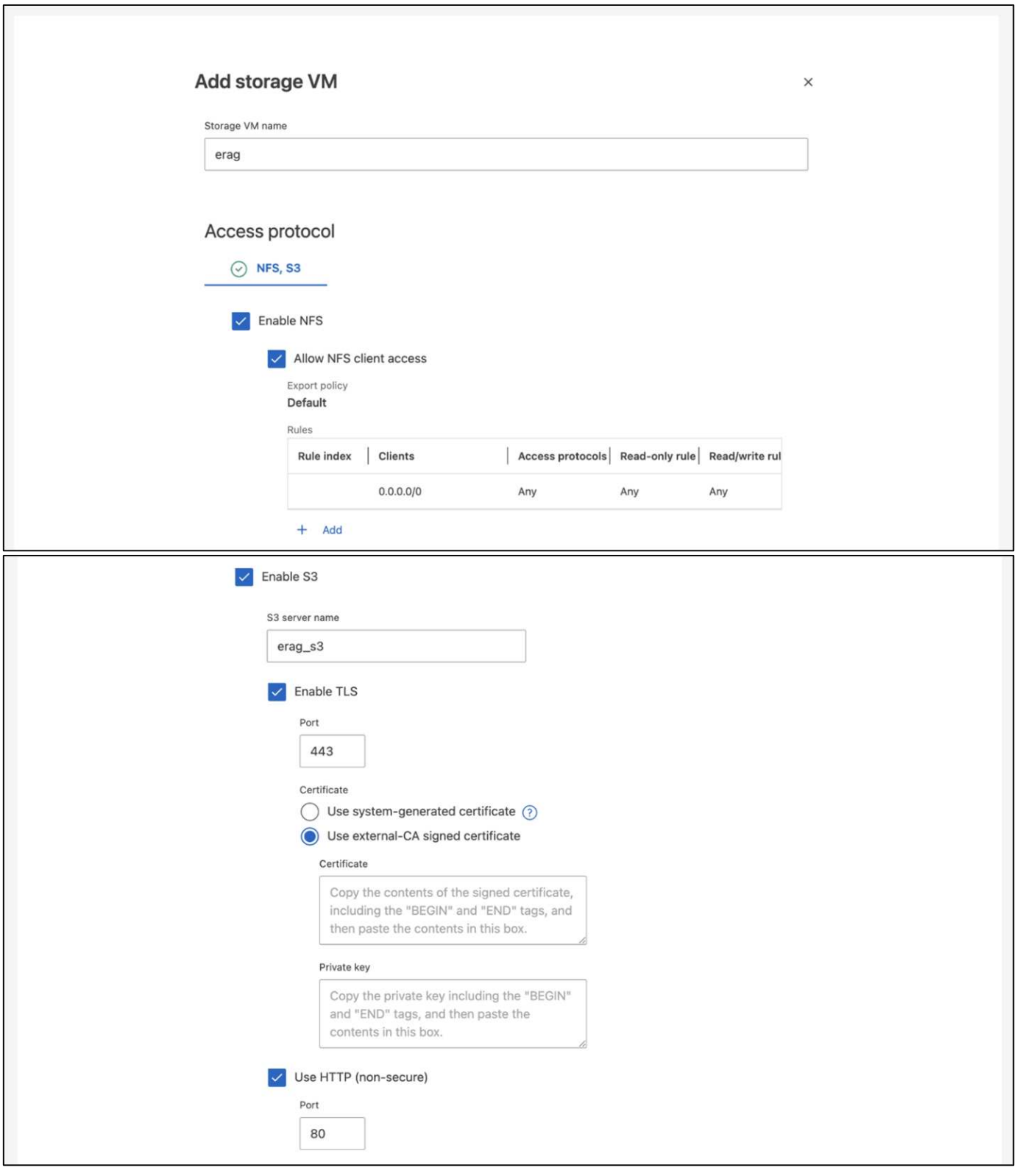
Kubernetes ノードからアクセス可能なネットワーク上で、NFS および S3 アクセス用のONTAPストレージ仮想マシン (SVM) を構成します。

