



시작하십시오 GenAI

NetApp
October 06, 2025

목차

시작하십시오	1
GenAI 커넥터를 위한 빠른 시작	1
GenAI 커넥터 요구 사항	1
기본 GenAI 요구 사항	2
Amazon Q Business용 NetApp 커넥터 요구 사항	2
커넥터에 추가할 데이터 소스 식별	3
최대 데이터 원본 수입니다	3
데이터 소스의 위치입니다	3
지원되는 프로토콜	3
지원되는 데이터 소스 파일 형식	3
GenAI 인프라를 구축합니다	4
인프라 세부 정보	4
GenAI 인프라를 구축합니다	5

시작하십시오

GenAI 커넥터를 위한 빠른 시작

Amazon FSx for NetApp ONTAP 파일 시스템에 있는 조직의 데이터를 사용하여 Amazon Q Business용 NetApp Connector를 만들어 보세요. 커넥터를 생성한 후 최종 사용자는 Amazon Q Business Assistant에 액세스하여 질문에 대한 조직 중심의 응답을 받을 수 있습니다.

1

Workload Factory에 로그인하세요

당신은 필요합니다 ["Workload Factory에 계정을 설정하세요"](#) 다음 중 하나를 사용하여 로그인하세요. **"콘솔 환경"**.

2

GenAI 요구사항을 충족하도록 환경을 설정합니다

AWS 인프라, 배포되고 검색된 FSx for ONTAP 파일 시스템, 커넥터에 통합하려는 데이터 소스 목록, Amazon Q Business 애플리케이션에 대한 액세스 등을 배포하려면 AWS 자격 증명이 필요합니다.

["GenAI 요구 사항에 대해 자세히 알아보십시오"](#)..

3

데이터 소스가 포함된 **FSx for ONTAP** 파일 시스템을 식별합니다

커넥터에 통합할 데이터 원본은 단일 FSx for ONTAP 파일 시스템 또는 여러 FSx for ONTAP 파일 시스템에 있을 수 있습니다. 이러한 시스템이 서로 다른 VPC에 있는 경우 동일한 네트워크 내에서 액세스할 수 있거나 VPC를 피어링하여 AI 엔진과 동일한 지역 및 AWS 계정을 사용해야 합니다.

["데이터 소스를 식별하는 방법에 대해 알아보십시오"](#)..

4

GenAI 인프라를 구축합니다

인프라 배포 마법사를 시작하여 AWS 환경에 GenAI 인프라를 구축합니다. 이 프로세스는 NetApp GenAI 엔진용 EC2 인스턴스와 NetApp AI Engine 데이터베이스를 포함할 FSx for ONTAP 파일 시스템의 볼륨을 배포합니다. 볼륨은 커넥터에 대한 정보를 저장하는 데 사용됩니다.

["GenAI 인프라를 구축하는 방법에 대해 알아보십시오"](#)..

다음 단계

이제 Amazon Q Business용 커넥터를 만들어 최종 사용자에게 조직 중심의 응답을 제공할 수 있습니다.

GenAI 커넥터 요구 사항

Amazon Q Business용 NetApp Connector를 생성하기 전에 Workload Factory와 AWS가 올바르게 설정되었는지 확인하세요.

기본 GenAI 요구 사항

GenAI는 시작하기 전에 환경에 필요한 일반적인 요구 사항을 충족해야 합니다.

Workload Factory 로그인 및 계정

당신은 필요합니다 ["Workload Factory에 계정을 설정하세요"](#) 다음 중 하나를 사용하여 로그인하세요. ["콘솔 환경"](#) .

AWS 자격 증명 및 권한

Workload Factory에 AWS 자격 증명을 읽기/쓰기 권한으로 추가해야 합니다. 즉, GenAI에 대해 Workload Factory를 읽기/쓰기 모드로 사용하게 됩니다.

현재 기본 모드와 읽기 전용 모드 권한은 지원되지 않습니다.

