



はじめに GenAI

NetApp
October 06, 2025

目次

はじめに	1
生成AIコネクタのクイックスタート	1
生成AI Connectorの要件	2
生成AIの基本要件	2
Amazon Q Business 向け NetApp コネクタの要件	2
コネクタに追加するデータソースを特定する	3
データソースの最大数	3
データソースの場所	3
サポートされるプロトコル	3
サポートされるデータソースファイル形式	4
GenAIインフラの導入	4
インフラの詳細	5
GenAIインフラの導入	6

はじめに

生成AIコネクタのクイックスタート

Amazon FSx for NetApp ONTAP ファイルシステムに存在する組織のデータを使用して、Amazon Q Business 用の NetApp コネクタの作成を開始します。コネクタを作成すると、エンドユーザーはAmazon Q Businessアシスタントにアクセスして、質問に対する組織中心の回答を得ることができます。

1

ワークロードファクトリーにログイン

必要となるのは ["Workload Factoryでアカウントを設定する"](#) 次のいずれかを使用してログインします ["コンソールエクスペリエンス"](#)。

2

生成AIの要件を満たす環境のセットアップ

AWSインフラの導入、導入および検出されたFSx for ONTAPファイルシステム、コネクタに統合するデータソースのリスト、Amazon Q Businessアプリケーションへのアクセスなどには、AWSのクレデンシャルが必要です。

["生成AI要件の詳細"](#)です。

3

データソースが格納されているFSx for ONTAPファイルシステムを特定する

コネクタに統合するデータソースは、単一のFSx for ONTAPファイルシステムに配置することも、複数のFSx for ONTAPファイルシステムに配置することもできます。これらのシステムが異なるVPCにある場合は、同じネットワーク内でアクセス可能であるか、またはAIエンジンと同じリージョンとAWSアカウントを使用してVPC間でピア関係を確立しておく必要があります。

["データソースを特定する方法"](#)です。

4

生成AIインフラの導入

インフラ導入ウィザードを起動して、AWS環境に 生成AIインフラを導入します。このプロセスでは、NetApp 生成AIエンジンのEC2インスタンスと、FSx for ONTAPファイルシステムにNetApp AIエンジンのデータベースを格納するボリュームを導入します。ボリュームは、コネクタに関する情報を格納するために使用されます。

["生成AIインフラの導入方法をご確認ください"](#)です。

次のステップ

Amazon Q Business用のコネクタを作成して、組織に焦点を当てた回答をエンドユーザーに提供できるようになりました。

生成AI Connectorの要件

Amazon Q Business 用のNetAppコネクタを作成する前に、Workload Factory と AWS が適切にセットアップされていることを確認してください。

生成AIの基本要件

生成AIには、作業を開始する前に環境が満たす必要のある一般的な要件があります。

Workload Factoryのログインとアカウント

必要となるのは ["Workload Factoryでアカウントを設定する"](#) 次のいずれかを使用してログインします ["コンソールエクスペリエンス"](#)。

AWS のクレデンシャルと権限

Workload Factory に読み取り/書き込み権限を持つ AWS 認証情報を追加する必要があります。つまり、Workload Factory を GenAI の 読み取り/書き込み モードで使用するようになります。

現時点では、_基本_モードと_読み取り専用_モードの権限はサポートされていません。

クレデンシャルを設定する際に、以下に示すように権限を選択すると、FSx for ONTAPファイルシステムの管理、GenAI EC2インスタンスおよびナレッジベースとチャットボットに必要なその他のAWSリソースの導入と管理を行うためのフルアクセスが提供されます。

["Workload FactoryにAWS認証情報を追加する方法を学ぶ"](#)

Amazon Q Business 向け NetApp コネクタの要件

Amazon Q Business Connectorの次の特定の要件を環境が満たしていることを確認します。

Amazon Q Businessアプリケーション

Amazon Q Businessアプリケーションを作成するか、既存のアプリケーションを使用する必要があります。

- アプリケーションがいずれかのAWSリージョンに存在することを確認します。
- アプリケーションにがあることを確認し ["インデックスを作成しました"](#)ます。
- アプリケーションが失敗した状態でないことを確認します。