ONTAP System Manager を使用して SVM を作成するには、[ストレージ] > [ストレージ VM] に移動し、[+ 追加] ボタンをクリックします。 SVM の S3 アクセスを有効にするときは、システム生成の証明書ではなく、外部 CA (証明機関) 署名付き証明書を使用するオプションを選択します。自己署名証明書または公的に信頼された CA によって署名された証明書のいずれかを使用できます。詳細については、 "ONTAP のドキュメント。"

次のスクリーンショットは、 ONTAP System Manager を使用して SVM を作成する様子を示しています。環

境に応じて必要に応じて詳細を変更します。

図5 - ONTAP System Managerを使用したSVMの作成。



S3の権限を設定する

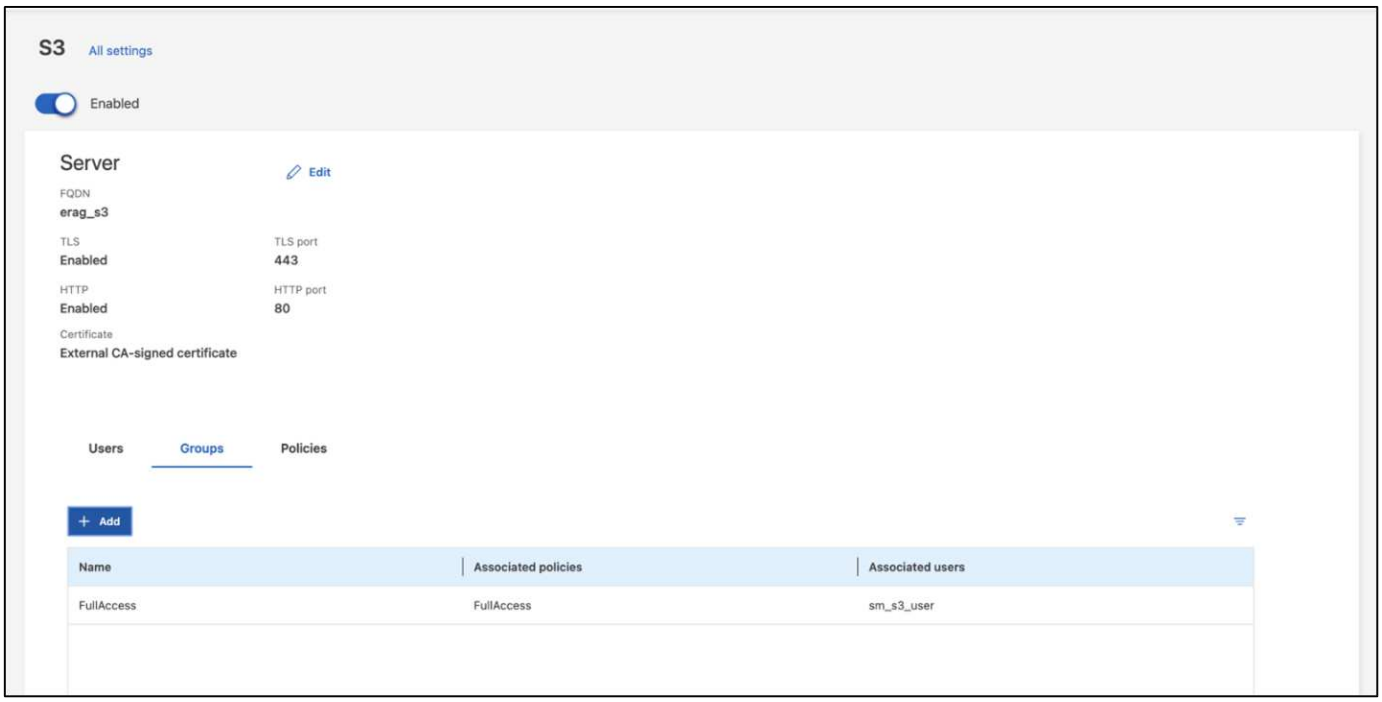
前の手順で作成した SVM の S3 ユーザー/グループ設定を構成します。その SVM のすべての S3 API 操作へ