자격 증명을 설정할 때 아래 표시된 권한을 선택하면 FSx for ONTAP 파일 시스템을 관리하고 기술 자료 및 챗봇에 필요한 GenAI EC2 인스턴스 및 기타 AWS 리소스를 배포 및 관리할 수 있는 모든 권한을 얻을 수 있습니다.

["Workload Factory에 AWS 자격 증명을 추가하는 방법을 알아보세요."](#)

Amazon Q Business용 NetApp 커넥터 요구 사항

환경이 아마존 Q 비즈니스 커넥터에 대한 다음과 같은 특정 요구 사항을 충족하는지 확인하십시오.

Amazon Q Business 응용 프로그램

아마존 Q 비즈니스 응용 프로그램을 생성하거나 기존 응용 프로그램을 사용해야 합니다.

- 애플리케이션이 AWS 지역 중 하나에 존재하는지 확인합니다.
- 응용 프로그램에 대해 이 ["인덱스를 만들었습니다"](#) 있는지 확인합니다.
- 응용 프로그램이 오류 상태가 아닌지 확인합니다.

FSx for ONTAP 파일 시스템

FSx for ONTAP 파일 시스템이 하나 이상 필요합니다.

- 커넥터에 대한 정보를 저장하기 위해 NetApp GenAI 엔진에서 하나의 파일 시스템을 사용(또는 없는 경우 생성)합니다.

이 FSx for ONTAP 파일 시스템은 FlexVol 볼륨을 사용해야 합니다. FlexGroup 볼륨은 지원되지 않습니다.

- 하나 이상의 파일 시스템에 커넥터에 추가할 데이터 원본이 포함됩니다.

하나의 FSx for ONTAP 파일 시스템을 두 가지 용도로 사용하거나 여러 FSx for ONTAP 파일 시스템을 사용할 수 있습니다.

- AWS FSx for ONTAP 파일 시스템이 상주하는 AWS 지역, VPC 및 서브넷을 알아야 합니다.
- 이 배포의 일부인 AWS 리소스에 적용할 태그 키/값 쌍을 고려해야 합니다(선택 사항).
- NetApp AI 엔진 인스턴스에 안전하게 연결할 수 있는 키 쌍 정보를 알아야 합니다.

["FSx for ONTAP 파일 시스템을 구축 및 관리하는 방법에 관해 알아보십시오"](#)

커넥터에 추가할 데이터 소스 식별

커넥터에 통합할 FSx for ONTAP 파일 시스템에 있는 문서(데이터 소스)를 식별하거나 생성합니다. 이러한 데이터 소스를 통해 Amazon Q Business는 조직과 관련된 데이터를 기반으로 사용자 쿼리에 정확하고 맞춤화된 답변을 제공할 수 있습니다.

최대 데이터 원본 수입니다

지원되는 최대 데이터 원본 수는 10개입니다.

데이터 소스의 위치입니다

데이터 소스는 Amazon FSx for NetApp ONTAP 파일 시스템의 SMB 공유 또는 NFS 내보내기의 단일 볼륨 또는 볼륨 내의 폴더에 저장될 수 있습니다. 데이터 소스는 NetApp SnapMirror 데이터 보호 관계에 있는 Amazon FSx for NetApp ONTAP 볼륨에 저장할 수도 있습니다.

볼륨이나 폴더 내에서 개별 문서를 선택할 수 없으므로 데이터 원본이 포함된 각 볼륨이나 폴더에 기술 문서와 통합되지 않아야 하는 관련 문서가 포함되어 있지 않은지 확인해야 합니다.

각 커넥터에 여러 데이터 소스를 추가할 수 있지만, 모두 AWS 계정에서 액세스할 수 있는 FSx for ONTAP 파일 시스템에 상주해야 합니다.

각 데이터 원본의 최대 파일 크기는 50MB입니다.

지원되는 프로토콜

커넥터는 NFS 또는 SMB/CIFS 프로토콜을 사용하는 볼륨의 데이터를 지원합니다. SMB 프로토콜을 사용하여 저장된 파일을 선택할 때 커넥터가 해당 볼륨의 파일에 액세스할 수 있도록 Active Directory 정보를 입력해야 합니다. 여기에는 Active Directory 도메인, IP 주소, 사용자 이름 및 암호가 포함됩니다.

