관리 및 모니터링

GenAI

NetApp
October 06, 2025

This PDF was generated from https://docs.netapp.com/ko-kr/workload-genai/general/manage-infrastructure.html on October 06, 2025. Always check docs.netapp.com for the latest.

목차

관리 및 모니터링 . . . 1
GenAI 인프라를 관리합니다 . . . 1
인프라에 대한 정보를 봅니다 . . . 1
인프라를 제거합니다 . . . 1
GenAI 지식 기반 관리 . . . 2
기술 문서에 대한 정보를 봅니다 . . . 2
기술 문서를 편집합니다 . . . 2
스냅샷으로 기술 자료 보호 . . . 3
기술 문서에 데이터 원본을 추가합니다 . . . 5
데이터 원본을 기술 문서와 동기화합니다 . . . 9
기술 문서를 생성하기 전에 채팅 모델을 평가합니다 . . . 10
기술 문서 게시를 취소합니다 . . . 10
기술 문서를 삭제합니다 . . . 11
Amazon Q Business 커넥터를 관리합니다 . . . 11
커넥터에 대한 정보를 봅니다 . . . 11
연결선을 편집합니다 . . . 11
커넥터에 데이터 원본을 추가합니다 . . . 12
데이터 원본을 커넥터와 동기화합니다 . . . 16
연결선을 삭제합니다 . . . 17
GenAI 데이터 소스를 관리합니다 . . . 17
데이터 원본에 대한 정보를 봅니다 . . . 17
데이터 원본 설정을 편집합니다 . . . 18
기존 데이터 원본의 내용을 업데이트합니다 . . . 18
데이터 원본을 삭제합니다 . . . 18
NetApp Workload Factory의 Tracker를 사용하여 워크로드 작업 모니터링 . . . 19
작업을 추적하고 모니터링합니다 . . . 19
API 요청을 봅니다 . . . 20
실패한 작업을 다시 시도하십시오 . . . 20
실패한 작업을 편집한 후 다시 시도하십시오 . . . 20

관리 및 모니터링

GenAI 인프라를 관리합니다

구축된 GenAI RAG 인프라에 대한 세부 정보를 보거나 더 이상 필요하지 않을 경우 챗봇 인프라를 제거할 수 있습니다.

인프라에 대한 정보를 봅니다

챗봇 인프라에 대한 정보를 볼 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 인프라 메뉴를 선택하세요.
4. 다음 구성 요소에 대한 세부 정보가 포함된 인프라에 대한 정보를 봅니다.
    - AWS 설정
    - 인프라 설정
    - 있습니다
    - 벡터 데이터베이스

인프라를 제거합니다

하나 이상의 챗봇에 배포한 챗봇 인프라가 더 이상 필요하지 않으면 Workload Factory에서 제거할 수 있습니다.

이 인프라에 배포된 모든 챗봇이 비활성화되고 모든 채팅 기록이 삭제됩니다.

이 작업은 Workload Factory에서 AI 인프라에 대한 링크만 제거할 뿐, AWS에서 모든 구성 요소를 제거하지는 않습니다. AWS에서 다음 인프라 구성 요소를 수동으로 삭제해야 합니다.

- VM 인스턴스
- 전용 끝점
- AI 데이터베이스가 포함된 FSx for ONTAP 파일 시스템의 볼륨입니다
- IAM 역할
- 정책에 동의하게 됩니다
- 보안 그룹입니다

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 인프라 메뉴를 선택하세요.

4. 를 ••• 선택하고 * 챗봇 인프라 제거 * 를 선택합니다.
5. 인프라를 삭제할 것인지 확인하고 * 제거 * 를 선택합니다.

결과

챗봇 인프라 구성 요소가 Workload Factory에서 제거되었습니다.

GenAI 지식 기반 관리

기술 문서를 만든 후에는 기술 문서 세부 정보를 보거나 기술 문서를 수정하거나 추가 데이터 원본을 통합하거나 기술 문서를 삭제할 수 있습니다.

기술 문서에 대한 정보를 봅니다

통합된 기술 자료 및 데이터 원본에 대한 설정에 대한 정보를 볼 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 보려는 기술 문서를 선택합니다.

   정의된 경우 현재 사용 중인 대화 시작점이 오른쪽 창에 표시됩니다.

4. 기술 문서 세부 정보를 보려면 * 기술 자료 관리 * 를 선택하고 ••• 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

   작업 * 메뉴를 사용하면 필요한 경우 기술 자료를 관리할 수 있습니다.

기술 문서를 편집합니다

일부 설정을 변경하여 기술 자료를 업데이트하거나 데이터 원본을 추가 또는 제거할 수 있습니다.

기술 문서에서 데이터 원본을 추가, 수정 또는 제거할 때마다 데이터 원본이 기술 문서에 다시 인덱싱되도록 데이터 원본을 동기화해야 합니다. 동기화는 증분 방식이므로 Amazon Bedrock은 마지막 동기화 이후 추가, 수정 또는 삭제된 FSx for ONTAP 볼륨의 개체만 처리합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 업데이트할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

5. 조치 * 메뉴를 선택하고 * 기술 문서 편집 * 을 선택합니다.

