



의료 AWS 용에서 검색 증강 생성 솔루션 생성

AWS 권장 가이드



AWS 권장 가이드: 의료 AWS 용에서 검색 증강 생성 솔루션 생성

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon의 상표 및 트레이드 드레스는 Amazon 외 제품 또는 서비스와 함께, Amazon 브랜드 이미지를 떨어뜨리거나 고객에게 혼동을 일으킬 수 있는 방식으로 사용할 수 없습니다. Amazon이 소유하지 않은 기타 모든 상표는 Amazon과 제휴 관계이거나 관련이 있거나 후원 관계와 관계없이 해당 소유자의 자산입니다.

Table of Contents

소개	1
환자 치료 및 생산성	1
인재 관리	2
기회 및 과제	3
의료 분야의 생성형 AI 애플리케이션 기회	3
고급 이미지 분석	3
솔루션 산업화와 관련된 문제	3
사용 사례: 의료 인텔리전스 애플리케이션 구축	5
솔루션 개요	5
1단계: 데이터 검색	7
2단계: 의료 지식 그래프 작성	7
3단계: 컨텍스트 검색 에이전트 구축	14
Amazon Bedrock 에이전트	14
LangChain 에이전트	16
4단계: 지식 기반 생성	17
OpenSearch Service 사용	17
RAG 아키텍처 생성	18
5단계: 응답 생성	21
AWS Well-Architected 프레임워크에 맞게 조정	23
사용 사례: 재입원률 예측	24
솔루션 개요	24
1단계: 환자 결과 예측	26
2단계: 환자 행동 예측	28
3단계: 환자 재입원 예측	30
4단계: 성향 점수 계산	32
AWS Well-Architected 프레임워크에 맞게 조정	34
사용 사례: 인재 관리	35
솔루션 개요	35
1단계: 기술 프로필 구축	37
2단계: role-to-skill 관련성 발견	38
3단계: 교육 권장	40
AWS Well-Architected 프레임워크에 맞게 조정	41
솔루션 개발	42
Amazon Q Developer	42

다중 검색기 RAG 설계	42
ReAct 에이전트	44
솔루션 평가	46
정보 추출 평가	46
여러 리트리버 평가	47
LLM 사용	47
리소스	48
AWS 설명서	48
AWS 블로그 게시물	48
기타 리소스	48
기여자	49
작성	49
검토	49
기술 작성	49
문서 기록	50
용어집	51
#	51
A	52
B	54
C	56
D	59
E	63
F	65
G	66
H	67
정보	69
L	71
M	72
O	76
P	78
Q	81
R	81
S	84
T	87
U	89
V	89

W	90
Z	91
.....	xcii

의료용에서 AWS 검색 증강 세대 솔루션 생성

Amazon Web Services, Accenture 및 Cadiem ([기여자](#))

2025년 3월([문서 기록](#))

대규모 언어 모델(LLMs)과 생성형 AI 이전에는 의료 산업에서 자동화된 고정밀 애플리케이션을 개발하는 작업이 어려웠습니다. 기존 방법은 수동 데이터 입력 및 분석에 크게 의존했습니다. 의료 영상 및 환자 기록을 분석하는 복잡성으로 인해 광범위한 사람의 개입이 필요했으며, 이로 인해 종종 조각화되고 비효율적인 워크플로가 발생했습니다. AI 기술의 발전은 대규모로 개인화된 애플리케이션을 구축하는 데 도움이 됩니다. 이제 의료 애플리케이션은 의료 지식 기반과 통합하고, 진단 이미지를 더 정확하게 해석하고, 예측 모델을 사용하여 환자 결과를 예측할 수 있습니다.

이 가이드에서는 LLMs 빌드할 수 있는 Retrieval Augmented Generation 애플리케이션을 통해 어떻게 의료를 혁신하고 있는지 살펴봅니다 AWS 서비스. Retrieval Augmented Generation(RAG)은 응답을 생성하기 전에 LLM이 훈련 데이터 소스 외부에 있는 신뢰할 수 있는 데이터 소스를 참조하는 생성형 AI 기술입니다. RAG 애플리케이션은 모델의 출력을 실제 지식에 기반을 두고 할루시네이션을 줄이고 응답 관련성을 높입니다. 의료 분야에서 RAG를 사용하여 정확하고 up-to-date 의료 정보를 제공할 수 있으므로 의료 서비스 제공자가 최신 연구 및 임상 지침을 이용할 수 있습니다. 데이터를 실행 가능한 인사이트로 변환하고 복잡한 프로세스를 자동화함으로써 이러한 기술은 환자 치료를 개선하고, 운영을 간소화하고, 의료 전문가의 생산성을 개선하는 데 도움이 됩니다.

[Amazon Bedrock](#)에서는 LLMs 미세 조정하고 지능형 에이전트와 통합하여 고급 의료 솔루션을 만들 수 있습니다. [Amazon OpenSearch Service](#)와 [Amazon Neptune](#) 간의 시너지 효과를 강조하는 이 가이드는 이러한 서비스가 향상된 검색 관련성과 고급 다중 소스 데이터 검색을 통해 RAG 솔루션을 개선하는 방법을 보여줍니다. Amazon Bedrock 에이전트를 사용하고 다양한 데이터 리포지토리 간의 상호 작용을 원활하게 조정하는 포괄적인 Amazon Bedrock 솔루션을 오케스트레이션 [LangChain](#) 할 수 있습니다. 이 통합은 특수 서비스를 결합하여 보다 효과적이고 효율적인 AI 기반 시스템을 생성하는 능력을 보여줍니다.

환자 치료 및 생산성

이 가이드에서는 환자 치료 및 생산성에 대한 두 가지 실제 사용 사례, 즉 [환자 데이터 증강](#)과 [재입원 위험 예측](#)을 제공합니다. 이러한 솔루션을 대규모로 구현하기 위한 전략적 청사진을 제공하여 의료 조직에 AI 기반 프로세스를 산업화하기 위한 명확한 경로를 제공합니다. 이러한 인사이트를 통해 의료 기관은 고급 AI 기술을 사용하여 보다 효율적이고 지능적인 워크플로를 생성할 수 있습니다.

인재 관리

또한이 가이드에서는 의료 종사자가 생성형 AI를 일상 업무에 원활하게 통합할 수 있도록 기술을 재교육하고 권한을 부여하는 전략을 간략하게 설명합니다. 이렇게 하면 생산성과 환자 치료 품질이 모두 향상될 수 있습니다. 의료 기관은 인력에게 고급 AI 도구를 효과적으로 사용할 수 있는 기술을 제공하여 투자 수익을 극대화하고 환자 치료 혁신을 주도할 수 있습니다.

이 AI 기반 [인재 관리 솔루션](#)에는 다음과 같은 주요 기능이 포함되어 있습니다.

- 지능형 인재 재개 구문 분석기 -이 도구는 Amazon Bedrock에서 사용할 수 있는 고급 LLMs 사용하여 재개에서 중요한 인재 기술과 속성을 효율적으로 추출하고 분석합니다. 이 도구는 채용 프로세스를 간소화할 수 있습니다.
- 인재 지식 기반 - Amazon Neptune으로 구동되는이 동적 데이터베이스는 인력 배치 수준, 기술 분포 및 업계 추세에 대한 실시간 인사이트를 제공합니다. 이를 통해 인력 관리에 대한 데이터 기반 결정을 내릴 수 있습니다.
- 학습 권장 사항 엔진 -이 AI 기반 도구는 조직 내 기술 격차를 식별하고 의료진을 위한 맞춤형 훈련 프로그램을 권장합니다. 이 도구는 지속적인 전문 개발을 촉진하고 인력이 진화하는 의료 기술에 적응하는 데 도움이 됩니다.

이러한 AI 기반 기능을 함께 사용하면 인력 성과를 최적화하여 인텔리전스와 효율성을 높여 인재 관리를 혁신할 수 있습니다.

기회 및 과제

Amazon Bedrock은 향상된 생산성, 확장성, 비용 효율성 및 데이터 기반 인사이트를 제공할 수 있습니다. Amazon Bedrock은 의료 조직이 콘텐츠 생성 및 데이터 분석부터 자동화된 의사 결정에 이르기까지 다양한 사용 사례에서 LLMs을 효과적으로 사용할 수 있도록 지원합니다. 이 가이드에서는 개념 증명에서 프로덕션으로 전환하는 동안 데이터 품질 문제, 인프라 확장성, 모델 성능 유지 관리, 지속적인 개선 요구 사항과 같은 일반적인 생성형 AI 문제를 극복하기 위한 접근 방식을 제공합니다.

의료 분야의 생성형 AI 애플리케이션 기회

의료 산업은 생성형 AI 애플리케이션에서 제공하는 기회에 따라 변화하는 변화에 대비하고 있습니다. 생성형 AI는 환자 관리를 개선하고, 운영을 간소화하고, 의료 연구를 가속화할 수 있는 잠재력이 있습니다. 의료 공급자는 고급 AI 모델을 사용하여 의료 기록의 증강을 자동화할 수 있습니다. 포괄적인 up-to-date 환자 기록을 통해 보다 정확한 진단 및 치료 계획을 수립할 수 있습니다. 초음파 및 기타 의료 영상 해석과 같은 AI 기반 이미지 분석은 빠르고 정확한 인사이트를 제공하여 의료 전문가의 워크로드를 줄이고 인적 오류 위험을 최소화할 수 있습니다.

생성형 AI는 진단 및 치료 외에도 예측 분석에서 중요한 역할을 할 수 있습니다. 예측 분석은 의료 조직이 환자 결과를 예측하고 그에 따라 의료 계획을 개인화하는 데 도움이 됩니다. 또한 이 기술은 환자 데이터 관리부터 공급자와 환자 간의 통신 간소화에 이르기까지 관리 프로세스를 최적화할 수 있습니다. 생성형 AI 솔루션을 기존 의료 시스템과 통합하면 의료 기관은 효율성을 높이고 비용을 절감하며 궁극적으로 더 높은 품질의 의료 서비스를 제공할 수 있습니다. AI와 의료의 통합은 단지 개선일 뿐만 아니라 보다 지능적이고 응답성이 뛰어나며 환자 중심적인 의료를 향한 근본적인 전환입니다.

고급 이미지 분석

Amazon Bedrock을 Amazon Neptune 및 Amazon OpenSearch Service와 같은 데이터 스토어와 결합하면 의료 분야의 고급 이미지 분석의 복잡성을 해결하는 데 도움이 될 수 있습니다. 정보 검색 솔루션은 진단 이미지를 평가하고 초음파를 해석하여 질병 발견 프로세스를 강화하고 해석 정확도를 높일 수 있습니다. 이 솔루션은 시각적 및 텍스트 평가 데이터를 의사에 의한 수동 환자 평가 검토와 통합할 수 있습니다.