SMB를 통해 액세스되는 공유(파일 또는 디렉토리)에 데이터 소스를 저장하는 경우, 해당 공유에 액세스할 수 있는 권한이 있는 챗봇 사용자 또는 그룹만 데이터에 액세스할 수 있습니다. 이 "권한 인식 기능"이 활성화되면 AI 시스템은 auth0의 사용자 이메일을 SMB 공유에서 파일을 보거나 사용할 수 있는 사용자와 비교합니다. 챗봇은 포함된 파일에 대한 사용자 권한을 기반으로 해답을 제공합니다.

예를 들어, 10개의 파일(데이터 소스)을 커넥터에 통합하고 2개의 파일이 제한된 정보를 포함하는 인적 리소스 파일인 경우, 이 두 파일에 액세스할 수 있도록 인증된 챗봇 사용자만 이러한 파일의 데이터를 포함하는 챗봇으로부터 응답을 받게 됩니다.



Amazon Q Business 커넥터에 데이터 원본을 추가하면 데이터 원본 파일에 사용자 권한만 적용됩니다. 그룹 권한이 적용되지 않습니다.



데이터 원본의 파일에 텍스트가 없는 경우(예: 텍스트가 없는 이미지) Amazon Q Business는 해당 파일을 인덱싱하지 않고 텍스트가 없음을 나타내는 항목을 Amazon CloudWatch 로그에 기록합니다.

지원되는 데이터 소스 파일 형식

현재 Amazon Q Business용 NetApp Connector에서 지원되는 데이터 소스 파일 형식은 다음과 같습니다.

파일 형식	연장
쉼표로 구분된 값 파일입니다	.csv입니다
JSON 및 JSONP	제이슨
마크다운	진료 부서
Microsoft Word를	.docx입니다
일반 텍스트	.txt입니다
휴대용 문서 형식	PDF로 이동합니다
Microsoft PowerPoint를 클릭합니다	.ppt 또는 .pptx
하이퍼텍스트 마크업 언어	.html을 참조하십시오
확장 가능한 마크업 언어	XML
XSLT를 참조하십시오	.XSLT입니다
Microsoft Excel을 선택합니다	xls입니다
서식 있는 텍스트	.RTF를 클릭합니다

GenAI 인프라를 구축합니다

조직을 위해 FSx for ONTAP 지식 베이스, 커넥터 및 애플리케이션을 구축하기 전에 환경에 RAG 프레임워크용 GenAI 인프라를 구축해야 합니다. 기본 인프라 구성요소는 Amazon Bedrock 서비스, NetApp GenAI 엔진의 가상 머신 인스턴스 및 FSx for ONTAP 파일 시스템입니다.

구축된 인프라는 여러 지식 베이스, 챗봇, 커넥터를 지원할 수 있으므로 일반적으로 이 작업은 한 번만 수행하면 됩니다.

인프라 세부 정보

GenAI 배포는 Amazon Bedrock이 지원되는 AWS 지역에 있어야 합니다. ["지원되는 지역 목록을 봅니다"](#)

인프라는 다음과 같은 구성 요소로 이루어집니다.

아마존 **Bedrock** 서비스

Amazon Bedrock은 단일 API를 통해 선도적인 AI 회사의 기반 모델(FMS)을 사용할 수 있는 완전 관리형 서비스입니다. 또한 안전한 생성 AI 애플리케이션을 구축하는 데 필요한 기능을 제공합니다.

["아마존 Bedrock에 대해 자세히 알아보십시오"](#)

아마존 **Q** 비즈니스

Amazon Q는 Amazon Bedrock을 기반으로 구축되어 질문에 답하고 데이터 소스의 정보를 기반으로 콘텐츠를 생성하는 데 사용할 수 있는 완전 관리형 생성 AI 도우미를 제공합니다.