6. 지식 기반 편집 페이지에서는 지식 기반 이름, 설명, 내장 모델, 채팅 모델, 기능 활성화를 변경할 수 있으며, 대화 시작자를 자동으로 생성할지 수동으로 생성할지 선택할 수 있으며, 지식 기반이 포함된 볼륨에 사용되는 스냅샷 정책을 선택할 수 있습니다.

   대화를 시작하는 데 수동 모드를 사용하는 경우 여기에서 대화 시작도 변경할 수 있습니다.

   임베딩, 비용 등 모든 기술 자료 검사 기술 문서를 만든 후 데이터 가드레일을 활성화하면 기술 문서가 다시 검색되어 비용이 발생합니다. 마찬가지로 채팅 모델을 변경하면 GenAI가 관련 데이터 소스를 다시 스캔합니다(비용 발생).

7. 변경한 후 * 저장 * 을 선택합니다.

스냅샷으로 기술 자료 보호

기술 자료 볼륨의 스냅샷을 생성하고 복원하여 기술 자료 데이터를 보호할 수 있습니다. 언제든지 스냅샷에서 복원하여 이전 버전의 기술 자료로 되돌릴 수 있습니다.

스냅샷은 백업보다 더 빠르고 스토리지 효율적이며, 다른 보호 정책을 사용하여 각 기술 자료를 보호할 수 있습니다. 스냅샷이 유용한 시나리오는 다음과 같습니다.

- 우발적인 데이터 손실 또는 손상
- 기술 문서에 수집되는 잘못된 데이터를 복구하는 중입니다
- 다양한 데이터 소스 또는 청킹 전략을 테스트하고 테스트가 완료되면 신속하게 되돌릴 수 있습니다

기술 자료 볼륨의 스냅샷을 생성합니다

기술 자료 볼륨의 수동 스냅샷을 생성하여 기술 문서의 상태를 저장할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 보호할 기술 자료를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

5. Actions * 메뉴를 선택하고 * Snapshot > Create new snapshot * 을 선택합니다.
6. 필요에 따라 * 스냅샷 이름 정의 * 를 선택하고 스냅샷에 대한 사용자 정의 이름을 입력합니다.

   사용자 지정 이름을 정의하면 나중에 복원해야 할 경우 스냅샷의 내용을 더 잘 확인할 수 있습니다.

7. Create * 를 선택합니다.

   기술 자료의 스냅샷이 생성됩니다.

기술 자료 볼륨의 스냅샷을 복원합니다

언제든지 기술 자료 볼륨의 수동 또는 예약 스냅샷을 복원할 수 있습니다.

> 볼륨에 저장된 데이터베이스가 손상되었거나 삭제된 경우 Generative AI 워크로드 UI를 사용하여 스냅샷을 복원할 수 없습니다. 이 문제를 해결하려면 볼륨이 호스팅되는 ONTAP 클러스터에서 을 사용하여 스냅샷을 복원할 수 "ONTAP CLI 를 참조하십시오" 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 복원할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

5. Actions * 메뉴를 선택하고 * Snapshot > Restore snapshot * 을 선택합니다.

   스냅샷 선택 대화 상자가 나타나고 이 기술 자료에 대해 생성된 스냅샷 목록을 볼 수 있습니다.

6. (선택 사항) 스냅샷이 복원된 후 예약 및 현재 실행 중인 데이터 소스 스캔을 계속하려면 * Pause running and scheduled scans after the snapshot Restore * (스냅숏 복원 후 실행 및 예약된 스캔 일시 중지) 옵션을 선택 취소합니다.

   이 옵션은 기본적으로 활성화되어 있으므로 기술 문서가 부분적으로 복원된 상태일 때 검사가 수행되지 않거나 검사가 새로 복원된 기술 문서를 이전 데이터로 업데이트하지 않습니다.

7. 목록에서 복구할 스냅샷을 선택합니다.
8. Restore * 를 선택합니다.

기술 문서를 복제합니다

기술 자료 스냅샷에서 새로운 기술 자료를 생성할 수 있습니다. 이 기능은 원본 기술 문서가 손상되었거나 손실된 경우에 유용합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 복원할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

5. Actions * 메뉴를 선택하고 * Snapshot > Clone Knowledge Base * 를 선택합니다.

   클론 대화 상자가 나타납니다.

6. 스냅숏이 클론 생성된 후 예약 및 현재 실행 중인 데이터 원본 스캔을 계속하려면 * 스냅숏 클론 생성 후 실행 및

예약된 검사 일시 중지 * 옵션을 선택 취소합니다.

이 옵션은 기본적으로 활성화되어 있으므로 기술 문서가 부분적으로 복원된 상태일 때 검사가 수행되지 않거나 검사가 새로 복원된 기술 문서를 이전 데이터로 업데이트하지 않습니다.

7. 목록에서 복제할 스냅샷을 선택합니다.
8. Continue * 를 선택합니다.
9. 새 기술 문서의 이름을 입력합니다.
10. 새 기술 자료에 사용할 파일 시스템 SVM 및 볼륨 이름을 선택합니다.
11. 클론 * 을 선택합니다.