FSx for ONTAPファイルシステム

少なくとも1つのFSx for ONTAPファイルシステムが必要です。

- NetAppの 生成AIエンジンは、コネクタに関する情報を格納するために1つのファイルシステムを使用（存在しない場合は作成）します。

このFSx for ONTAPファイルシステムでは、FlexVolボリュームを使用する必要があります。FlexGroupボリュームはサポートされません。

- 1つまたは複数のファイルシステムには、コネクタに追加するデータソースが含まれます。

1つのFSx for ONTAPファイルシステムを両方の目的に使用することも、複数のFSx for ONTAPファイ

ルシステムを使用することもできます。

- AWS FSx for ONTAPファイルシステムが配置されているAWSリージョン、VPC、サブネットを把握しておく必要があります。
- この導入に含まれるAWSリソースに適用するタグのキーと値のペアを検討する必要があります（オプション）。
- NetApp AIエンジンインスタンスに安全に接続するためのキーペア情報を知っておく必要があります。

["FSx for ONTAPファイルシステムの導入と管理の方法をご確認ください"](#)

コネクタに追加するデータソースを特定する

コネクタに統合するFSx for ONTAPファイルシステムに存在するドキュメント（データソース）を特定または作成します。これらのデータソースを使用すると、Amazon Q Businessは、組織に関連するデータに基づいて、ユーザークエリに対して正確でパーソナライズされた回答を提供できます。

データソースの最大数

サポートされるデータソースの最大数は10です。

データソースの場所

データソースは、単一のボリュームに格納することも、ボリューム内のフォルダに格納することも、Amazon FSx for NetApp ONTAPファイルシステム上のSMB共有やNFSエクスポートに格納することもできます。また、NetApp SnapMirrorデータ保護関係にあるAmazon FSx for NetApp ONTAPボリュームにデータソースを保存することもできます。

ボリュームまたはフォルダ内の個々のドキュメントを選択することはできません。したがって、データソースを含む各ボリュームまたはフォルダに、ナレッジベースと統合すべきではない無関係なドキュメントが含まれていないことを確認する必要があります。

各コネクタには複数のデータソースを追加できますが、すべてのデータソースをAWSアカウントからアクセスできるFSx for ONTAPファイルシステムに配置する必要があります。

各データソースの最大ファイルサイズは50MBです。

サポートされるプロトコル

コネクタは、NFSまたはSMB / CIFSプロトコルを使用するボリュームのデータをサポートします。SMBプロトコルを使用して保存されているファイルを選択する場合は、コネクタがこれらのボリューム上のファイルにアクセスできるように、Active Directory情報を入力する必要があります。これには、Active Directoryドメイン、IPアドレス、ユーザ名、パスワードが含まれます。

SMB経由でアクセスされる共有（ファイルまたはディレクトリ）にデータソースを格納する場合、その共有にアクセスする権限を持つチャットボットのユーザまたはグループのみがデータにアクセスできます。この「権限認識機能」が有効になっている場合、AIシステムはAuth0内のユーザのEメールを、SMB共有上のファイルの表示または使用を許可されているユーザと比較します。チャットボットは、埋め込まれたファイルのユーザ権限に基づいて回答を提供します。

たとえば、コネクタに10個のファイル(データソース)を統合し、そのうちの2個が制限された情報を含む人事ファイルである場合、これら2つのファイルへのアクセスを認証されたチャットボットユーザーのみが、それらのファイルからのデータを含むチャットボットから応答を受け取ります。



Amazon Q Business Connectorにデータソースを追加する場合、データソースファイルにはユーザー権限のみが適用されます。グループ権限は適用されません。



データソース内のファイルにテキストがない場合(テキストのない画像など)、Amazon Q Businessはインデックスを作成しませんが、テキストがないことを示すエントリをAmazon CloudWatch Logsに記録します。

サポートされるデータソースファイル形式

NetApp Connector for Amazon Q Business では現在、次のデータ ソース ファイル形式がサポートされています。

ファイル形式	エクステンション
カンマ区切り値ファイル	.csv
JSONとJSONP	.json
マークダウン	.md
Microsoft Word	.docx
プレーンテキスト	.txt
ポータブルドキュメントフォーマット	.pdf
Microsoft PowerPoint	.pptまたは.pptx
ハイパーテキストマークアップ言語	.html からのアクセスが可能です
拡張マークアップ言語	.xml に保存されます
XSLT	.XSLT
Microsoft Excel	.xls
リッチテキスト形式	.rtf

GenAIインフラの導入

組織のFSx for ONTAPナレッジベース、コネクタ、アプリケーションを構築する前に、生成AI Infrastructure for RAGフレームワークを環境に導入する必要があります。主要なインフラコンポーネントは、Amazon Bedrockサービス、NetApp生成AIエンジンの仮想マシンインスタンス、FSx for ONTAPファイルシステムです。