["아마존 Q 비즈니스에 대해 자세히 알아보십시오"](#)

NetApp GenAI 엔진용 가상 머신입니다

이 프로세스 중에 NetApp GenAI 엔진이 배포됩니다. 데이터 소스에서 데이터를 수집한 다음 해당 데이터를 벡터 데이터베이스에 쓸 수 있는 처리 능력을 제공합니다.

FSx for ONTAP 파일 시스템

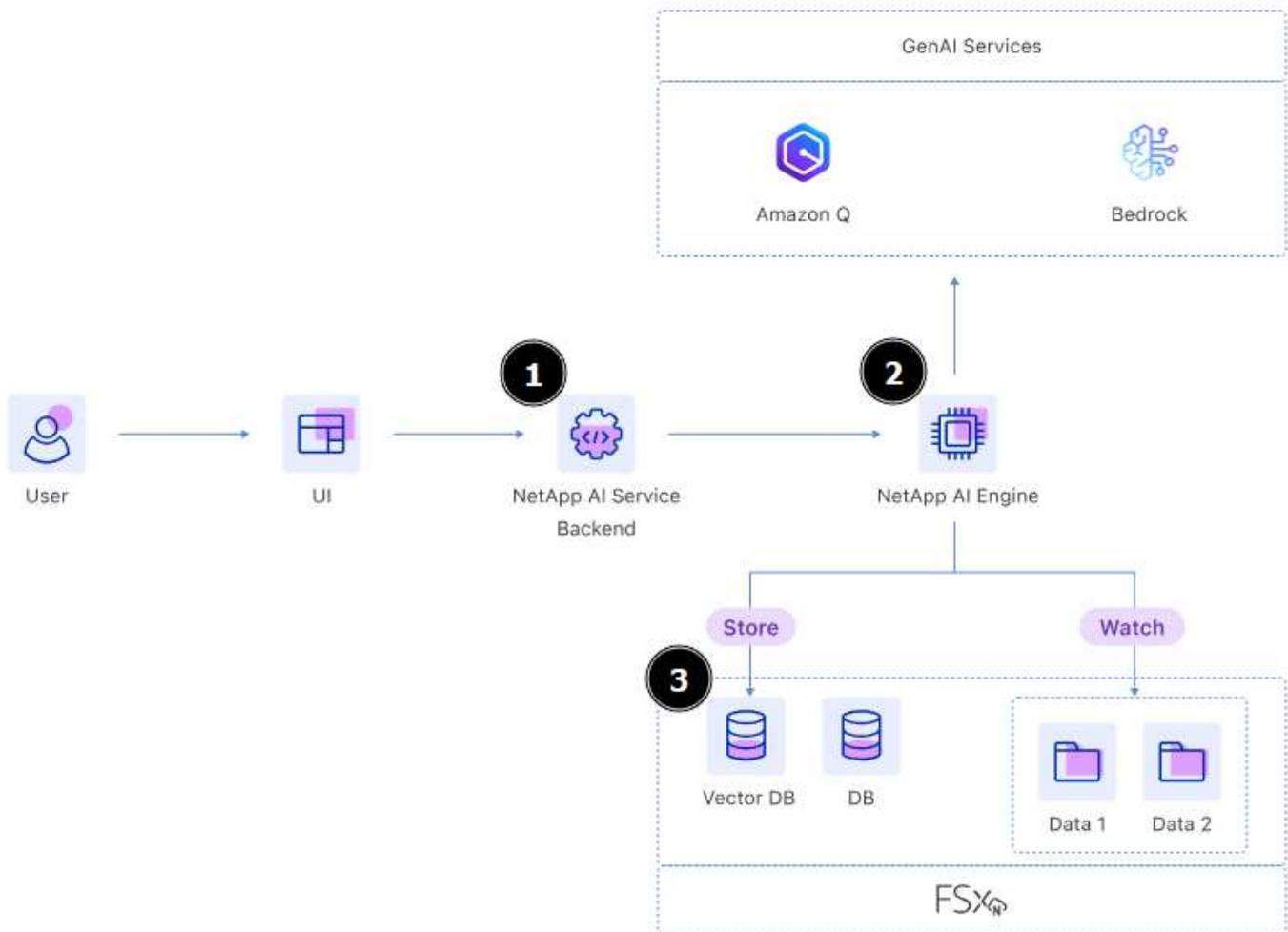
FSx for ONTAP 파일 시스템은 GenAI 시스템을 위한 스토리지를 제공합니다.