のフルアクセス権を持つユーザーがいることを確認します。詳細については、 ONTAP S3 のドキュメントを参照してください。

注：このユーザーは、Intel® AI for Enterprise RAGアプリケーションのデータ取り込みサービスに必要になります。ONTAP System Managerを使用してSVMを作成した場合、System Managerは、SVMの作成時に`sm_s3_user`という名前のユーザーと `FullAccess`という名前のポリシーを自動的に作成しますが、`sm_s3_user`には権限が割り当てられません。

このユーザーの権限を編集するには、[ストレージ] > [ストレージ VM] に移動し、前の手順で作成した SVM の名前をクリックして、[設定] をクリックし、[S3] の横にある鉛筆アイコンをクリックします。与える`sm_s3_user`すべてのS3 API操作へのフルアクセスを付与するには、関連付ける新しいグループを作成します。 `sm_s3_user`と `FullAccess`次のスクリーンショットに示すようなポリシーです。

図6 - S3のアクセス許可

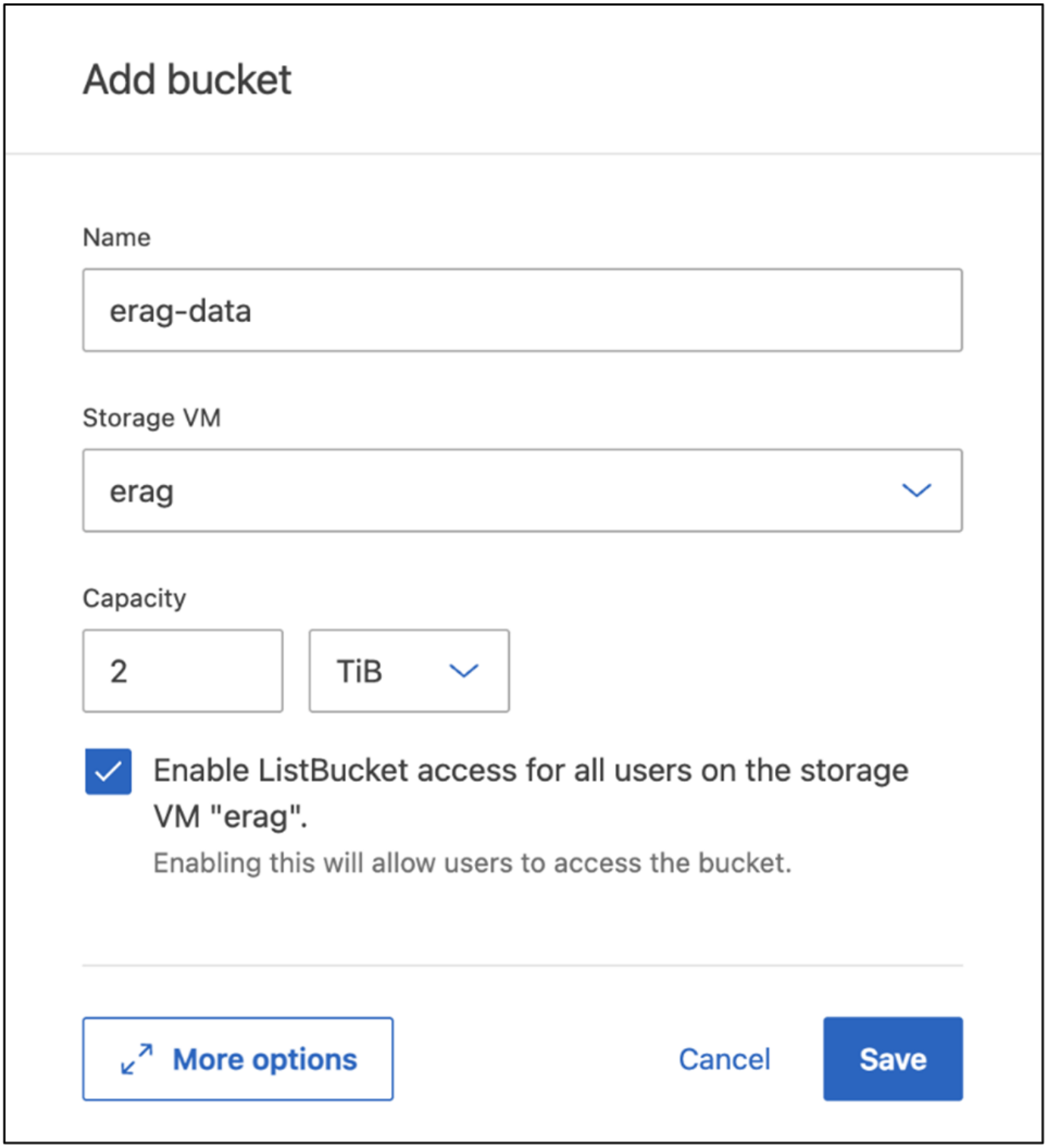


S3バケットを作成する

先ほど作成した SVM 内に S3 バケットを作成します。 ONTAP System Manager を使用して SVM を作成するには、[ストレージ] > [バケット] に移動し、[+ 追加] ボタンをクリックします。詳細については、 ONTAP S3 のドキュメントを参照してください。

次のスクリーンショットは、 ONTAP System Manager を使用して S3 バケットを作成する様子を示しています。

図 7 - S3 バケットを作成する。



S3バケットの権限を設定する

前の手順で作成した S3 バケットのアクセス許可を設定します。前の手順で設定したユーザーに次の権限があることを確認します。 `GetObject, PutObject, DeleteObject, ListBucket, GetBucketAcl, GetObjectAcl, ListBucketMultipartUploads, ListMultipartUploadParts, GetObjectTagging, PutObjectTagging, DeleteObjectTagging, GetBucketLocation, GetBucketVersioning, PutBucketVersioning, ListBucketVersions, GetBucketPolicy, PutBucketPolicy, DeleteBucketPolicy, PutLifecycleConfiguration, GetLifecycleConfiguration, GetBucketCORS, PutBucketCORS.`

ONTAP System Manager を使用して S3 バケットの権限を編集するには、[ストレージ] > [バケット] に移動し、バケットの名前をクリックして、[権限] をクリックし、[編集] をクリックします。参照 "ONTAP S3 ドキュメント"詳細については、こちらをご覧ください。

次のスクリーンショットは、 ONTAP System Manager で必要なバケット権限を示しています。

図8：S3バケットの権限。

バケットのクロスオリジンリソース共有ルールを作成する

ONTAP CLI を使用して、前の手順で作成したバケットのバケット クロスオリジン リソース共有 (CORS) ルールを作成します。

```
ontap::> bucket cors-rule create -vserver erag -bucket erag-data -allowed
-origins *erag.com -allowed-methods GET,HEAD,PUT,DELETE,POST -allowed
-headers *
```

このルールにより、OPEA for Intel® AI for Enterprise RAG Webアプリケーションは、Webブラウザー内からバケットと対話できるようになります。

サーバーの展開

サーバーを展開し、各サーバーに Ubuntu 22.04 LTS をインストールします。 Ubuntu をインストールしたら、すべてのサーバーに NFS ユーティリティをインストールします。 NFS ユーティリティをインストールするには、次のコマンドを実行します。

```
apt-get update && apt-get install nfs-common
```