솔루션 산업화와 관련된 문제

의료 분야에서 AI 솔루션을 산업화할 때 해결해야 할 주요 장애물은 데이터 품질과 가용성입니다. 의료 데이터는 종종 조각화되고 일관되지 않은 형식으로 존재합니다. AI 모델이 깔끔하고 구조화된 대표

데이터에 액세스할 수 있도록 하는 것은 실제 시나리오에서 성능을 유지하는 데 매우 중요합니다. 프로덕션 환경 때문에 인프라 확장성이 어려울 수 있습니다. 이러한 환경은 빠른 응답 시간을 제공하고 HIPAA(Health Insurance Portability and Accountability Act)와 같은 데이터 개인 정보 보호 규정을 준수하면서 대량의 실시간 환자 데이터를 처리해야 합니다. 또한 시간이 지남에 따라 새로운 의료 정보와 환자 데이터가 발전함에 따라 AI 모델을 재학습하고 업데이트하여 관련성을 유지하고 정확한 권장 사항을 제공해야 합니다. 마지막으로 상호 운용성 문제와 현재 임상 워크플로에 맞게 조정해야 하기 때문에 이러한 AI 솔루션을 기존 의료 시스템에 통합하는 것이 복잡할 수 있습니다. 이 통합에는 기술 및 운영 변경이 모두 필요합니다.

사용 사례: 증강된 환자 데이터를 사용하여 의료 인텔리전스 애플리케이션 구축

생성형 AI는 임상 및 관리 기능을 모두 개선하여 환자 치료 및 직원 생산성을 높이는 데 도움이 될 수 있습니다. 소노그램 해석과 같은 AI 기반 이미지 분석은 진단 프로세스를 가속화하고 정확도를 개선합니다. 적시에 의료 개입을 지원하는 중요한 인사이트를 제공할 수 있습니다.

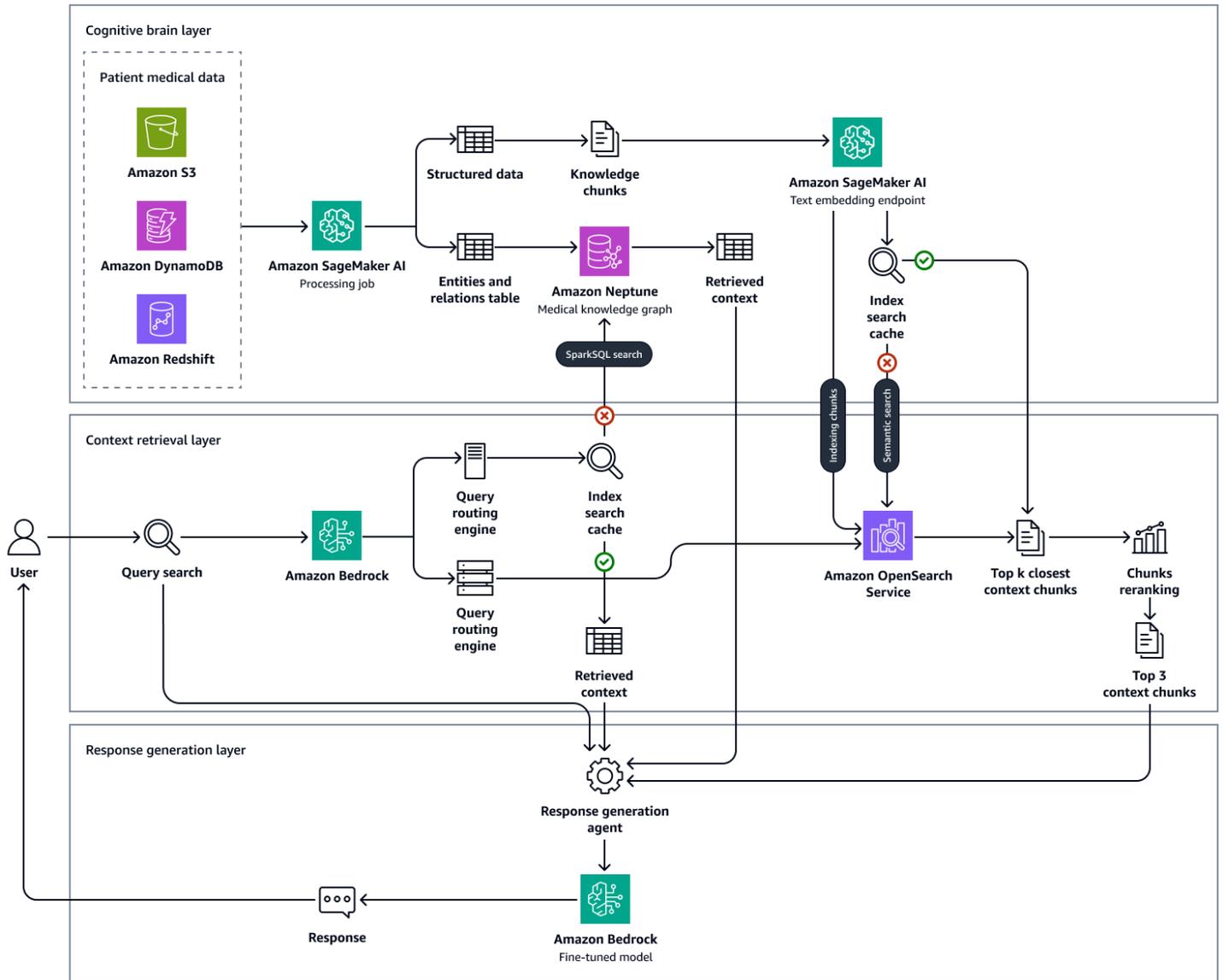
생성형 AI 모델을 지식 그래프와 결합하면 전자 환자 레코드의 시간순 구성을 자동화할 수 있습니다. 이를 통해 의사와 환자의 상호 작용, 증상, 진단, 랩 결과 및 이미지 분석의 실시간 데이터를 통합할 수 있습니다. 이를 통해 의사가 포괄적인 환자 데이터를 확보할 수 있습니다. 이 데이터는 의사가 보다 정확하고 시기적절한 의료 결정을 내리는 데 도움이 되어 환자 결과와 의료 서비스 제공자 생산성을 모두 향상시킵니다.

솔루션 개요

AI는 환자 데이터와 의료 지식을 합성하여 중요한 인사이트를 제공하여 의사와 임상 의에게 권한을 부여할 수 있습니다. 이 Retrieval Augmented Generation(RAG) 솔루션은 수백만 개의 임상 상호 작용에서 얻은 포괄적인 환자 데이터 및 지식을 소비하는 의료 인텔리전스 엔진입니다. 생성형 AI의 기능을 활용하여 환자 치료 개선을 위한 증거 기반 인사이트를 생성합니다. 임상 워크플로를 개선하고, 오류를 줄이고, 환자 결과를 개선하도록 설계되었습니다.

솔루션에는 LLMs. 이 기능은 의료 담당자가 유사한 진단 이미지를 수동으로 검색하고 진단 결과를 분석하는 데 소요되는 시간을 줄입니다.

다음 이미지는 이 솔루션의 end-to-end-workflow를 보여줍니다. Amazon Neptune, Amazon SageMaker AI, Amazon OpenSearch Service 및 Amazon Bedrock의 파운데이션 모델을 사용합니다. Neptune의 의료 지식 그래프와 상호 작용하는 컨텍스트 검색 에이전트의 경우 Amazon Bedrock 에이전트와 LangChain 에이전트 중에서 선택할 수 있습니다.



샘플 의료 질문을 사용한 실험에서 Neptune, OpenSearch 벡터 데이터베이스의 임상 지식 기반 및 Amazon Bedrock LLMs에서 유지 관리하는 지식 그래프를 사용하여 접근 방식으로 생성된 최종 응답은 사실성을 기반으로 하며 거짓 긍정을 줄이고 참 긍정을 부스팅하여 훨씬 더 정확하다는 것을 관찰했습니다. 이 솔루션은 환자의 상태에 대한 증거 기반 인사이트를 생성할 수 있으며 임상 워크플로를 개선하고 오류를 줄이며 환자 결과를 개선하는 것을 목표로 합니다.

이 솔루션 구축은 다음 단계로 구성됩니다.

- [1단계: 데이터 검색](#)
- [2단계: 의료 지식 그래프 작성](#)
- [3단계: 의료 지식 그래프를 쿼리하기 위한 컨텍스트 검색 에이전트 구축](#)

- [4단계: 설명이 포함된 실시간 데이터의 지식 기반 생성](#)
- [5단계: LLMs 사용하여 의료 질문에 답변](#)

1단계: 데이터 검색

의료 AI 기반 솔루션 개발을 지원하는 데 사용할 수 있는 오픈 소스 의료 데이터 세트는 많습니다. 이러한 데이터 세트 중 하나는 [MIMIC-IV 데이터 세트](#)로, 이는 의료 연구 커뮤니티에서 널리 사용되는 공개적으로 사용 가능한 전자 건강 기록(EHR) 데이터 세트입니다. MIMIC-IV에는 환자 레코드의 자유 텍스트 방출 메모를 포함한 자세한 임상 정보가 포함되어 있습니다. 이러한 레코드를 사용하여 텍스트 요약 및 개체 추출 기술을 실험할 수 있습니다. 이러한 기법은 구조화되지 않은 텍스트에서 의료 정보(예: 환자 증상, 약물 및 처방된 치료)를 추출하는 데 도움이 됩니다.

또한 연구 목적으로 특별히 큐레이션된 식별 정보가 제거된 주석이 달린 환자 퇴원 요약을 제공하는 데이터 세트를 사용할 수도 있습니다. 배출 요약 데이터 세트는 개체 추출을 실험하는 데 도움이 되므로 텍스트에서 주요 의료 개체(예: 상태, 절차 및 약물)를 식별할 수 있습니다. 이 가이드 [2단계: 의료 지식 그래프 작성](#)에서는 MIMIC-IV 및 배출 요약 데이터 세트에서 추출한 구조화된 데이터를 사용하여 의료 지식 그래프를 생성하는 방법을 설명합니다. 이 의료 지식 그래프는 의료 전문가를 위한 고급 쿼리 및 의사 결정 지원 시스템의 백본 역할을 합니다.

텍스트 기반 데이터 세트 외에도 이미지 데이터 세트를 사용할 수 있습니다. 예를 들어, 뼈의 다중 보기 영상 이미지의 포괄적인 데이터베이스인 [Musculoskeletal Radiographs\(MURA\) 데이터 세트](#)입니다. 이러한 이미지 데이터 세트를 사용하여 의료 이미지 디코딩 기술을 통해 진단 평가를 실험합니다. 이러한 디코딩 기법은 musculoskeletal disease, 내화성 질병, 내화성 질병 등의 질병 조기 진단에 매우 중요합니다. 의료 이미지 데이터 세트에서 비전 및 언어 파운데이션 모델을 미세 조정하여 진단 이미지의 이상을 감지할 수 있습니다. 이를 통해 시스템은 임상 의에게 빠르고 정확한 진단 인사이트를 제공할 수 있습니다. 이미지 및 텍스트 데이터 세트를 사용하면 텍스트 및 이미지 데이터를 모두 처리하여 환자 관리를 개선할 수 있는 AI 기반 의료 애플리케이션을 생성할 수 있습니다.