기술 문서에 데이터 원본을 추가합니다

추가 데이터 원본을 기술 문서에 포함시켜 추가 조직 데이터로 채울 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 데이터 소스를 추가할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 데이터 원본 추가 * 를 선택합니다.
5. 추가하려는 데이터 소스 유형을 선택하세요.
   - ONTAP 파일 시스템용 FSx 추가(기존 ONTAP 볼륨용 FSx의 파일 사용)
   - 파일 시스템 추가(일반 SMB 또는 NFS 공유의 파일 사용)

ONTAP 파일 시스템에 **FSx** 추가

1. * 파일 시스템 선택 *: 데이터 소스 파일이 있는 FSx for ONTAP 파일 시스템을 선택하고 * 다음 * 을 선택합니다.
2. * 볼륨 선택 *: 데이터 원본 파일이 있는 볼륨을 선택하고 * 다음 * 을 선택합니다.

   SMB 프로토콜을 사용하여 저장된 파일을 선택할 때 도메인, IP 주소, 사용자 이름 및 암호를 포함한 Active Directory 정보를 입력해야 합니다.

3. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.
4. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.

   ◦ * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.

   ▪ * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.

   ▪ * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

   > ⓘ 높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   ◦ * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.

   ▪ 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

   이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

   이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든 PII가 GenAI 데이터베이스에 마스킹됩니다.

   > ⓘ 이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

   ◦ 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

   > ⓘ 수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

5. 선택한 데이터 원본이 SMB 프로토콜을 사용하는 볼륨에 있을 때만 사용할 수 있는 * 권한 인식 * 섹션에서 권한 인식 응답을 활성화하거나 비활성화할 수 있습니다.

- 사용: 이 기술 자료에 액세스하는 챗봇 사용자는 액세스 권한이 있는 데이터 원본에서 쿼리에 대한 응답만 받습니다.
- * 사용 안 함 * : 챗봇 사용자는 모든 통합 데이터 소스의 콘텐츠를 사용하여 응답을 받습니다.

6. 이 데이터 소스를 기술 문서에 추가하려면 * 추가 * 를 선택하십시오.

일반 **NFS** 파일 시스템 추가

1. 파일 시스템 선택: 데이터 소스 파일이 있는 파일 시스템 호스트의 IP 주소 또는 FQDN을 입력하고, 네트워크 공유에 대한 NFS 프로토콜을 선택하고 *다음*을 선택합니다.
2. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.

> 경우에 따라 NFS 내보내기 이름을 직접 입력하고 *디렉터리 검색*을 선택하여 사용 가능한 디렉터리를 표시해야 할 수도 있습니다. 내보내기 전체 또는 내보내기에서 특정 폴더만 선택할 수 있습니다.

3. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.
   - * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.
     - * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.
     - * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

     > 높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   - * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.
     - 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

       이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

     이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든 PII가 GenAI 데이터베이스에 마스킹됩니다.

     > 이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

   - 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

4. *데이터 소스 추가*를 선택하여 이 데이터 소스를 지식 기반에 추가하세요.

일반 **SMB** 파일 시스템 추가

1. 파일 시스템 선택:
   a. 데이터 소스 파일이 있는 파일 시스템 호스트의 IP 주소나 FQDN을 입력하세요.
   b. 네트워크 공유에 SMB 프로토콜을 선택합니다.
   c. 도메인, IP 주소, 사용자 이름, 비밀번호 등 Active Directory 정보를 입력합니다.
   d. 다음 * 을 선택합니다.
2. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.

   경우에 따라 SMB 공유 이름을 직접 입력하고 *디렉터리 검색*을 선택하여 사용 가능한 디렉터리를 표시해야 할 수도 있습니다. 전체 공유를 선택하거나 공유에서 특정 폴더만 선택할 수 있습니다.

3. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.
   - * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.
     - * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.
     - * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

       높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   - 권한 인식: 권한 인식 응답을 활성화하거나 비활성화합니다.
     - 사용: 이 기술 자료에 액세스하는 챗봇 사용자는 액세스 권한이 있는 데이터 원본에서 쿼리에 대한 응답만 받습니다.
     - * 사용 안 함 * : 챗봇 사용자는 모든 통합 데이터 소스의 콘텐츠를 사용하여 응답을 받습니다.
   - * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.
     - 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

       이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

   이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든

PII가 GenAI 데이터베이스에 마스킹됩니다.

> 이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

◦ 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

> 수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

4. *데이터 소스 추가*를 선택하여 이 데이터 소스를 지식 기반에 추가하세요.

결과

데이터 원본은 기술 자료에 통합됩니다.

데이터 원본을 기술 문서와 동기화합니다

데이터 소스는 하루에 한 번 관련 기술 자료와 자동으로 동기화되므로 데이터 소스 변경 사항이 챗봇에 반영됩니다. 데이터 원본을 변경하고 데이터를 즉시 동기화하려는 경우 필요 시 동기화를 수행할 수 있습니다.

동기화는 증분 동기화이므로 Amazon Bedrock은 마지막 동기화 이후 추가, 수정 또는 삭제된 데이터 원본의 객체만 처리합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 동기화할 기술 자료를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.
5. 조치 * 메뉴를 선택하고 * 지금 스캔 * 을 선택합니다.