Enterprise RAG 2.0を導入する

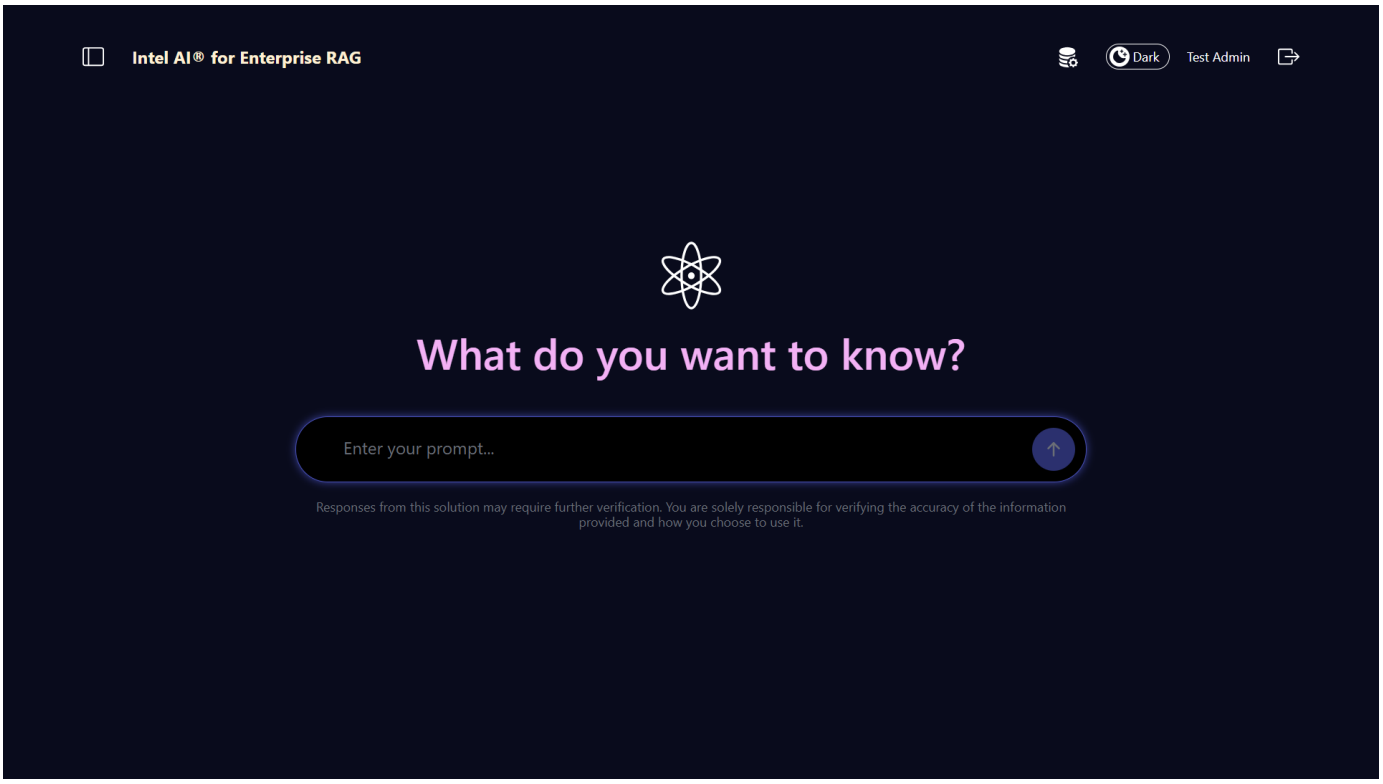
完全なステップバイステップの導入ワークフローについては、次のドキュメントを参照してください：NetApp AIPod Mini for ERAG - 導入手順すべての前提条件、インフラストラクチャの準備、構成パラメータ、および導入手順は、上記の導入ガイドに記載されています。

Intel® AI for Enterprise RAG UI の **OPEA** にアクセスする

Intel® AI for Enterprise RAG UI の OPEA にアクセスします。詳細については、 "Intel® AI for Enterprise RAG

導入ドキュメント"を参照してください。

図 9 - Intel® AI for Enterprise RAG UI 向け OPEA
。



RAGのデータを取り込む

RAG ベースのクエリ拡張に含めるファイルを取り込むことができるようになりました。ファイルの取り込みには複数のオプションがあります。ニーズに応じて適切なオプションを選択してください。