展開されたインフラストラクチャは、複数のナレッジベース、チャットボット、コネクタをサポートできるため、通常はこのタスクを一度だけ実行する必要があります。

インフラの詳細

GenAI環境は、Amazon Bedrockが有効になっているAWSリージョンに配置する必要があります。 ["サポートされているリージョンの一覧を表示する"](#)

インフラストラクチャは、次のコンポーネントで構成されています。

Amazon Bedrockサービス

Amazon Bedrockは、業界をリードするAI企業の基盤モデル（FMS）を単一のAPIで利用できるフルマネージドサービスです。また、セキュアなジェネレーティブAIアプリケーションの構築に必要な機能も提供します。

["Amazon Bedrockの詳細"](#)

Amazon Q Business

Amazon QはAmazon Bedrock上に構築されており、フルマネージドのジェネレーティブAIアシスタントを使用して、データソースからの情報に基づいて質問に答えたりコンテンツを生成したりできます。

["Amazon Q Businessの詳細"](#)

NetApp GenAIエンジン用の仮想マシン

このプロセスでは、NetApp GenAIエンジンがデプロイされます。データソースからデータを取り込み、そのデータをベクターデータベースに書き込む処理能力を提供します。

FSx for ONTAP ファイルシステム

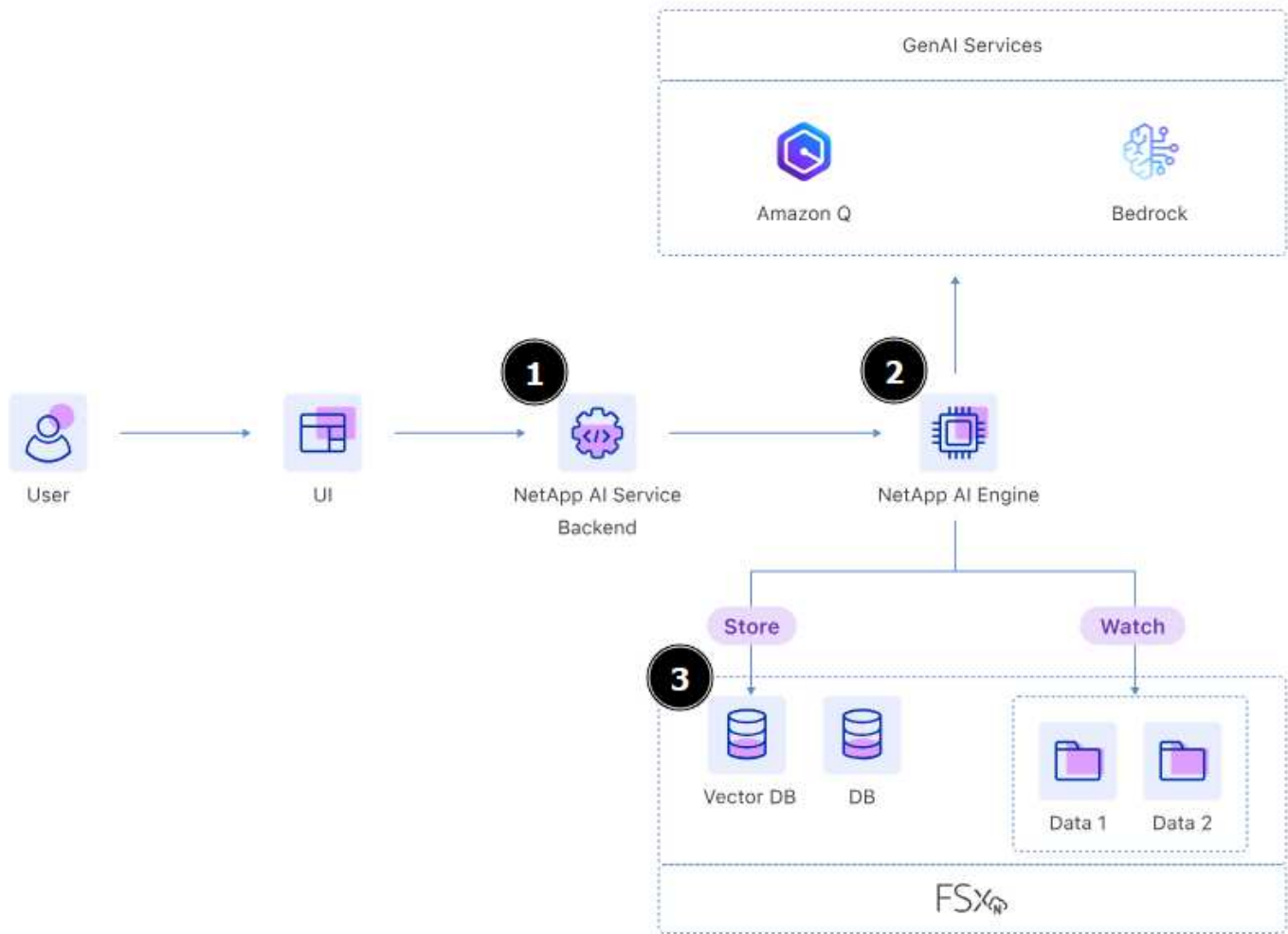
FSx for ONTAP ファイルシステムは、GenAIシステムにストレージを提供します。