   데이터 원본을 스캔한다는 메시지와 검사가 완료되면 최종 메시지가 표시됩니다.

결과

기술 자료는 첨부된 데이터 원본과 동기화되며 활성 챗봇은 데이터 원본의 최신 정보를 사용하기 시작합니다.

예약된 동기화를 일시 중지하거나 다시 시작합니다

데이터 원본의 다음 동기화(스캔)를 일시 중지하거나 다시 시작하려면 언제든지 다시 시작할 수 있습니다. 데이터 원본을 변경하고 변경 기간 동안 동기화를 실행하지 않으려면 다음 예약된 동기화를 일시 중지해야 할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.

3. 지식 기반 및 커넥터 메뉴에서 스캔을 일시 중지하거나 다시 시작할 지식 기반을 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.
5. Actions * 메뉴를 선택하고 * Scan > Pause Scheduled Scan * 또는 * Scan > Resume Scheduled Scan * 을 선택합니다.

   다음 예약된 스캔이 일시 중지되었거나 다시 시작되었다는 메시지가 표시됩니다.

기술 문서를 생성하기 전에 채팅 모델을 평가합니다

기술 자료를 작성하기 전에 사용 가능한 기본 채팅 모델을 평가할 수 있으므로 구현에 가장 적합한 모델을 확인할 수 있습니다. 모델 지원은 AWS 지역에 따라 다르므로, 기술 자료를 배포할 계획이 있는 지역에서 사용할 수 있는 모델을 확인하려면 을 참조하십시오 "이 AWS 설명서 페이지" .

> 이 기능은 Knowledge Base가 생성되지 않은 경우(Knowledge Base 인벤토리 페이지에 Knowledge Base가 없는 경우에만) 사용할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 Chatbot 페이지 오른쪽에 채팅 모델을 선택하는 옵션이 표시됩니다.
4. 목록에서 채팅 모델을 선택하고 프롬프트 영역에 질문 집합을 입력하여 챗봇이 어떻게 응답하는지 확인합니다.
5. 여러 모델을 시도하여 구현에 가장 적합한 모델을 확인하십시오.

결과

기술 문서를 작성할 때 이 채팅 모델을 사용하십시오.

기술 문서 게시를 취소합니다

챗봇 애플리케이션과 통합될 수 있도록 기술 자료를 게시한 후, 챗봇 애플리케이션이 기술 자료에 액세스하지 못하도록 하려면 게시를 취소할 수 있습니다.

기술 문서의 게시를 취소하면 모든 채팅 응용 프로그램이 작동하지 않습니다. 기술 자료에 액세스할 수 있는 고유 API 끝점이 비활성화됩니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 게시를 취소할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드 및 포함된 모든 데이터 원본의 목록이 표시됩니다.

5. 작업 * 메뉴를 선택하고 * 게시 취소 * 를 선택합니다.

결과

기술 문서가 비활성화되고 챗봇 애플리케이션에서 더 이상 액세스할 수 없습니다.

기술 문서를 삭제합니다

지식 기반이 더 이상 필요하지 않으면 삭제할 수 있습니다. 지식 기반을 삭제하면 Workload Factory에서도 제거되고, 해당 지식 기반이 포함된 볼륨도 삭제됩니다. 지식 기반을 사용하는 모든 애플리케이션이나 챗봇은 작동을 멈춥니다. 지식 기반을 삭제하면 되돌릴 수 없습니다.

기술 문서를 삭제할 때 기술 문서와 관련된 모든 상담원과의 연결을 해제하여 기술 문서와 연결된 모든 리소스를 완전히 삭제해야 합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 삭제할 기술 문서를 선택합니다.
4. 를 ••• 선택하고 * 기술 자료 관리 * 를 선택합니다.
5. 조치 * 메뉴를 선택하고 * 기술 문서 삭제 * 를 선택합니다.
6. 기술 자료 삭제 대화 상자에서 삭제할 내용을 확인하고 * 삭제 * 를 선택합니다.

결과

지식 기반이 Workload Factory에서 제거되고 관련 볼륨이 삭제됩니다.

Amazon Q Business 커넥터를 관리합니다

Amazon Q Business용 커넥터를 만든 후에는 커넥터 세부 정보를 보거나 커넥터를 수정하거나 추가 데이터 원본을 통합하거나 커넥터를 삭제할 수 있습니다.

커넥터에 대한 정보를 봅니다

통합된 커넥터 및 데이터 원본에 대한 설정 정보를 볼 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 지식 기반 및 커넥터 인벤토리 페이지에서 보려는 커넥터를 선택합니다.
4. 커넥터 세부 정보를 보려면 * 커넥터 관리 * 를 선택하고 ••• 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

   조치 * 메뉴를 사용하면 커넥터를 원하는 경우 관리할 수 있습니다.

연결선을 편집합니다

일부 설정을 변경하여 연결선을 업데이트하거나 데이터 원본을 추가 또는 제거할 수 있습니다.