注：ファイルが取り込まれた後、OPEA for Intel® AI for Enterprise RAG アプリケーションはファイルの更新を自動的にチェックし、それに応じて更新を取り込みます。

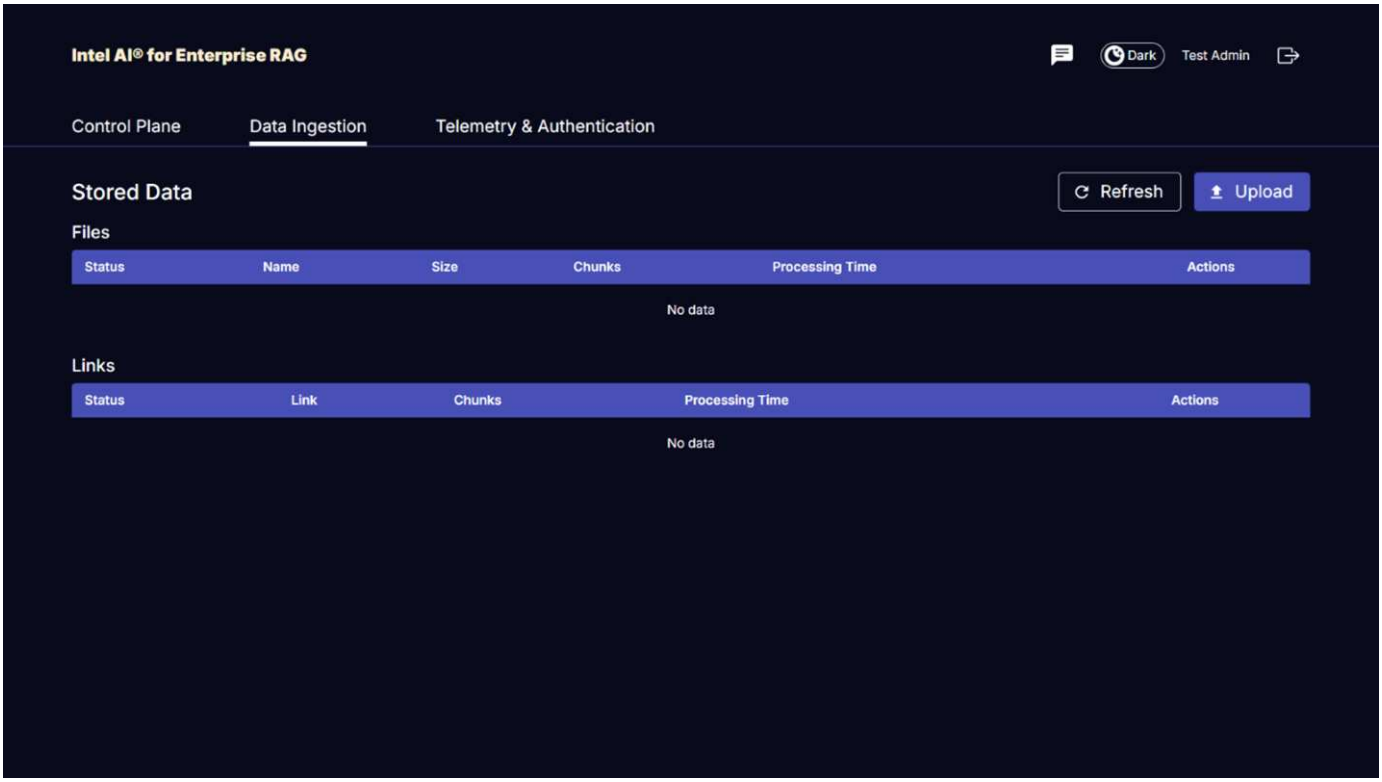
*オプション1：S3バケットに直接アップロードする 一度に多くのファイルを取り込むには、任意のS3クライアントを使用して、ファイルをS3バケット（以前に作成したバケット）にアップロードすることをお勧めします。一般的なS3クライアントには、AWS CLI、Amazon SDK for Python（Boto3）、s3cmd、S3 Browser、Cyberduck、Commander Oneなどがあります。ファイルがサポートされているタイプである場合、S3バケットにアップロードしたファイルは、OPEA for Intel® AI for Enterprise RAGアプリケーションによって自動的に取り込まれます。