2단계: 의료 지식 그래프 작성

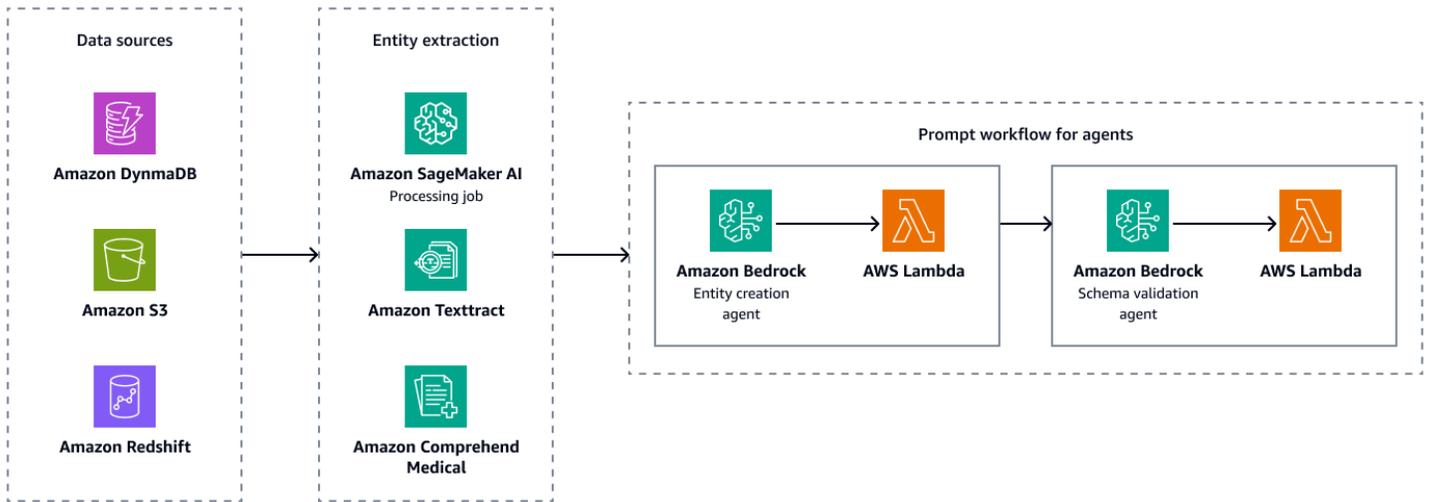
방대한 지식 기반을 기반으로 의사 결정 지원 시스템을 구축하려는 의료 조직의 경우 주요 과제는 임상 기록, 의학 저널, 배출 요약 및 기타 데이터 소스에 있는 의료 엔터티를 찾고 추출하는 것입니다. 또한 추출된 엔터티, 속성 및 관계를 효과적으로 사용하려면 이러한 의료 기록에서 시간 관계, 주제 및 확실성 평가를 캡처해야 합니다.

첫 번째 단계는 Amazon Bedrock의 Llama 3와 같은 파운데이션 모델에 대한 몇 번의 샷 프롬프트를 사용하여 비정형 의료 텍스트에서 의료 개념을 추출하는 것입니다. 퓨샷 프롬프트는 유사한 작업을 수행

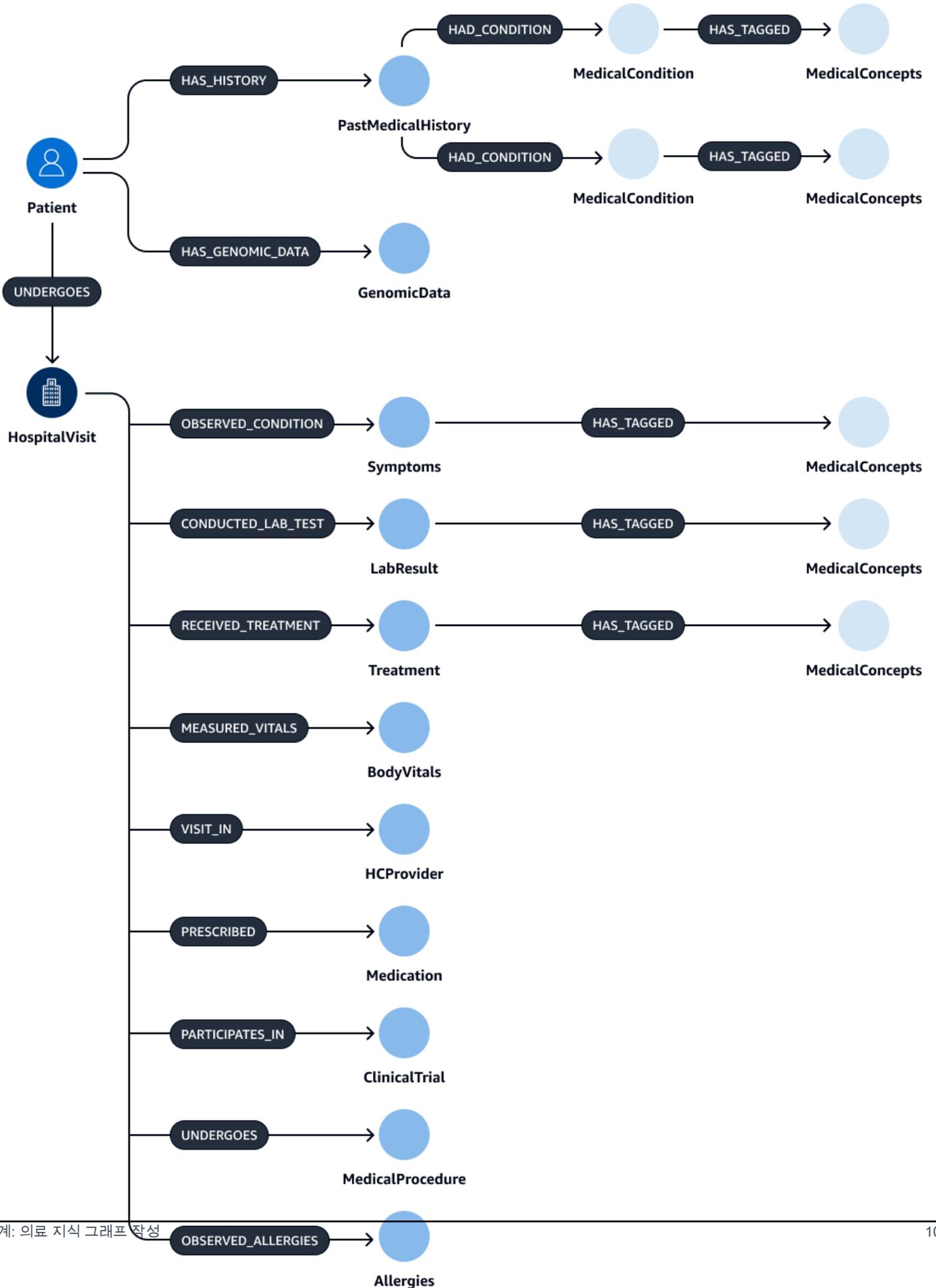
하도록 요청하기 전에 작업과 원하는 출력을 보여주는 몇 가지 예제를 LLM에 제공하는 것입니다. LLM 기반 의료 엔터티 추출기를 사용하여 구조화되지 않은 의료 텍스트를 구문 분석한 다음 의료 지식 엔터티의 구조화된 데이터 표현을 생성할 수 있습니다. 다운스트림 분석 및 자동화를 위해 환자 속성을 저장할 수도 있습니다. 개체 추출 프로세스에는 다음 작업이 포함됩니다.

- 질병, 약물, 의료 기기, 용량, 약물 빈도, 약물 지속 시간, 증상, 의료 절차 및 임상적으로 관련된 속성과 같은 의료 개념에 대한 정보를 추출합니다.
- 추출된 엔터티, 주제 및 확실성 평가 간의 시간적 관계와 같은 기능적 특성을 캡처합니다.
- 다음과 같은 표준 의학 어휘를 확장합니다.
 - RxNorm 데이터베이스의 개념 식별자(RxCUI) [RxNorm](#)
 - [국제 질병 분류, 10차 개정, 임상 수정\(ICD-10-CM\)의 코드](#)
 - [의학 주제 제목\(MeSH\)의 용어](#)
 - [체계화된 의학 명명법의 개념, 임상 용어\(SNOMED CT\)](#)
 - [UMLS\(Unified Medical Language System\)의 코드](#)
- 배출 정보를 요약하고 트랜스크립트에서 의료 인사이트를 도출합니다.

다음 그림은 엔터티, 속성 및 관계의 유효한 쌍 조합을 생성하기 위한 엔터티 추출 및 스키마 검증 단계를 보여줍니다. Amazon Simple Storage Service(Amazon S3)에 배출 요약 또는 환자 메모와 같은 비정형 데이터를 저장할 수 있습니다. 엔터프라이즈 리소스 계획(ERP) 데이터, 전자 환자 레코드, 랩 정보 시스템과 같은 구조화된 데이터를 Amazon Redshift 및 Amazon DynamoDB에 저장할 수 있습니다. Amazon Bedrock 개체 생성 에이전트를 빌드할 수 있습니다. 이 에이전트는 Amazon SageMaker AI 데이터 추출 파이프라인, Amazon Textract 및 Amazon Comprehend Medical과 같은 서비스를 통합하여 정형 및 비정형 데이터 소스에서 엔터티, 관계 및 속성을 추출할 수 있습니다. 마지막으로 Amazon Bedrock 스키마 검증 에이전트를 사용하여 추출된 엔터티 및 관계가 사전 정의된 그래프 스키마를 준수하는지 확인하고 노드 엣지 연결 및 관련 속성의 무결성을 유지합니다.



개체, 관계 및 속성을 추출하고 검증한 후 이를 연결하여 subject-object-predicate 트리플릿을 생성할 수 있습니다. 다음 그림과 같이 데이터를 Amazon Neptune 그래프 데이터베이스에 수집합니다. [그래프 데이터베이스](#)는 데이터 항목 간의 관계를 저장하고 쿼리하도록 최적화되어 있습니다.



이 데이터를 사용하여 포괄적인 지식 그래프를 생성할 수 있습니다. [지식 그래프](#)는 모든 종류의 연결된 정보를 구성하고 쿼리하는 데 도움이 됩니다. 예를 들어, HospitalVisit, PastMedicalHistory, 및 메이저 노드가 있는 지식 그래프를 생성할 수 Symptoms Medication MedicalProcedures 있습니다Treatment.

다음 표에는 배출 노트에서 추출할 수 있는 엔터티와 해당 속성이 나열되어 있습니다.

개체	속성
Patient	PatientID , Name, Age, Gender, Address, ContactInformation
HospitalVisit	VisitDate , Reason, Notes
HealthcareProvider	ProviderID , Name, Specialty , ContactInformation , Address, AffiliatedInstitution
Symptoms	Description , RiskFactors
Allergies	AllergyType , Duration
Medication	MedicationID , Name, Description , Dosage, SideEffects , Manufacturer
PastMedicalHistory	ContinuingMedicines
MedicalCondition	ConditionName , Severity, Treatment Received , DoctorinCharge , HospitalName , MedicinesFollowed
BodyVitals	HeartRate , BloodPressure , RespiratoryRate , BodyTemperature , BMI
LabResult	LabResultID , PatientID , TestName, Result, Date
ClinicalTrial	TrialID, Name, Description , Phase, Status, StartDate , EndDate

개체	속성
GenomicData	GenomicDataID , PatientID , Sequencedata , VariantInformation
Treatment	TreatmentID , Name, Description , Type, SideEffects
MedicalProcedure	ProcedureID , Name, Description , Risks, Outcomes
MedicalConcepts	UMLSCodes , MedicalVocabularies

다음 표에는 엔터티가 가질 수 있는 관계와 해당 속성이 나열되어 있습니다. 예를 들어 Patient 엔터티는 [UNDERGOES] 관계를 사용하여 HospitalVisit 엔터티에 연결할 수 있습니다. 이 관계의 속성은 VisitDate입니다.