커넥터에서 데이터 원본을 추가, 수정 또는 제거할 때마다 GenAI는 다시 인덱싱되도록 데이터 원본 정보를 Amazon Q Business로 전송해야 합니다. 동기화는 증분 방식이므로 Amazon Q Business는 마지막 동기화 이후 추가, 수정 또는 삭제된 FSx for ONTAP 볼륨의 객체만 처리합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base & Connectors 인벤토리 페이지에서 업데이트할 커넥터를 선택합니다.
4. 을 선택하고 * 커넥터 관리 * 를 선택합니다.

   이 페이지에는 게시된 상태, 데이터 원본의 포함 상태, 포함 모드, 포함된 모든 데이터 원본의 목록 등이 표시됩니다.

5. 작업 * 메뉴를 선택하고 * 커넥터 편집 * 을 선택합니다.
6. 커넥터 편집 페이지에서 커넥터가 포함된 볼륨에 사용되는 커넥터 이름, 설명, 포함 모델, 데이터 가드레일 사용 및 스냅샷 정책을 변경할 수 있습니다.

   임베딩을 포함한 모든 데이터 소스 스캔에는 비용이 소요됩니다. 커넥터를 만든 후 데이터 가드레일을 활성화하면 데이터 소스가 다시 검사되어 비용이 발생합니다.

7. 변경한 후 * 저장 * 을 선택합니다.

커넥터에 데이터 원본을 추가합니다

커넥터에 추가 데이터 원본을 포함시켜 추가 조직 데이터로 채울 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base & Connectors 인벤토리 페이지에서 데이터 원본을 추가할 커넥터를 선택합니다.
4. 를 선택하고 * 데이터 원본 추가 * 를 선택합니다.
5. 추가하려는 데이터 소스 유형을 선택하세요.
   - ONTAP 파일 시스템용 FSx 추가(기존 ONTAP 볼륨용 FSx의 파일 사용)
   - 파일 시스템 추가(일반 SMB 또는 NFS 공유의 파일 사용)

ONTAP 파일 시스템에 **FSx** 추가

1. * 파일 시스템 선택 *: 데이터 소스 파일이 있는 FSx for ONTAP 파일 시스템을 선택하고 * 다음 * 을 선택합니다.
2. * 볼륨 선택 *: 데이터 원본 파일이 있는 볼륨을 선택하고 * 다음 * 을 선택합니다.

   SMB 프로토콜을 사용하여 저장된 파일을 선택할 때 도메인, IP 주소, 사용자 이름 및 암호를 포함한 Active Directory 정보를 입력해야 합니다.

3. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.
4. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.

   ◦ * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.

   ▪ * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.

   ▪ * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

   높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   ◦ * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.

   ▪ 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

   이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

   이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든 PII가 GenAI 데이터베이스에 마스킹됩니다.

   이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

   ◦ 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

   수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

5. 선택한 데이터 원본이 SMB 프로토콜을 사용하는 볼륨에 있을 때만 사용할 수 있는 * 권한 인식 * 섹션에서 권한 인식 응답을 활성화하거나 비활성화할 수 있습니다.

- 사용: 이 기술 자료에 액세스하는 챗봇 사용자는 액세스 권한이 있는 데이터 원본에서 쿼리에 대한 응답만 받습니다.
- * 사용 안 함 * : 챗봇 사용자는 모든 통합 데이터 소스의 콘텐츠를 사용하여 응답을 받습니다.

6. 이 데이터 소스를 기술 문서에 추가하려면 * 추가 * 를 선택하십시오.

일반 **NFS** 파일 시스템 추가

1. 파일 시스템 선택: 데이터 소스 파일이 있는 파일 시스템 호스트의 IP 주소 또는 FQDN을 입력하고, 네트워크 공유에 대한 NFS 프로토콜을 선택하고 *다음*을 선택합니다.
2. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.

> 경우에 따라 NFS 내보내기 이름을 직접 입력하고 *디렉터리 검색*을 선택하여 사용 가능한 디렉터리를 표시해야 할 수도 있습니다. 내보내기 전체 또는 내보내기에서 특정 폴더만 선택할 수 있습니다.

3. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.
   - * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.
     - * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.
     - * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

     > 높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   - * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.
     - 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

       이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

     이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든 PII가 GenAI 데이터베이스에 마스킹됩니다.

     > 이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

   - 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

4. *데이터 소스 추가*를 선택하여 이 데이터 소스를 지식 기반에 추가하세요.

일반 **SMB** 파일 시스템 추가

1. 파일 시스템 선택:
   a. 데이터 소스 파일이 있는 파일 시스템 호스트의 IP 주소나 FQDN을 입력하세요.
   b. 네트워크 공유에 SMB 프로토콜을 선택합니다.
   c. 도메인, IP 주소, 사용자 이름, 비밀번호 등 Active Directory 정보를 입력합니다.
   d. 다음 * 을 선택합니다.
2. * 데이터 소스 선택 *: 파일을 저장한 위치를 기준으로 데이터 소스 위치를 선택합니다. 전체 볼륨일 수도 있고 볼륨의 특정 폴더 또는 하위 폴더일 수도 있고 * 다음 * 을 선택합니다.

   경우에 따라 SMB 공유 이름을 직접 입력하고 *디렉터리 검색*을 선택하여 사용 가능한 디렉터리를 표시해야 할 수도 있습니다. 전체 공유를 선택하거나 공유에서 특정 폴더만 선택할 수 있습니다.