注：このドキュメントの執筆時点では、PDF、HTML、TXT、DOC、DOCX、ADOC、PPT、PPTX、MD、XML、JSON、JSONL、YAML、XLS、XLSX、CSV、TIFF、JPG、JPEG、PNG、SVG のファイルタイプがサポートされています。

OPEA for Intel® AI for Enterprise RAG UI を使用して、ファイルが適切に取り込まれたことを確認できます。詳細については、Intel® AI for Enterprise RAG UI のドキュメントを参照してください。アプリケーションが大量のファイルを取り込むには時間がかかる場合があることに注意してください。

*オプション2：UIを使用してアップロードする 少数のファイルのみを取り込む必要がある場合は、OPEA for Intel® AI for Enterprise RAG UIを使用して取り込むことができます。詳細については、Intel® AI for Enterprise RAG UIのドキュメントを参照してください。

図 10 - データ取り込み UI
。



チャットクエリを実行する

付属のチャット UI を使用して、OPEA for Intel® AI for Enterprise RAG アプリケーションと「チャット」できるようになりました。クエリに応答する際、アプリケーションは取り込んだファイルを使用して RAG を実行します。つまり、アプリケーションは取り込んだファイル内の関連情報を自動的に検索し、クエリに応答するときにこの情報を組み込みます。

サイズガイド

検証作業の一環として、Intel と連携してパフォーマンス テストを実施しました。このテストの結果、次の表に示すサイズ設定のガイドラインが得られました。

特徴づけ	**Value**	コメント
モデルサイズ	200億のパラメータ	ラマ-8B、ラマ-13B、ミストラル7B、クウェン14B、ディープシーク・ディスティル8B
入力サイズ	約2,000トークン	約4ページ
出力サイズ	約2,000トークン	約4ページ
同時ユーザー数	32	「同時ユーザー」とは、同時にクエリを送信しているプロンプト要求を指します。

_注：上記のサイジングガイダンスは、96コアのIntel Xeon 6プロセッサーを使用して収集されたパフォーマンス検証およびテスト結果に基づいています。同様のI/Oトークンとモデルサイズ要件を持つお客様には、96コアのXeon 6プロセッサーを搭載したサーバーの使用をお勧めします。サイジングガイドの詳細について

は、"Intel® AI for Enterprise RAG サイジング ガイド"を参照してください。

まとめ

エンタープライズRAGシステムとLLMは連携して動作し、組織が正確でコンテキストに応じた応答を提供できるように支援するテクノロジーです。これらの応答には、膨大な量のプライベートデータと企業内部データに基づく情報検索が含まれます。RAG、API、ベクトル埋め込み、およびハイパフォーマンスストレージシステムを使用して企業データを含むドキュメントリポジトリを照会すると、データはより高速かつ安全に処理されます。NetApp AIPod Miniは、NetAppのインテリジェントデータインフラストラクチャとONTAPデータ管理機能、Intel Xeon 6プロセッサー、Intel® AI for Enterprise RAG、およびOPEAソフトウェアスタックを組み合わせることで、ハイパフォーマンスRAGアプリケーションの導入を支援し、組織をAIリーダーシップへの道へと導きます。