주체 엔터티	관계	객체 개체	속성
Patient	[UNDERGOES]	HospitalVisit	VisitDate
HospitalVisit	[VISIT_IN]	HealthcareProvider	ProviderName , Location, ProviderID , VisitDate
HospitalVisit	[OBSERVED_CONDITION]	Symptoms	Severity, CurrentStatus , VisitDate
HospitalVisit	[RECEIVED_TREATMENT]	Treatment	Duration, Dosage, VisitDate
HospitalVisit	[PRESCRIBED]	Medication	Duration, Dosage, Adherence , VisitDate

주체 엔터티	관계	객체 개체	속성
Patient	[HAS_HISTORY]	PastMedicalHistory	없음
PastMedicalHistory	[HAD_CONDITION]	MedicalCondition	DiagnosisDate , CurrentStatus
HospitalVisit	[PARTICIPATES_IN]	ClinicalTrial	VisitDate , Status, Outcomes
Patient	[HAS_GENOMIC_DATA]	GenomicData	CollectionDate
HospitalVisit	[OBSERVED_ALLERGIES]	Allergies	VisitDate
HospitalVisit	[CONDUCTED_LAB_TEST]	LabResult	VisitDate , AnalysisDate , Interpretation
HospitalVisit	[UNDERGOES]	MedicalProcedure	VisitDate , Outcome
MedicalCondition	[HAS_TAGGED]	MedicalConcepts	없음
LabResult	[HAS_TAGGED]	MedicalConcepts	없음
Treatment	[HAS_TAGGED]	MedicalConcepts	없음
Symptoms	[HAS_TAGGED]	MedicalConcepts	없음

3단계: 의료 지식 그래프를 쿼리하기 위한 컨텍스트 검색 에이전트 구축

의료 그래프 데이터베이스를 빌드한 후 다음 단계는 그래프 상호 작용을 위한 에이전트를 빌드하는 것입니다. 에이전트는 의사 또는 임상가가 입력하는 쿼리에 대해 정확하고 필요한 컨텍스트를 검색합니다. 지식 그래프에서 컨텍스트를 검색하는 이러한 에이전트를 구성하는 몇 가지 옵션이 있습니다.

- [Amazon Bedrock 에이전트](#)
- [LangChain 에이전트](#)

그래프 상호 작용을 위한 Amazon Bedrock 에이전트

Amazon Bedrock [에이전트](#)는 Amazon Neptune 그래프 데이터베이스와 원활하게 작동합니다. Amazon Bedrock [작업 그룹](#)을 통해 고급 상호 작용을 수행할 수 있습니다. 작업 그룹은 Neptune openCypher 쿼리를 실행하는 AWS Lambda 함수를 호출하여 프로세스를 시작합니다.

지식 그래프를 쿼리하는 경우 직접 쿼리 실행 또는 컨텍스트 임베딩을 사용한 쿼리라는 두 가지 고유한 접근 방식을 사용할 수 있습니다. 이러한 접근 방식은 특정 사용 사례 및 순위 기준에 따라 독립적으로 적용하거나 결합할 수 있습니다. 두 접근 방식을 결합하면 LLM에 보다 포괄적인 컨텍스트를 제공하여 결과를 개선할 수 있습니다. 다음은 두 가지 쿼리 실행 접근 방식입니다.

- 임베딩 없이 직접 Cypher 쿼리 실행 - Lambda 함수는 임베딩 기반 검색 없이 Neptune에 대해 직접 쿼리를 실행합니다. 다음은 이 접근 방식의 예입니다.

```
MATCH (p:Patient)-[u:UNDERGOES]->(h:HospitalVisit) WHERE h.Reason = 'Acute Diabetes'
AND date(u.VisitDate) > date('2024-01-01')
RETURN p.PatientID, p.Name, p.Age, p.Gender, p.Address, p.ContactInformation
```

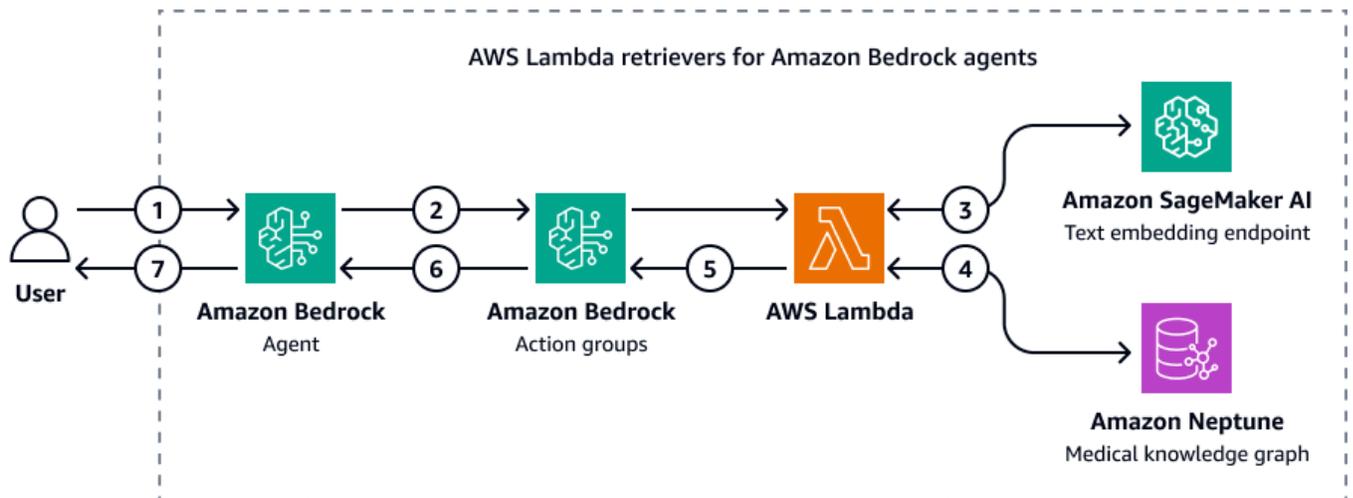
- 임베딩 검색을 사용한 직접 Cypher 쿼리 실행 - Lambda 함수는 임베딩 검색을 사용하여 쿼리 결과를 개선합니다. 이 접근 방식은 데이터의 밀집 벡터 표현인 임베딩을 통합하여 쿼리 실행을 개선합니다. 임베딩은 쿼리에 의미론적 유사성이나 정확한 일치 이상의 광범위한 이해가 필요한 경우에 특히 유용합니다. 사전 훈련된 모델 또는 사용자 지정 훈련된 모델을 사용하여 각 의학적 상태에 대한 임베딩을 생성할 수 있습니다. 다음은 이 접근 방식의 예입니다.

```
CALL { WITH "Acute Diabetes" AS query_term RETURN search_embedding(query_term) AS
similar_reasons }
```

```
MATCH (p:Patient)-[u:UNDERGOES]->(h:HospitalVisit) WHERE h.Reason IN similar reasons
AND date(u.VisitDate) > date('2024-01-01')
RETURN p.PatientID, p.Name, p.Age, p.Gender, p.Address, p.ContactInformation
```

이 예제에서 `search_embedding("Acute Diabetes")` 함수는 의미상 "Acute™"에 가까운 조건을 검색합니다. 이렇게 하면 쿼리에서 당뇨병 전증 또는 예측성 신증과 같은 상태가 있는 환자도 찾을 수 있습니다.

다음 이미지는 의료 지식 그래프의 Cypher 쿼리를 수행하기 위해 Amazon Bedrock 에이전트가 Amazon Neptune와 상호 작용하는 방법을 보여줍니다.



이 다이어그램은 다음 워크플로를 보여줍니다.

1. 사용자가 Amazon Bedrock 에이전트에게 질문을 제출합니다.
2. Amazon Bedrock 에이전트는 질문 및 입력 필터 변수를 Amazon Bedrock 작업 그룹에 전달합니다. 이러한 작업 그룹에는 Amazon SageMaker AI 텍스트 임베딩 엔드포인트 및 Amazon Neptune 의료 지식 그래프와 상호 작용하는 AWS Lambda 함수가 포함되어 있습니다.
3. Lambda 함수는 SageMaker AI 텍스트 임베딩 엔드포인트와 통합되어 openCypher 쿼리 내에서 의미 검색을 수행합니다. 기본 LangChain 에이전트를 사용하여 자연어 쿼리를 openCypher 쿼리로 변환합니다.
4. Lambda 함수는 Neptune 의료 지식 그래프에서 올바른 데이터 세트를 쿼리하고 Neptune 의료 지식 그래프에서 출력을 수신합니다.
5. Lambda 함수는 Neptune의 결과를 Amazon Bedrock 작업 그룹에 반환합니다.
6. Amazon Bedrock 작업 그룹은 검색된 컨텍스트를 Amazon Bedrock 에이전트로 전송합니다.

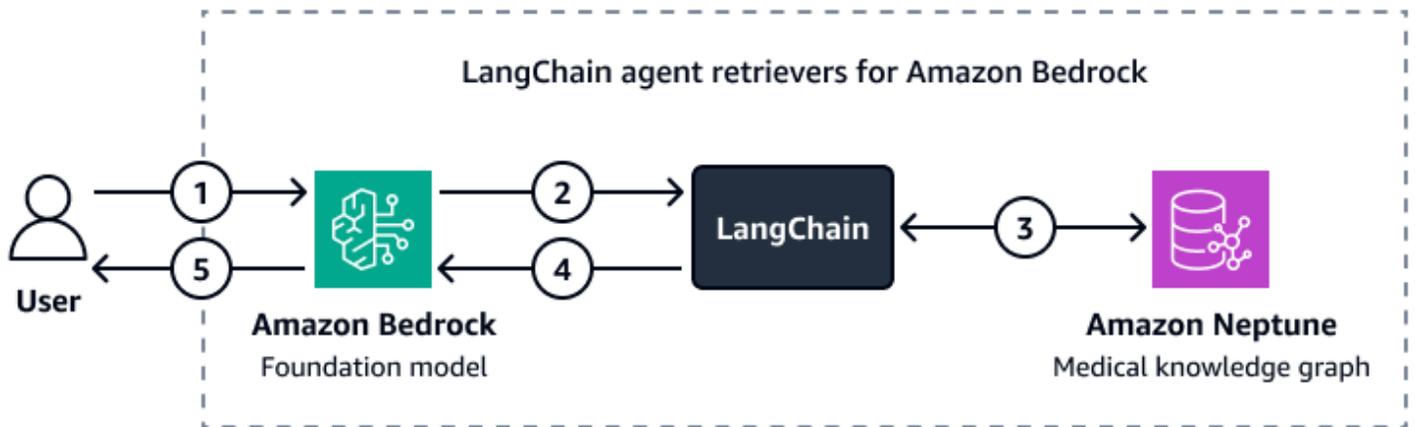
7. Amazon Bedrock 에이전트는 원래 사용자 쿼리와 지식 그래프에서 검색된 컨텍스트를 사용하여 응답을 생성합니다.