3. * 구성 *: 데이터 소스가 파일에서 정보를 수집하는 방법과 검색에 포함할 파일을 구성합니다.
   - * 데이터 소스 정의 *: * 청크 전략 * 섹션에서 데이터 소스가 기술 문서에 통합될 때 GenAI 엔진이 데이터 소스 컨텐츠를 청크로 분할하는 방법을 정의합니다. 다음 전략 중 하나를 선택할 수 있습니다.
     - * 다중 문장 청킹 *: 데이터 소스의 정보를 문장 정의 청크로 정리합니다. 각 청크를 구성하는 문장의 수(최대 100개)를 선택할 수 있습니다.
     - * 오버랩 기반 청크 *: 데이터 소스의 정보를 인접 청크와 겹칠 수 있는 문자 정의 청크로 구성합니다. 각 청크의 크기를 문자 단위로 선택하고 각 청크가 인접한 청크와 겹치는 정도를 선택할 수 있습니다. 청크 크기는 50자에서 3000자 사이이고 겹치는 비율은 1 ~ 99%로 구성할 수 있습니다.

       높은 중복 비율을 선택하면 검색 정확도가 약간 개선되어 저장소 요구 사항이 크게 증가할 수 있습니다.

   - 권한 인식: 권한 인식 응답을 활성화하거나 비활성화합니다.
     - 사용: 이 기술 자료에 액세스하는 챗봇 사용자는 액세스 권한이 있는 데이터 원본에서 쿼리에 대한 응답만 받습니다.
     - * 사용 안 함 * : 챗봇 사용자는 모든 통합 데이터 소스의 콘텐츠를 사용하여 응답을 받습니다.
   - * 파일 필터링 *: 검색에 포함할 파일을 구성합니다.
     - 파일 형식 지원 * 섹션에서 모든 파일 형식을 포함하거나 데이터 원본 검색에 포함할 개별 파일 형식을 선택합니다.

       이미지나 PDF 파일을 포함하는 경우 NetApp Workload Factory for GenAI는 이미지(PDF 문서의 이미지 포함)의 텍스트를 구문 분석하므로 비용이 더 많이 발생합니다.

   이미지의 텍스트 데이터를 포함할 경우, 스캔된 텍스트 데이터가 사용자 환경에서 AWS로 전송되기 때문에 GenAI는 이미지에서 PII(개인 식별 정보)를 마스킹할 수 없습니다. 그러나 데이터가 저장되면 모든

PII가 GenAI 데이터베이스에 마스킹됩니다.

이미지 파일을 스캔에 포함할지 여부는 기술 자료 채팅 모델과 관련이 있습니다. 스캔에 이미지 파일을 포함할 경우 채팅 모델은 이미지를 지원해야 합니다. 여기에서 이미지 파일 형식을 선택하면 기술 문서를 이미지 파일을 지원하지 않는 채팅 모델로 전환할 수 없습니다.

- 파일 수정 시간 필터 * 섹션에서 수정 시간에 따라 파일 포함을 활성화 또는 비활성화하도록 선택합니다. 수정 시간 필터링을 사용하는 경우 목록에서 날짜 범위를 선택합니다.

수정 날짜 범위를 기준으로 파일을 포함하는 경우 날짜 범위가 충족되지 않으면(지정한 날짜 범위 내에서 파일이 수정되지 않음) 파일이 정기 검색에서 제외되고 데이터 원본에 이러한 파일이 포함되지 않습니다.

4. *데이터 소스 추가*를 선택하여 이 데이터 소스를 지식 기반에 추가하세요.

결과

데이터 원본이 커넥터에 통합되어 있습니다.

데이터 원본을 커넥터와 동기화합니다

데이터 원본은 하루에 한 번 연결된 커넥터와 자동으로 동기화되므로 데이터 원본 변경 내용이 Amazon Q Business에 반영됩니다. 데이터 원본을 변경하고 데이터를 즉시 동기화(검사)하려는 경우 필요 시 동기화를 수행할 수 있습니다.

동기화는 증분 동기화이므로 Amazon Q Business는 마지막 동기화 이후 추가, 수정 또는 삭제된 데이터 원본의 객체만 처리합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 지식 기반 및 커넥터 메뉴에서 동기화할 커넥터를 선택합니다.
4. 을 선택하고 * 커넥터 관리 * 를 선택합니다.
5. 조치 * 메뉴를 선택하고 * 지금 스캔 * 을 선택합니다.

   데이터 원본을 스캔한다는 메시지와 검사가 완료되면 최종 메시지가 표시됩니다.

결과

커넥터는 첨부된 데이터 원본과 동기화되며 Amazon Q Business는 데이터 원본의 최신 정보를 사용하기 시작합니다.

예약된 동기화를 일시 중지하거나 다시 시작합니다

데이터 원본의 다음 동기화(스캔)를 일시 중지하거나 다시 시작하려면 언제든지 다시 시작할 수 있습니다. 데이터 원본을 변경하고 변경 기간 동안 동기화를 실행하지 않으려면 다음 예약된 동기화를 일시 중지해야 할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.