了承

この文書は、NetApp ソリューション エンジニアリング チームのメンバーである Sathish Thyagarajan 、Michael Oglesby、Arpita Mahajan によって執筆されました。著者らはまた、Intel のエンタープライズ AI 製品チーム（Ajay Mungara、Mikolaj Zyczynski、Igor Konopko、Ramakrishna Karamsetty、Michal Prostko 、Anna Alberska、Maciej Cichocki、Shreejan Mistry、Nicholas Rago、Ned Fiori）、および NetApp の他のチームメンバー（Lawrence Bunka、Bobby Oommen、Jeff Liborio）に、ソリューション検証プロセス中の継続的なサポートと支援に対して感謝の意を表します。

部品表

以下は、このソリューションの機能検証に使用された BOM であり、参照として使用できます。次の構成に適合する任意のサーバーまたはネットワーク コンポーネント (または、100GbE 帯域幅が望ましい既存のネットワーク) を使用できます。

アプリサーバーの場合:

部品番号	製品説明	量
222HA-TN-OTO-37	ハイパースーパーサーバー SYS-222HA-TN /2U	2
P4X-GNR6972P-SRPL2-UC	Intel® Xeon® 6972P プロセッサー96コア2.40GHz 480MBキャッシュ（500W）	4
RAM	MEM-DR564MC-ER64(x16)64GB DDR5-6400 2RX4 (16Gb) ECC RDIMM	32
	HDS-M2N4-960G0-E1-TXD-NON-080(x2) SSD M.2 NVMe PCIe4 960GB 1DWPD TLC D、80mm	2
	WS-1K63A-1R(x2)1U 692W/1600W 冗長シングル出力電源。熱放散は 2361 BTU/時、最高温度は 59 °C（約）	4

制御サーバーの場合:

部品番号	製品説明	量
511R-M-OTO-17	1U X13SCH-SYS、CSE-813MF2TS-R0RCNBP、PWS-602A-1Rまで最適化	1
	RPL-E 6369P IP 8C/16T 3.3G 24MB 95W 1700 BO	1
RAM	MEM-DR516MB-EU48(x2)16GB DDR5-4800 1Rx8 (16Gb) ECC UDIMM	1
	HDS-M2N4-960G0-E1-TXD-NON-080(x2) SSD M.2 NVMe PCIe4 960GB 1DWPD TLC D、80mm	2

ネットワーク スイッチの場合:

部品番号	製品説明	量
DCS-7280CR3A	アリスタ 7280R3A 28x100 GbE	1

NetApp AFFストレージ:

部品番号	製品説明	量
AFF-A20A-100-C	AFF A20 HAシステム、-C	1
X800-42U-R6-C	ジャンパーカード、インキャブ、C13-C14、-C	2
X97602A-C	電源、1600W、チタン、-C	2
X66211B-2-NC	ケーブル、100GbE、QSFP28-QSFP28、Cu、2m、-C	4
X66240A-05-NC	ケーブル、25GbE、SFP28-SFP28、Cu、0.5m、-C	2
X5532A-NC	レール、4ポスト、薄型、Rnd/Sq-Hole、Sm、Adj、24-32、-C	1
X4024A-2-AC	ドライブ パック 2X1.92TB 、NVMe4、SED、-C	6
X60130A-C	IO モジュール、2PT、100GbE、-C	2
X60132A-C	IOモジュール、4PT、10/25GbE、-C	2
SW-ONTAPB-FLASH-A20-C	SW、 ONTAPベース パッケージ、TB 単位、フラッシュ、A20、-C	23

インフラ準備チェックリスト

詳細については、NetApp AIPod Mini - インフラ準備を参照してください。

詳細情報の入手方法

このドキュメントに記載されている情報の詳細については、次のドキュメントや Web サイトを参照してください。

"NetApp製品ドキュメント"

"OPEAプロジェクト"

"Intel® AI ERAG ドキュメント"

"OPEA エンタープライズ RAG 導入プレイブック" == バージョン履歴

バージョン	日付	ドキュメントのバージョン履歴
バージョン1.0	2025年9月	初版
バージョン2.0	2026年2月	OPEA-Intel® AI for Enterprise RAG 2.0でアップデート

著作権に関する情報

商標に関する情報