LangChain 그래프 상호 작용을 위한 에이전트

를 NeptuneLangChain과 통합하여 그래프 기반 쿼리 및 검색을 활성화할 수 있습니다. 이 접근 방식은 Neptune의 그래프 데이터베이스 기능을 사용하여 AI 기반 워크플로를 개선할 수 있습니다. 사용자 지정 LangChain리트리버는 중개자 역할을 합니다. Amazon Bedrock의 기본 모델은 직접 Cypher 쿼리와 더 복잡한 그래프 알고리즘을 모두 사용하여 Neptune과 상호 작용할 수 있습니다.

사용자 지정 리트리버를 사용하여 LangChain 에이전트가 Neptune 그래프 알고리즘과 상호 작용하는 방식을 세분화할 수 있습니다. 예를 들어, 특정 패턴 또는 예제를 기반으로 파운데이션 모델의 응답을 조정하는 데 도움이 되는 몇 개의 샷 프롬프트를 사용할 수 있습니다. LLM 식별 필터를 적용하여 컨텍스트를 구체화하고 응답의 정밀도를 개선할 수도 있습니다. 이렇게 하면 복잡한 그래프 데이터와 상호 작용할 때 전체 검색 프로세스의 효율성과 정확도가 향상될 수 있습니다.

다음 이미지는 사용자 지정 LangChain 에이전트가 Amazon Bedrock 파운데이션 모델과 Amazon Neptune 의료 지식 그래프 간의 상호 작용을 오케스트레이션하는 방법을 보여줍니다.



이 다이어그램은 다음 워크플로를 보여줍니다.

1. 사용자가 Amazon Bedrock과 LangChain 에이전트에게 질문을 제출합니다.
2. Amazon Bedrock 파운데이션 모델은 LangChain 에이전트가 제공하는 Neptune 스키마를 사용하여 사용자의 질문에 대한 쿼리를 생성합니다.
3. LangChain 에이전트는 Amazon Neptune 의료 지식 그래프에 대해 쿼리를 실행합니다.
4. LangChain 에이전트는 검색된 컨텍스트를 Amazon Bedrock 파운데이션 모델로 보냅니다.

5. Amazon Bedrock 파운데이션 모델은 검색된 컨텍스트를 사용하여 사용자의 질문에 대한 답변을 생성합니다.

4단계: 설명이 포함된 실시간 데이터의 지식 기반 생성

그런 다음 설명이 포함된 실시간 의사-환자 상호 작용 메모, 진단 이미지 평가 및 랩 분석 보고서의 지식 기반을 생성합니다. 이 지식 기반은 [벡터 데이터베이스](#)입니다. 의료 공급자는 인덱싱된 벡터화된 형식으로 설명이 포함된 의료 지식을 저장할 수 있는 벡터 데이터베이스를 사용하여 방대한 리포지토리에서 관련 정보를 효율적으로 쿼리하고 액세스할 수 있습니다. 이러한 벡터화된 표현은 의미상 유사한 데이터를 검색하는 데 도움이 됩니다. 의료 공급자는 임상 기록, 의료 이미지 및 랩 결과를 빠르게 탐색할 수 있습니다. 이를 통해 상황에 맞는 관련 정보에 즉시 액세스할 수 있도록 하여 정보에 입각한 의사 결정을 가속화하고 진단 및 치료 계획의 정확성과 속도를 높일 수 있습니다.

OpenSearch Service 의료 지식 기반 사용

[Amazon OpenSearch Service](#)는 대량의 고차원 의료 데이터를 관리할 수 있습니다. 고성능 검색 및 실시간 분석을 용이하게 하는 관리형 서비스입니다. RAG 애플리케이션을 위한 벡터 데이터베이스로 적합합니다. OpenSearch Service는 의료 기록, 연구 문서, 임상 기록과 같은 방대한 양의 비정형 또는 반정형 데이터를 관리하는 백엔드 도구 역할을 합니다. 고급 의미 체계 검색 기능은 상황에 맞는 정보를 검색하는 데 도움이 됩니다. 따라서 임상 의사 결정 지원 시스템, 환자 쿼리 해결 도구, 의료 지식 관리 시스템과 같은 애플리케이션에서 특히 유용합니다. 예를 들어 임상 의는 특정 증상 또는 치료 프로토콜과 일치하는 관련 환자 데이터 또는 조사 연구를 빠르게 찾을 수 있습니다. 이를 통해 임상 의는 up-to-date의 관련 정보를 바탕으로 결정을 내릴 수 있습니다.

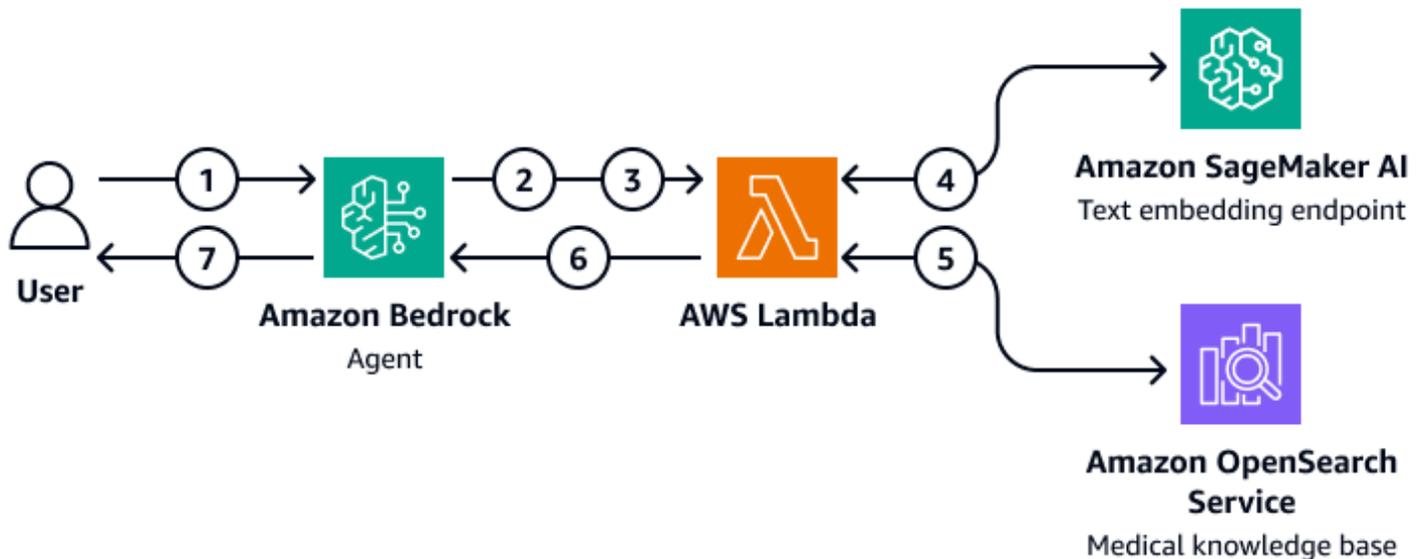
OpenSearch Service는 실시간 데이터 인덱싱 및 쿼리를 확장하고 처리할 수 있습니다. 따라서 정확한 정보에 적시에 액세스하는 것이 중요한 동적 의료 환경에 적합합니다. 또한 의료 이미지 및 의사 메모와 같이 여러 입력이 필요한 검색에 적합한 다중 모달 검색 기능이 있습니다. 의료 애플리케이션용 OpenSearch Service를 구현할 때는 데이터 인덱싱 및 검색을 최적화하기 위해 정확한 필드와 매핑을 정의하는 것이 중요합니다. 필드는 환자 기록, 의료 기록, 진단 코드와 같은 개별 데이터 부분을 나타냅니다. 매핑은 이러한 필드를 저장(임베딩 양식 또는 원본 양식)하고 쿼리하는 방법을 정의합니다. 의료 애플리케이션의 경우 정형 데이터(예: 수치적 테스트 결과), 반정형 데이터(예: 환자 메모), 비정형 데이터(예: 의료 이미지) 등 다양한 데이터 유형을 수용하는 매핑을 설정하는 것이 중요합니다.

OpenSearch Service에서는 큐레이션된 프롬프트를 통해 전체 텍스트 [신경 검색](#) 쿼리를 수행하여 의료 기록, 임상 기록 또는 연구 논문을 검색하여 특정 증상, 치료 또는 환자 기록에 대한 관련 정보를 빠르게 찾을 수 있습니다. 신경 검색 쿼리는 내장 신경망 모델을 사용하여 입력 프롬프트와 이미지의 임베딩을 자동으로 처리합니다. 이를 통해 다중 모달 데이터에서 더 깊은 의미 관계를 이해하고 캡처하여

k-Nearest Neighbor(k-NN) 검색과 같은 다른 검색 쿼리 알고리즘에 비해 컨텍스트 인식 및 정확한 검색 결과를 제공할 수 있습니다.

RAG 아키텍처 생성

Amazon Bedrock 에이전트를 사용하여 OpenSearch Service에서 의료 지식 기반을 쿼리하는 사용자 지정 RAG 솔루션을 배포할 수 있습니다. 이를 위해 OpenSearch Service와 상호 작용하고 쿼리할 수 있는 AWS Lambda 함수를 생성합니다. Lambda 함수는 SageMaker AI 텍스트 임베딩 엔드포인트에 액세스하여 사용자의 입력 질문을 임베딩합니다. Amazon Bedrock 에이전트는 추가 쿼리 파라미터를 Lambda 함수에 대한 입력으로 전달합니다. 함수는 OpenSearch Service에서 의료 지식 기반을 쿼리하여 관련 의료 콘텐츠를 반환합니다. Lambda 함수를 설정한 후 Amazon Bedrock 에이전트 내에서 작업 그룹으로 추가합니다. Amazon Bedrock 에이전트는 사용자의 입력을 받아 필요한 변수를 식별하고 변수와 질문을 Lambda 함수에 전달한 다음 함수를 시작합니다. 함수는 파운데이션 모델이 사용자의 질문에 보다 정확한 답변을 제공하는 데 도움이 되는 컨텍스트를 반환합니다.