3. 커넥터 인벤토리 페이지에서 스캔을 일시 중지하거나 다시 시작할 커넥터를 선택합니다.
4. 을 선택하고 * 커넥터 관리 * 를 선택합니다.
5. Actions * 메뉴를 선택하고 * Scan > Pause Scheduled Scan * 또는 * Scan > Resume Scheduled Scan * 을 선택합니다.

   다음 예약된 스캔이 일시 중지되었거나 다시 시작되었다는 메시지가 표시됩니다.

연결선을 삭제합니다

더 이상 커넥터가 필요하지 않으면 삭제할 수 있습니다. 커넥터를 삭제하면 Workload Factory에서 제거되고 커넥터가 포함된 볼륨도 삭제됩니다. 커넥터를 삭제하면 되돌릴 수 없습니다.

커넥터를 삭제할 때 커넥터와 관련된 모든 에이전트에서 커넥터를 연결 해제하여 커넥터와 연결된 모든 리소스를 완전히 삭제해야 합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base & Connectors 인벤토리 페이지에서 삭제할 커넥터를 선택합니다.
4. 을 선택하고 * 커넥터 관리 * 를 선택합니다.
5. 조치 * 메뉴를 선택하고 * 커넥터 삭제 * 를 선택합니다.
6. 커넥터 삭제 대화 상자에서 삭제할 커넥터를 확인하고 * 삭제 * 를 선택합니다.

결과

커넥터가 Workload Factory에서 제거되고 연관된 볼륨이 삭제됩니다.

GenAI 데이터 소스를 관리합니다

FSx for ONTAP 파일 시스템에서 데이터 소스를 사용하여 기술 자료 또는 커넥터를 생성한 후에는 데이터 소스 세부 정보를 보거나, 데이터 소스 콘텐츠를 업데이트 또는 변경하거나, 데이터 소스 설정을 편집하거나, 데이터 소스를 삭제할 수 있습니다.

데이터 원본에 대한 정보를 봅니다

데이터 원본의 내용에 대한 정보를 볼 수 있으며 기술 문서 또는 커넥터를 사용하여 포함 상태를 볼 수 있습니다. 데이터 원본은 기술 문서 또는 커넥터와 연결되어 있으므로 데이터 원본 세부 정보를 보려면 먼저 기술 문서나 커넥터를 선택해야 합니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요. "콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 데이터 원본이 있는 기술 문서 또는 커넥터를 선택한 다음 * 지식 기반 관리 * 또는 * 커넥터 관리 * 를 선택하고 선택합니다.

페이지의 아래쪽에는 연결된 데이터 원본이 나열됩니다.

4. 을 선택하여 각 행을 확장하여 FSx for ONTAP 파일 시스템, 볼륨 및 데이터 소스가 있는 경로와 같은 각 데이터 소스에 대한 자세한 정보를 확인합니다.

   또한 포함 정보 및 해당 데이터 원본이 현재 기술 문서나 커넥터에 포함되어 있는지 여부도 나열합니다.

데이터 원본 설정을 편집합니다

기술 자료 또는 연결선과 통합된 데이터 원본에 대한 정보를 편집할 수 있습니다. 대부분의 정보는 데이터 원본을 추가한 후에 수정되지만 일부 구성(예: 청킹 정의 또는 권한 인식)을 변경할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. Knowledge Base 인벤토리 페이지에서 데이터 원본이 있는 기술 문서를 선택한 다음 * Manage Knowledge Base * 를 선택하고 ••• 선택합니다.

   페이지 아래쪽에는 이 기술 자료에 포함된 데이터 원본이 나열됩니다.

4. 편집할 데이터 원본의 행에서 * 데이터 원본 편집 * 을 선택하고 ••• 선택합니다.
5. 데이터 원본 편집 페이지에서 청크 정의에 대한 행을 확장하도록 선택합니다.
6. 청킹 전략 및 구성, 사용 권한 인식(SMB 볼륨의 경우)에 대한 설정을 업데이트하고 * 저장 * 을 선택합니다.

결과

데이터 소스 설정이 업데이트되고 AI 시스템이 데이터 소스를 동기화하여 기술 자료에 다시 인덱싱되도록 합니다.

기존 데이터 원본의 내용을 업데이트합니다

언제든지 데이터 원본의 내용을 변경하여 조직 데이터를 추가하거나 업데이트할 수 있습니다. 이 데이터 원본이 기술 문서에서 활발하게 사용되는 경우 데이터 원본이 기술 문서에 다시 인덱싱되도록 데이터 원본을 동기화해야 합니다. 동기화는 증분 방식이므로 Amazon Bedrock은 마지막 동기화 이후 추가, 수정 또는 삭제된 FSx for ONTAP 볼륨의 개체만 처리합니다.

데이터 소스는 하루에 한 번 기술 자료와 자동으로 동기화되므로 데이터 소스 변경 사항이 챗봇에 반영됩니다. 데이터 원본을 변경하고 데이터를 즉시 동기화하려는 경우 할 수 "필요 시 동기화를 수행합니다"있습니다.