데이터 소스를 기반으로 기본 모델에 의해 생성된 데이터를 저장하는 벡터 데이터베이스를 포함하는 단일 볼륨이 배포됩니다.

기술 자료에 통합할 데이터 소스는 동일한 FSx for ONTAP 파일 시스템 또는 다른 시스템에 상주할 수 있습니다.

NetApp GenAI 엔진은 이 두 볼륨을 모두 모니터링하고 상호 작용합니다.

다음 이미지는 GenAI 인프라를 보여 줍니다. 이 절차를 수행하는 동안 번호가 1, 2, 3인 구성 요소가 전개됩니다. 배포를 시작하기 전에 다른 요소가 있어야 합니다.



GenAI 인프라를 구축합니다

AWS 자격 증명을 입력하고 FSx for ONTAP 파일 시스템을 선택하여 검색 증강 생성(RAG) 인프라를 배포해야 합니다.

시작하기 전에

이 절차를 시작하기 전에 사용자의 환경이 지식 베이스 또는 커넥터의 요구 사항을 충족하는지 확인하십시오.

- "기술 자료 요구 사항"
- "커넥터 요구 사항"

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요. "콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 인프라 다이어그램을 검토하고 * Next * 를 선택합니다.
4. AWS 설정 * 섹션의 항목을 완료합니다.
 - a. * AWS 자격 증명 *: AWS 리소스 배포 권한을 제공하는 AWS 자격 증명을 선택하거나 추가합니다.
 - b. * 위치 *: AWS 지역, VPC 및 서브넷을 선택합니다.

GenAI 배포는 Amazon Bedrock이 활성화된 AWS 지역에 있어야 합니다. "지원되는 지역 목록을 봅니다"

5. 인프라 설정 * 섹션의 항목을 완료합니다.
 - a. 태그: 이 배포에 포함된 모든 AWS 리소스에 적용할 태그 키/값 쌍을 입력합니다. 이러한 태그는 AWS Management Console과 Workload Factory 내의 인프라 정보 영역에서 볼 수 있으며, Workload Factory 리소스를 추적하는 데 도움이 될 수 있습니다.
6. Connectivity * 섹션을 완료합니다.
 - a. * 키 쌍 * : NetApp GenAI 엔진 인스턴스에 안전하게 연결할 수 있는 키 쌍을 선택하십시오.
7. AI 엔진 * 섹션을 완료하십시오.
 - a. 인스턴스 이름: 선택적으로 *인스턴스 이름 정의*를 선택하고 AI 엔진 인스턴스에 대한 사용자 지정 이름을 입력합니다. 인스턴스 이름은 AWS Management Console과 Workload Factory 내의 인프라 정보 영역에 표시되며, Workload Factory 리소스를 추적하는 데 도움이 됩니다.
8. 배포 * 를 선택하여 배포를 시작합니다.



자격 증명 오류로 인해 배포가 실패하는 경우 오류 메시지 내에서 하이퍼링크를 선택하여 오류 세부 정보를 확인할 수 있습니다. 누락 또는 차단된 권한 목록과 GenAI 워크로드를 배포하기 위해 필요한 권한 목록을 확인할 수 있습니다.

결과

Workload Factory가 첫 번째 인프라 구축을 시작합니다. 이 과정은 최대 10분이 걸릴 수 있습니다.

배포 프로세스 중에 다음 항목이 설정됩니다.