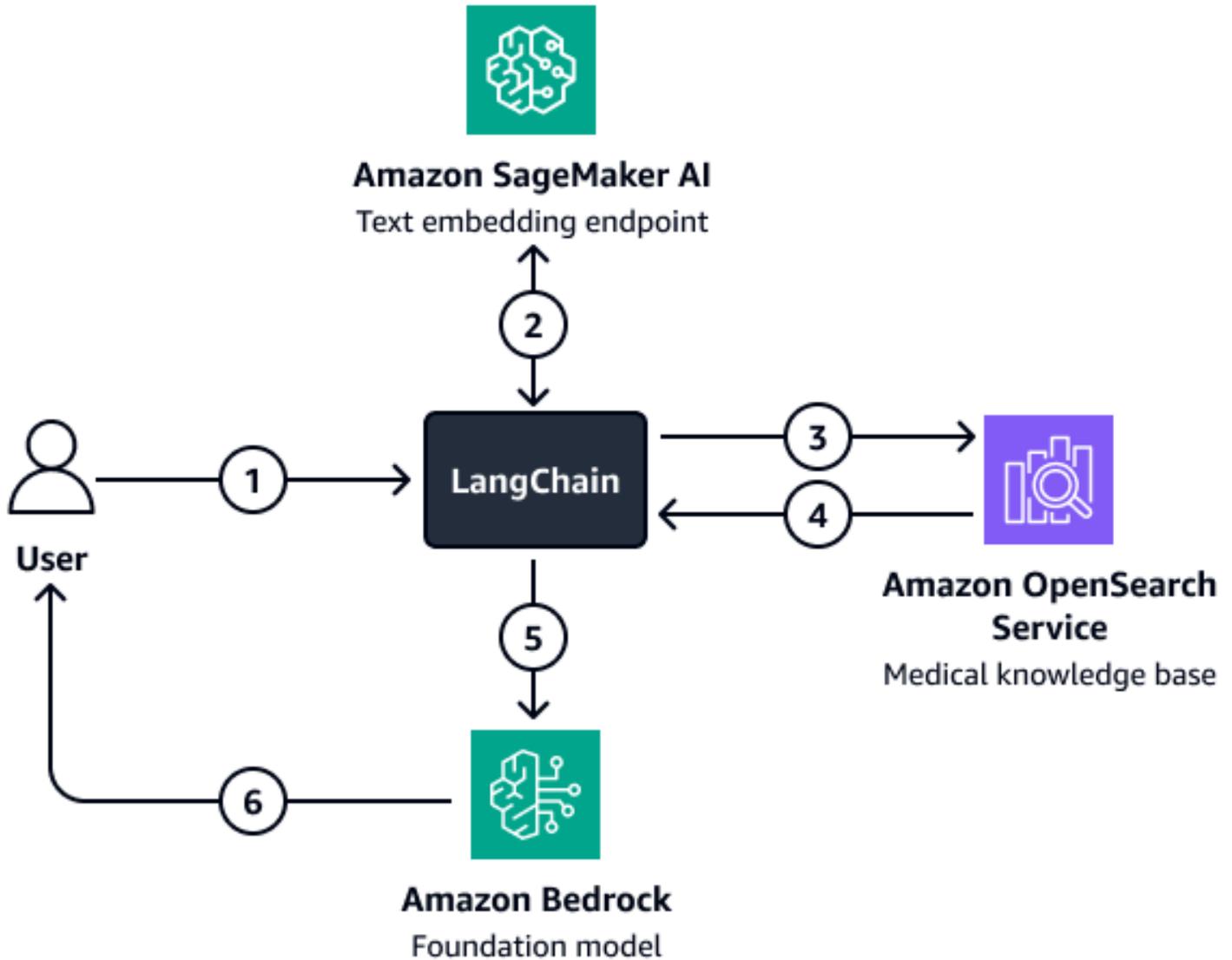


이 다이어그램은 다음 워크플로를 보여줍니다.

1. 사용자가 Amazon Bedrock 에이전트에게 질문을 제출합니다.
2. Amazon Bedrock 에이전트는 시작할 작업 그룹을 선택합니다.
3. Amazon Bedrock 에이전트는 AWS Lambda 함수를 시작하고 해당 함수에 파라미터를 전달합니다.
4. Lambda 함수는 Amazon SageMaker AI 텍스트 임베딩 모델을 시작하여 사용자 질문을 임베딩합니다.

5. Lambda 함수는 포함된 텍스트와 추가 파라미터 및 필터를 Amazon OpenSearch Service에 전달합니다. Amazon OpenSearch Service는 의료 지식 기반을 쿼리하고 Lambda 함수에 결과를 반환합니다.
6. Lambda 함수는 결과를 Amazon Bedrock 에이전트에 다시 전달합니다.
7. Amazon Bedrock 에이전트의 파운데이션 모델은 결과를 기반으로 응답을 생성하고 사용자에게 응답을 반환합니다.

보다 복잡한 필터링이 수반되는 상황에서는 사용자 지정 LangChain리트리버를 사용할 수 있습니다. 이 리트리버는 OpenSearch Service 벡터 검색 클라이언트를 설정하여 이 리트리버를 생성합니다. 이 리트리버를 사용하면 필터 파라미터를 생성하기 위해 더 많은 변수를 전달할 수 있습니다. 리트리버를 설정한 후 Amazon Bedrock 모델 및 리트리버를 사용하여 검색 질문 응답 체인을 설정합니다. 이 체인은 사용자 입력과 잠재적 필터를 리트리버에 전달하여 모델과 리트리버 간의 상호 작용을 오케스트레이션합니다. 리트리버는 파운데이션 모델이 사용자의 질문에 답변하는 데 도움이 되는 관련 컨텍스트를 반환합니다.



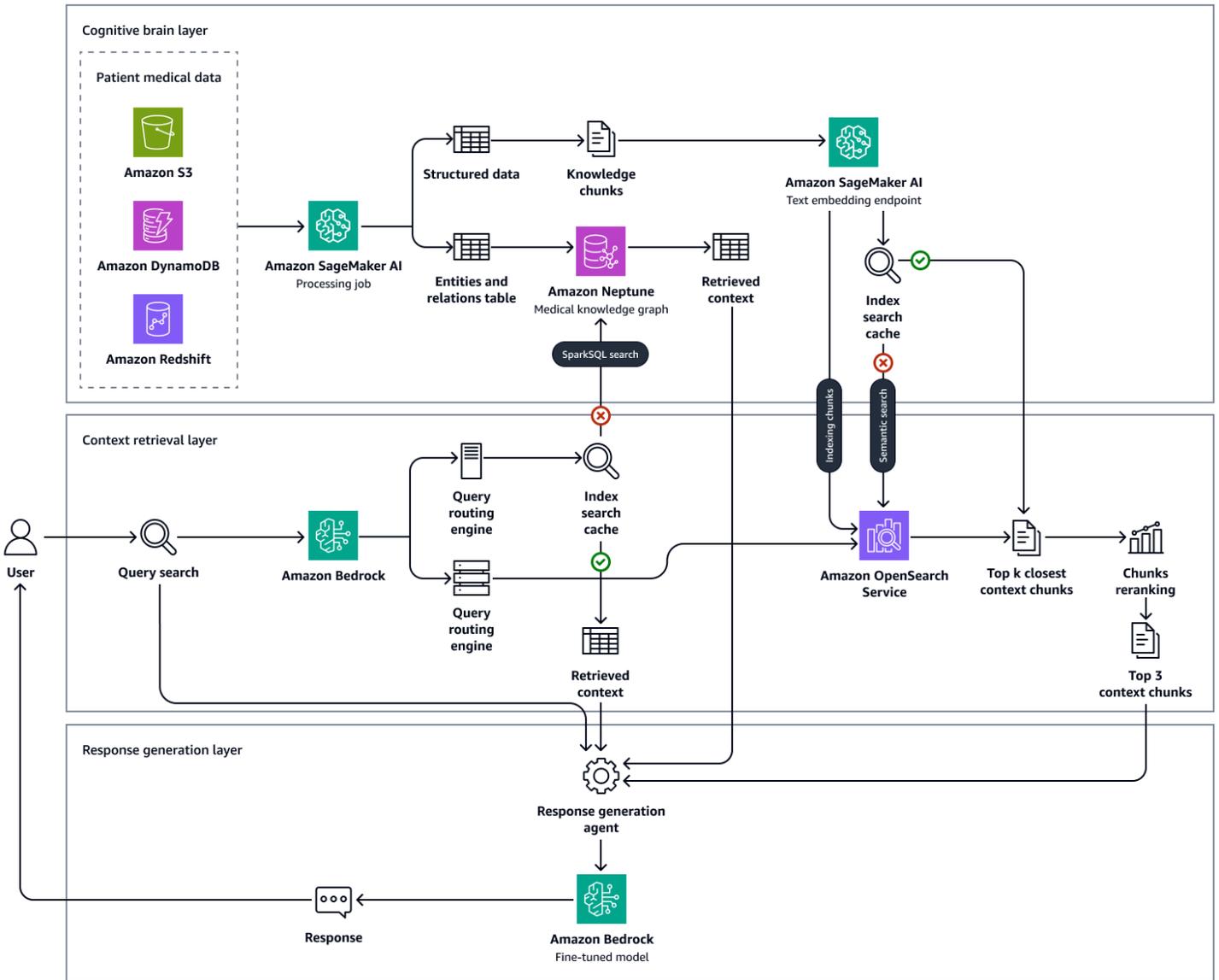
이 다이어그램은 다음 워크플로를 보여줍니다.

1. 사용자가 LangChain 리트리버 에이전트에게 질문을 제출합니다.
2. LangChain 리트리버 에이전트는 Amazon SageMaker AI 텍스트 임베딩 엔드포인트에 질문을 보내 질문을 임베드합니다.
3. LangChain 리트리버 에이전트는 임베디드 텍스트를 Amazon OpenSearch Service에 전달합니다.
4. Amazon OpenSearch Service는 검색된 문서를 LangChain 리트리버 에이전트에 반환합니다.
5. LangChain 리트리버 에이전트는 사용자 질문과 검색된 컨텍스트를 Amazon Bedrock 파운데이션 모델에 전달합니다.
6. 파운데이션 모델은 응답을 생성하여 사용자에게 전송합니다.

5단계: LLMs 사용하여 의료 질문에 답변

이전 단계는 환자의 의료 기록을 가져오고 관련 약물 및 잠재적 진단을 요약할 수 있는 의료 인텔리전스 애플리케이션을 구축하는 데 도움이 됩니다. 이제 생성 계층을 빌드합니다. 이 계층은 Llama 3과 같은 Amazon Bedrock에서 LLM의 생성 기능을 사용하여 애플리케이션의 출력을 보강합니다.

임상 의사 쿼리를 입력하면 애플리케이션의 컨텍스트 검색 계층은 지식 그래프에서 검색 프로세스를 수행하고 환자의 기록, 인구 통계, 증상, 진단 및 결과와 관련된 상위 레코드를 반환합니다. 또한 벡터 데이터베이스에서 설명이 포함된 실시간 의사-환자 상호 작용 메모, 진단 이미지 평가 인사이트, 랩 분석 보고서 요약, 방대한 의학 연구 및 학계의 인사이트를 검색합니다. 검색된 이러한 상위 결과, 임상 의사의 쿼리 및 프롬프트(쿼리의 특성에 따라 답변을 큐레이트하도록 조정됨)는 Amazon Bedrock의 파운데이션 모델로 전달됩니다. 응답 생성 계층입니다. LLM은 검색된 컨텍스트를 사용하여 임상 의사의 쿼리에 대한 응답을 생성합니다. 다음 그림은 이 솔루션 단계의 end-to-end 워크플로를 보여줍니다.