データソースに基づく基本モデルによって生成されたデータを格納するベクターデータベースを含む単一のボリュームが導入されます。

ナレッジベースに統合するデータソースは、同じFSx for ONTAP ファイルシステムに配置することも、別のシステムに配置することもできます。

NetApp GenAIエンジンは、これらのボリュームの両方を監視し、相互作用します。

次の図は、GenAIインフラストラクチャを示しています。この手順では、番号1、2、3のコンポーネントを展開します。展開を開始する前に、他の要素が配置されている必要があります。



GenAIインフラの導入

AWSのクレデンシャルを入力し、FSx for ONTAPファイルシステムを選択して、Retrieval-Augmented Generation (RAG) インフラを導入する必要があります。

開始する前に

この手順を開始する前に、使用している環境がナレッジベースまたはコネクタの要件を満たしていることを確認してください。

- ["ナレッジベースの要件"](#)
- ["コネクタの要件"](#)

手順

1. 次のいずれかを使用してWorkload Factoryにログインします。"[コンソールエクスペリエンス](#)"。
2. [AI Workloads]タイルで、*[Deploy & manage]*を選択します。
3. インフラの図を確認し、* Next *を選択します。
4. [AWS settings]セクションの項目を入力します。
 - a. * AWSクレデンシャル*：AWSリソースを導入する権限を提供するAWSクレデンシャルを選択または追加します。

- b. 場所: AWSのリージョン、VPC、サブネットを選択します。

GenAI環境は、Amazon Bedrockが有効になっているAWSリージョンに配置する必要があります。"[サポートされているリージョンの一覧を表示する](#)"

5. [インフラストラクチャー設定]セクションの項目を入力します。

- a. タグ: このデプロイメントの一部であるすべてのAWS リソースに適用するタグのキーと値のペアを入力します。これらのタグは、AWS マネジメントコンソールと Workload Factory 内のインフラストラクチャ情報領域に表示され、Workload Factory リソースを追跡するのに役立ちます。

6. [Connectivity]*セクションに入力します。

- a. キーペア: NetApp GenAIエンジンインスタンスに安全に接続できるキーペアを選択します。

7. 「AIエンジン*」セクションを完了します。

- a. インスタンス名: オプションで、*インスタンス名の定義*を選択し、AI エンジン インスタンスのカスタム名を入力します。インスタンス名はAWS マネジメントコンソールと Workload Factory 内のインフラストラクチャ情報領域に表示され、Workload Factory リソースを追跡するのに役立ちます。

8. 導入*を選択して導入を開始します。



導入がクレデンシャルエラーで失敗した場合は、エラーメッセージ内のハイパーリンクを選択すると、エラーの詳細を確認できます。見つからないかブロックされている権限のリストと、生成AIワークロードが生成AIインフラを導入するために必要な権限のリストを確認できます。

結果

Workload Factory がチャットボット インフラストラクチャの展開を開始します。このプロセスには最大 10 分かかる場合があります。

導入プロセスでは、次の項目が設定されます。

- ネットワークはプライベートエンドポイントとともにセットアップされます。
- IAMロール、インスタンスプロファイル、およびセキュリティグループが作成されます。
- GenAIエンジンの仮想マシンインスタンスが導入されている。
- Amazon Bedrockは、プレフィックスを持つロググループを使用して、Amazon CloudWatch Logsにログを送信するように構成されて `aws/bedrock/` います。
- GenAIエンジンは、次の名前のロググループを使用してAmazon CloudWatch Logsにログを送信するように設定されています。 `/netapp/wlmai/<tenancyAccountId>/randomId`、どこ `<tenancyAccountId>` は "[NetAppコンソールアカウントID](#)"現在のユーザーに対して。

著作権に関する情報

Copyright © 2025 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。