- 네트워크는 전용 끝점과 함께 설정됩니다.
- IAM 역할, 인스턴스 프로필 및 보안 그룹이 생성됩니다.
- GenAI 엔진의 가상 머신 인스턴스가 배포됩니다.
- Amazon Bedrock은 접두사가 있는 로그 그룹을 사용하여 Amazon CloudWatch 로그에 로그를 보내도록 구성되어 `aws/bedrock/` 있습니다.
- GenAI 엔진은 이름이 있는 로그 그룹을 사용하여 Amazon CloudWatch Logs에 로그를 보내도록 구성됩니다. `/netapp/wlmai/<tenancyAccountId>/randomId` , 어디 `<tenancyAccountID>` 이다 "NetApp 콘솔"

계정 ID" 현재 사용자에게 대해서.

저작권 정보

Copyright © 2025 NetApp, Inc. All Rights Reserved. 미국에서 인쇄된 본 문서의 어떠한 부분도 저작권 소유자의 사전 서면 승인 없이는 어떠한 형식이나 수단(복사, 녹음, 녹화 또는 전자 검색 시스템에 저장하는 것을 비롯한 그래픽, 전자적 또는 기계적 방법)으로도 복제될 수 없습니다.

NetApp이 저작권을 가진 자료에 있는 소프트웨어에는 아래의 라이선스와 고지사항이 적용됩니다.

본 소프트웨어는 NetApp에 의해 '있는 그대로' 제공되며 상품성 및 특정 목적에의 적합성에 대한 명시적 또는 묵시적 보증을 포함하여(이에 제한되지 않음) 어떠한 보증도 하지 않습니다. NetApp은 대체품 또는 대체 서비스의 조달, 사용 불능, 데이터 손실, 이익 손실, 영업 중단을 포함하여(이에 국한되지 않음), 이 소프트웨어의 사용으로 인해 발생하는 모든 직접 및 간접 손해, 우발적 손해, 특별 손해, 징벌적 손해, 결과적 손해의 발생에 대하여 그 발생 이유, 책임론, 계약 여부, 엄격한 책임, 불법 행위(과실 또는 그렇지 않은 경우)와 관계없이 어떠한 책임도 지지 않으며, 이와 같은 손실의 발생 가능성이 통지되었다 하더라도 마찬가지입니다.

NetApp은 본 문서에 설명된 제품을 언제든지 예고 없이 변경할 권리를 보유합니다. NetApp은 NetApp의 명시적인 서면 동의를 받은 경우를 제외하고 본 문서에 설명된 제품을 사용하여 발생하는 어떠한 문제에도 책임을 지지 않습니다. 본 제품의 사용 또는 구매의 경우 NetApp에서는 어떠한 특허권, 상표권 또는 기타 지적 재산권이 적용되는 라이선스도 제공하지 않습니다.

본 설명서에 설명된 제품은 하나 이상의 미국 특허, 해외 특허 또는 출원 중인 특허로 보호됩니다.

제한적 권리 표시: 정부에 의한 사용, 복제 또는 공개에는 DFARS 252.227-7013(2014년 2월) 및 FAR 52.227-19(2007년 12월)의 기술 데이터-비상업적 품목에 대한 권리(Rights in Technical Data -Noncommercial Items) 조항의 하위 조항 (b)(3)에 설명된 제한사항이 적용됩니다.

여기에 포함된 데이터는 상업용 제품 및/또는 상업용 서비스(FAR 2.101에 정의)에 해당하며 NetApp, Inc.의 독점 자산입니다. 본 계약에 따라 제공되는 모든 NetApp 기술 데이터 및 컴퓨터 소프트웨어는 본질적으로 상업용이며 개인 비용만으로 개발되었습니다. 미국 정부는 데이터가 제공된 미국 계약과 관련하여 해당 계약을 지원하는 데에만 데이터에 대한 전 세계적으로 비독점적이고 양도할 수 없으며 재사용이 불가능하며 취소 불가능한 라이선스를 제한적으로 가집니다. 여기에 제공된 경우를 제외하고 NetApp, Inc.의 사전 서면 승인 없이는 이 데이터를 사용, 공개, 재생산, 수정, 수행 또는 표시할 수 없습니다. 미국 국방부에 대한 정부 라이선스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

NETAPP, NETAPP 로고 및 <http://www.netapp.com/TM>에 나열된 마크는 NetApp, Inc.의 상표입니다. 기타 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다.