의료 인텔리전스 애플리케이션이 처리해야 하는 다양한 사용 사례에 대해 Llama 3과 같은 사전 훈련된 기본 모델을 Amazon Bedrock에서 사용할 수 있습니다. 지정된 작업에 가장 효과적인 LLM은 사용 사례에 따라 달라집니다. 예를 들어 사전 훈련된 모델은 환자-의사 대화를 요약하고, 약물 및 환자 기록을 검색하고, 내부 의료 데이터 세트와 과학적 지식의 본문에서 인사이트를 검색하는 데 충분할 수 있습니다. 그러나 실시간 실험실 평가, 의료 절차 권장 사항, 환자 결과 예측과 같은 다른 복잡한 사용 사례에는 미세 조정된 LLM이 필요할 수 있습니다. 의료 도메인 데이터 세트에 대해 학습하여 LLM을 미세 조정할 수 있습니다. 구체적이거나 복잡한 의료 및 생명과학 요구 사항은 이러한 미세 조정된 모델의 개발을 주도합니다.

LLM 미세 조정 또는 의료 도메인 데이터에 대해 훈련된 기존 LLM 선택에 대한 자세한 내용은 [의료 및 생명과학 사용 사례에 대규모 언어 모델 사용을 참조하세요](#).

AWS Well-Architected 프레임워크에 맞게 조정

이 솔루션은 다음과 같이 [AWS Well-Architected Framework](#)의 6가지 원칙에 모두 부합합니다.

- 운영 우수성 - 효율적인 모니터링 및 업데이트를 위해 아키텍처가 분리됩니다. Amazon Bedrock 에 이전트 및는 도구를 빠르게 배포하고 롤백하는 데 AWS Lambda 도움이 됩니다.
- 보안 -이 솔루션은 HIPAA와 같은 의료 규정을 준수하도록 설계되었습니다. 또한 암호화, 세분화된 액세스 제어 및 Amazon Bedrock 가드레일을 구현하여 환자 데이터를 보호할 수 있습니다.
- 신뢰성 - Amazon OpenSearch Service 및 Amazon Bedrock과 같은 AWS 관리형 서비스는 지속적인 모델 상호 작용을 위한 인프라를 제공합니다.
- 성능 효율성 - RAG 솔루션은 최적화된 시맨틱 검색 및 Cypher 쿼리를 사용하여 관련 데이터를 빠르게 검색하는 반면 에이전트 라우터는 사용자 쿼리를 위한 최적의 경로를 식별합니다.
- 비용 최적화 pay-per-token 모델은 추론 및 훈련 전 비용을 줄입니다.
- 지속 가능성 - 서버리스 인프라와 pay-per-token 컴퓨팅을 사용하면 리소스 사용량을 최소화하고 지속 가능성을 개선할 수 있습니다.

사용 사례: 환자 결과 및 재입원률 예측

AI 기반 예측 분석은 환자 결과를 예측하고 개인 맞춤형 치료 계획을 활성화하여 추가 이점을 제공합니다. 이를 통해 환자 만족도와 건강 결과를 개선할 수 있습니다. 이러한 AI 기능을 Amazon Bedrock 및 기타 기술과 통합하여 의료 공급자는 상당한 생산성 향상을 달성하고, 비용을 절감하고, 환자 진료의 전반적인 품질을 높일 수 있습니다.

환자 기록, 임상 기록, 약물, 치료 등의 의료 데이터를 [지식 그래프](#)에 저장할 수 있습니다. 의료 지식 그래프의 정형적이고 일시적인 데이터와 LLMs에 대한 심층적인 컨텍스트 이해를 결합하여 의료 공급자는 개별 환자 패턴에 대한 추가 인사이트를 얻을 수 있습니다. 예측 분석을 사용하면 잠재적 미준수 또는 치료 복잡성을 조기에 식별하고 개인화된 재입원 성향 점수를 생성할 수 있습니다.

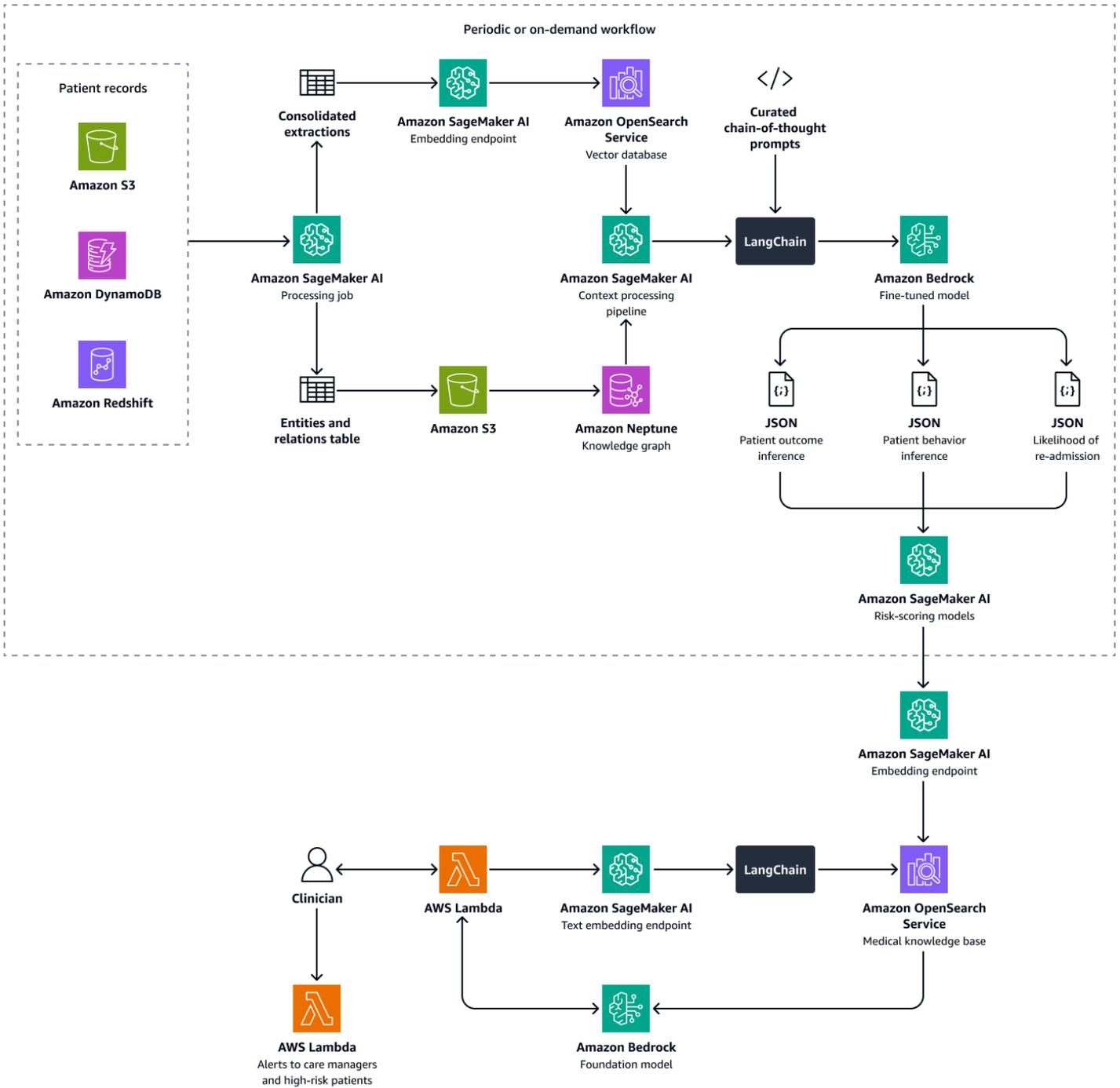
이 솔루션은 재입원 가능성을 예측하는 데 도움이 됩니다. 이러한 예측은 환자 결과를 개선하고 의료 비용을 절감할 수 있습니다. 또한 이 솔루션은 병원 의사와 관리자가 재입원 위험이 높은 환자에게 집중하는 데 도움이 될 수 있습니다. 또한 알림, 셀프 서비스 및 데이터 기반 작업을 통해 이러한 환자와 선제적 개입을 시작하는 데 도움이 됩니다.

솔루션 개요

이 솔루션은 다중 검색 검색 증강 생성(RAG) 프레임워크를 사용하여 환자 데이터를 분석합니다. 개별 환자의 병원 재입원 가능성을 예측하고 병원 수준의 재입원 성향 점수를 계산하는 데 도움이 됩니다. 이 솔루션은 다음 기능을 통합합니다.

- 지식 그래프 - 병원 방문, 이전 재입원, 증상, 실험실 결과, 처방된 치료, 약물 준수 기록과 같은 정형화된 시간순 환자 데이터를 저장합니다.
- 벡터 데이터베이스 - 퇴원 요약, 의사 기록, 예약 누락 또는 보고된 약물 부작용 기록과 같은 비정형 임상 데이터를 저장합니다.
- 미세 조정된 LLM - 환자의 행동, 치료 준수 및 재입원 가능성에 대한 추론을 생성하기 위해 지식 그래프의 정형 데이터와 벡터 데이터베이스의 비정형 데이터를 모두 사용합니다.

위험도 평가 모델은 LLM의 추론을 수치 점수로 정량화합니다. 점수를 병원 수준의 재입원 성향 점수로 집계할 수 있습니다. 이 점수는 각 환자의 위험 노출을 정의하며 주기적으로 또는 필요에 따라 계산할 수 있습니다. 모든 추론 및 위험 점수는 인덱싱되고 Amazon OpenSearch Service에 저장되므로 의료 관리자와 임상가가 이를 검색할 수 있습니다. 대화형 AI 에이전트들이 벡터 데이터베이스와 통합하여 임상가와 의료 관리자는 개별 환자 수준, 시설 전체 수준 또는 의료 전문 분야에서 인사이트를 원활하게 추출할 수 있습니다. 또한 위험 점수를 기반으로 자동 알림을 설정하여 선제적 개입을 장려할 수도 있습니다.



이 솔루션 구축은 다음 단계로 구성됩니다.