데이터 원본을 삭제합니다

데이터 원본이 더 이상 기술 문서에 포함되지 않아도 되는 경우 삭제할 수 있습니다.

단계

1. 다음 중 하나를 사용하여 Workload Factory에 로그인하세요."콘솔 환경" .
2. AI 워크로드 타일에서 * 배포 및 관리 * 를 선택합니다.
3. 기술 문서 인벤토리 페이지에서 데이터 원본이 있는 기술 문서를 선택한 다음 * 기술 자료 관리 * 를 선택하고 ••• 선택합니다.

페이지 아래쪽에는 이 기술 자료에 포함된 데이터 원본이 나열됩니다.

4. 삭제할 데이터 원본의 행에서 * 데이터 원본 삭제 * 를 선택하고 ••• 선택합니다.
5. 데이터 원본 삭제 대화 상자에서 삭제할지 확인하고 * 확인 * 을 선택합니다.

결과

데이터 원본이 기술 자료에서 제거되고 AI 시스템은 기술 자료에서 이 데이터 원본에 대한 인덱싱된 정보를 제거합니다. 해당 데이터 소스의 정보는 기술 자료를 사용하는 챗봇에 더 이상 사용할 수 없습니다.

NetApp Workload Factory의 Tracker를 사용하여 워크로드 작업 모니터링

NetApp Workload Factory의 Tracker를 사용하여 워크로드 작업 실행을 모니터링하고 추적하며 작업 진행 상황을 모니터링합니다.

이 작업에 대해

NetApp Workload Factory는 Tracker라는 모니터링 기능을 제공하여 워크로드 작업의 진행 상황과 상태를 모니터링하고 추적하고, 작업 작업과 하위 작업에 대한 세부 정보를 검토하고, 문제나 오류를 진단할 수 있습니다.

Tracker에서 몇 가지 작업을 사용할 수 있습니다. 시간 프레임(최근 24시간, 7일, 14일 또는 30일), 작업 부하, 상태 및 사용자별로 작업을 필터링하고 검색 기능을 사용하여 작업을 찾고 작업 테이블을 CSV 파일로 다운로드할 수 있습니다. 언제든지 Tracker를 새로 고치고 실패한 작업을 재시도하거나 실패한 작업에 대한 매개 변수를 편집하고 작업을 다시 시도할 수 있습니다.

추적기는 작업에 따라 두 가지 수준의 모니터링을 지원합니다. 파일 시스템 배포와 같은 각 작업은 작업 설명, 상태, 시작 시간, 작업 기간, 사용자, 영역, 프록시 리소스, 작업 ID 및 모든 관련 하위 작업을 표시합니다. API 응답을 보고 작업 중에 발생한 상황을 파악할 수 있습니다.

예를 포함한 추적기 작업 수준

- 레벨 1(작업): 파일 시스템 구축을 추적합니다.
- 레벨 2(하위 작업): 파일 시스템 구축과 관련된 하위 작업을 추적합니다.

작업 상태

Tracker의 작업 상태는 다음과 같습니다_in progress_,_success_및_failed_입니다.

작동 주파수

작업 빈도는 작업 유형 및 작업 일정을 기반으로 합니다.

이벤트 보존

이벤트는 사용자 인터페이스에서 30일 동안 유지됩니다.

작업을 추적하고 모니터링합니다

Tracker를 사용하여 Workload Factory 콘솔에서 작업을 추적하고 모니터링하세요.

단계

1. 중 하나를 사용하여 "콘솔 환경"로그인합니다.

2. 작업 부하 메뉴에서 *관리*를 선택한 다음 *추적기*를 선택합니다.
3. Tracker에서 필터나 검색을 사용하여 채용 공고 결과를 좁히세요. 또한, 일자리 보고서를 다운로드할 수도 있습니다.

API 요청을 봅니다

Tracker의 작업에 대한 코드상자에서 API 요청을 봅니다.

단계

1. Tracker에서 작업을 선택합니다.
2. 작업 메뉴를 선택한 다음 *API 요청 보기*를 선택합니다.

실패한 작업을 다시 시도하십시오

Tracker에서 실패한 작업을 재시도하십시오. 실패한 작업의 오류 메시지를 복사할 수도 있습니다.

실패한 작업을 최대 10회까지 재시도할 수 있습니다.

단계

1. Tracker에서 실패한 작업을 선택합니다.
2. 작업 메뉴를 선택한 다음 *다시 시도*를 선택하세요.

결과

작업이 다시 시작됩니다.

실패한 작업을 편집한 후 다시 시도하십시오

실패한 작업의 매개 변수를 편집하고 Tracker 외부에서 작업을 재시도하십시오.

단계

1. Tracker에서 실패한 작업을 선택합니다.
2. 작업 메뉴를 선택한 다음 *편집 및 다시 시도*를 선택하세요.

   매개 변수를 편집하고 작업을 다시 시도할 수 있는 작업 페이지로 리디렉션됩니다.

결과

작업이 다시 시작됩니다. Tracker로 이동하여 작업 상태를 확인합니다.

저작권 정보

Copyright © 2025 NetApp, Inc. All Rights Reserved. 미국에서 인쇄됨 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 라이센스와 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 라이센스도 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 라이센스를 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 라이센스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

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