- 1단계: 의료 지식 그래프를 사용하여 환자 결과 예측
- 2단계: 처방된 약물 또는 치료에 대한 환자 행동 예측
- 3단계: 환자 재입원 가능성 예측
- 4단계: 병원 재입원 성향 점수 계산

1단계: 의료 지식 그래프를 사용하여 환자 결과 예측

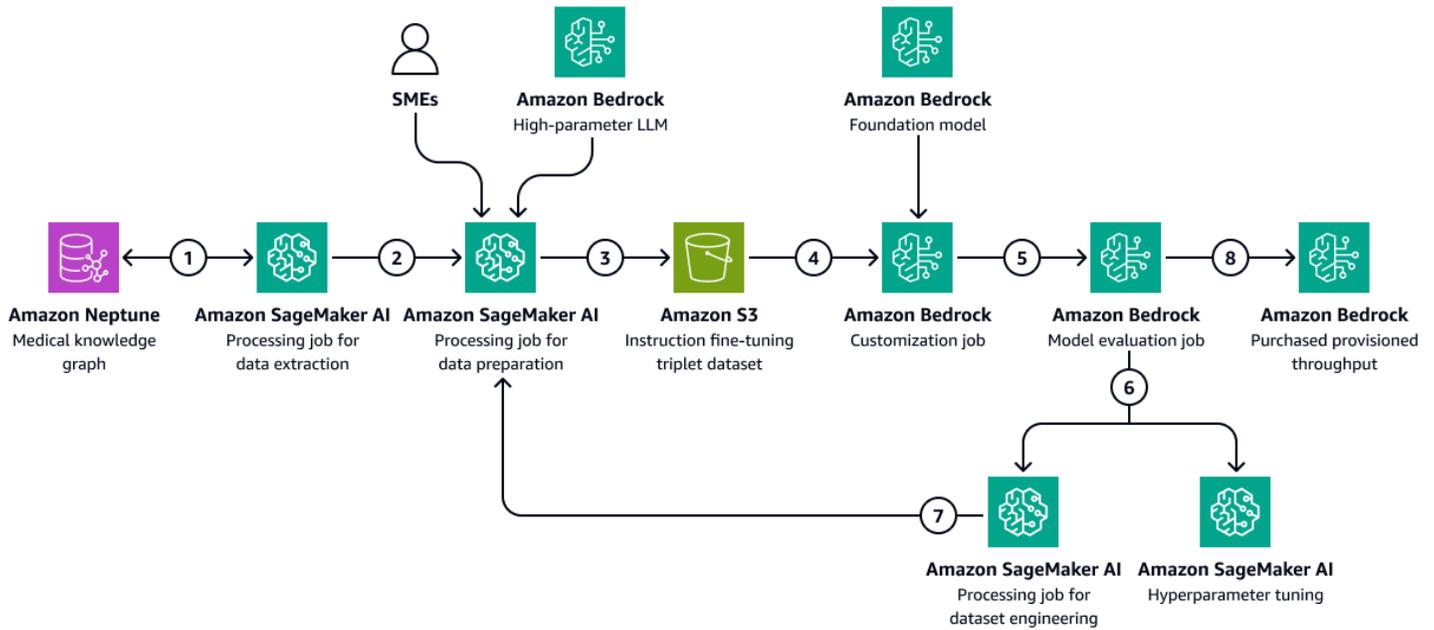
[Amazon Neptune](#)에서는 지식 그래프를 사용하여 시간 경과에 따른 환자 방문 및 결과에 대한 일시적 지식을 저장할 수 있습니다. 지식 그래프를 구축하고 저장하는 가장 효과적인 방법은 그래프 모델과 그래프 데이터베이스를 사용하는 것입니다. 그래프 데이터베이스는 관계를 저장하고 탐색하기 위해 특별히 구축되었습니다. 그래프 데이터베이스를 사용하면 고도로 연결된 데이터를 더 쉽게 모델링 및 관리하고 유연한 스키마를 보유할 수 있습니다.

지식 그래프는 시계열 분석을 수행하는 데 도움이 됩니다. 다음은 환자 결과의 일시적 예측에 사용되는 그래프 데이터베이스의 주요 요소입니다.

- 과거 데이터 - 이전 진단, 지속적인 약물, 이전에 사용한 약물 및 환자의 실험실 결과
- 환자 방문(시간별) - 방문 날짜, 증상, 관찰된 알러지, 임상 기록, 진단, 절차, 치료, 처방 약물 및 실험실 결과
- 증상 및 임상 파라미터 - 심각도, 진행 패턴, 의학에 대한 환자의 반응을 포함한 임상 및 증상 기반 정보

의료 지식 그래프의 인사이트를 사용하여 Llama 3과 같은 Amazon Bedrock의 LLM을 미세 조정할 수 있습니다. 시간 경과에 따른 약물 또는 치료 세트에 대한 환자의 응답에 대한 순차적 환자 데이터로 LLM을 미세 조정합니다. 의 약물 또는 치료 세트와 환자-클리닉 상호 작용 데이터를 환자의 상태를 나타내는 사전 정의된 범주로 분류하는 레이블이 지정된 데이터 세트를 사용합니다. 이러한 범주의 예로는 상태 저하, 개선 또는 안정적인 진행 상황이 있습니다. 임상 의가 환자 및 증상과 관련된 새 컨텍스트를 입력하면 미세 조정된 LLM은 훈련 데이터 세트의 패턴을 사용하여 잠재적 환자 결과를 예측할 수 있습니다.

다음 이미지는 의료별 훈련 데이터 세트를 사용하여 Amazon Bedrock에서 LLM을 미세 조정하는 것과 관련된 순차적 단계를 보여줍니다. 이 데이터에는 환자의 의학적 상태와 시간 경과에 따른 치료에 대한 응답이 포함될 수 있습니다. 이 훈련 데이터 세트는 모델이 환자 결과에 대한 일반화된 예측을 수행하는 데 도움이 됩니다.



이 다이어그램은 다음 워크플로를 보여줍니다.

1. Amazon SageMaker AI 데이터 추출 작업은 지식 그래프를 쿼리하여 시간 경과에 따른 약물 또는 치료 세트에 대한 다양한 환자의 응답에 대한 시간별 데이터를 검색합니다.
2. SageMaker AI 데이터 준비 작업은 Amazon Bedrock LLM과 주제 전문가(SMEs). 작업은 지식 그래프에서 검색된 데이터를 각 환자의 상태를 나타내는 사전 정의된 범주(예: 상태 저하, 개선 또는 안정적인 진행 상황)로 분류합니다.
3. 작업은 지식 그래프에서 추출한 정보, chain-of-thought 프롬프트 및 환자 결과 범주를 포함하는 미세 조정 데이터 세트를 생성합니다. 이 훈련 데이터 세트를 Amazon S3 버킷에 업로드합니다.
4. Amazon Bedrock 사용자 지정 작업은 이 훈련 데이터 세트를 사용하여 LLM을 미세 조정합니다.
5. Amazon Bedrock 사용자 지정 작업은 훈련 환경 내에서 선택한 Amazon Bedrock 기본 모델을 통합합니다. 미세 조정 작업을 시작하고 구성된 훈련 데이터 세트와 훈련 하이퍼파라미터를 사용합니다.
6. Amazon Bedrock 평가 작업은 사전 설계된 모델 평가 프레임워크를 사용하여 미세 조정된 모델을 평가합니다.
7. 모델을 개선해야 하는 경우 훈련 데이터세트를 신중하게 고려한 후 더 많은 데이터로 훈련 작업이 다시 실행됩니다. 모델이 점진적 성능 개선을 보이지 않는 경우 훈련 하이퍼파라미터 수정도 고려하세요.
8. 모델 평가가 비즈니스 이해관계자가 정의한 표준을 충족하면 미세 조정된 모델을 Amazon Bedrock 프로비저닝 처리량으로 호스팅합니다.

2단계: 처방된 약물 또는 치료에 대한 환자 행동 예측

미세 조정된 LLMs 임시 의료 지식 그래프에서 임상 기록, 퇴원 요약 및 기타 환자별 문서를 처리할 수 있습니다. 이들은 환자가 처방된 약물 또는 치료를 따를 가능성이 있는지 평가할 수 있습니다.

이 단계에서는에서 생성된 지식 그래프를 사용합니다 [1단계: 의료 지식 그래프를 사용하여 환자 결과 예측](#). 지식 그래프에는 노드로서 환자의 과거 준수 여부를 포함하여 환자 프로필의 데이터가 포함되어 있습니다. 또한 의약품 또는 치료를 준수하지 않는 사례, 의약품에 대한 부작용, 의약품에 대한 액세스 또는 비용 장벽 부족 또는 이러한 노드의 속성으로 복잡한 투여 방법이 포함됩니다.

미세 조정된 LLMs 의료 지식 그래프의 과거 처방 이행 데이터와 Amazon OpenSearch Service 벡터 데이터베이스의 임상 노트에 대한 설명 요약을 사용할 수 있습니다. 이러한 임상 기록에는 자주 예약 누락 또는 치료 미준수가 언급될 수 있습니다. LLM은 이러한 참고 사항을 사용하여 향후 규정 미준수 가능성을 예측할 수 있습니다.

1. 다음과 같이 입력 데이터를 준비합니다.

- 구조화된 데이터 - 의료 지식 그래프에서 지난 3회 방문 및 랩 결과와 같은 최근 환자 데이터를 추출합니다.
- 비정형 데이터 - Amazon OpenSearch Service 벡터 데이터베이스에서 최근 임상 정보를 검색합니다.

2. 환자 기록과 현재 컨텍스트가 포함된 입력 프롬프트를 구성합니다. 다음은 프롬프트의 예입니다.

```
You are a highly specialized AI model trained in healthcare predictive analytics.
Your task is to analyze a patient's historical medical records, adherence patterns,
and clinical context to predict the likelihood of future non-adherence to
prescribed medications or treatments.
```

```
### **Patient Details**
```

- ****Patient ID:**** {patient_id}
- ****Age:**** {age}
- ****Gender:**** {gender}
- ****Medical Conditions:**** {medical_conditions}
- ****Current Medications:**** {current_medications}
- ****Prescribed Treatments:**** {prescribed_treatments}

```
### **Chronological Medical History**
```

- ****Visit Dates & Symptoms:**** {visit_dates_symptoms}
- ****Diagnoses & Procedures:**** {diagnoses_procedures}
- ****Prescribed Medications & Treatments:**** {medications_treatments}
- ****Past Adherence Patterns:**** {historical_adherence}

```

- Instances of Non-Adherence: {past_non_adherence}
- Side Effects Experienced: {side_effects}
- Barriers to Adherence (e.g., Cost, Access, Dosing Complexity): {barriers}

### Patient-Specific Insights
- Clinical Notes & Discharge Summaries: {clinical_notes}
- Missed Appointments & Non-Compliance Patterns: {missed_appointments}

### Let's think Step-by-Step to predict the patient behaviour
1. You should first analyze past adherence trends and patterns of non-adherence.
2. Identify potential barriers, such as financial constraints, medication side effects, or complex dosing regimens.
3. Thoroughly examine clinical notes and documented patient behaviors that may hint at non-adherence.
4. Correlate adherence history with prescribed treatments and patient conditions.
5. Finally predict the likelihood of non-adherence based on these contextual insights.

### Output Format (JSON)
Return the prediction in the following structured format:
```json
{
 "patient_id": "{patient_id}",
 "likelihood_of_non_adherence": "{low | moderate | high}",
 "reasoning": "{detailed_explanation_based_on_patient_history}"
}

```

3. 프롬프트를 미세 조정된 LLM에 전달합니다. LLM은 프롬프트를 처리하고 결과를 예측합니다. 다음은 LLM의 응답 예제입니다.

```

{
 "patient_id": "P12345",
 "likelihood_of_non_adherence": "high",
 "reasoning": "The patient has a history of missed appointments, has reported side effects to previous medications. Additionally, clinical notes indicate difficulty following complex dosing schedules."
}

```

4. 모델의 응답을 구문 분석하여 예측 결과 범주를 추출합니다. 예를 들어 이전 단계의 예제 응답 범주는 미준수 가능성이 높을 수 있습니다.
5. (선택 사항) 모델 로짓 또는 추가 방법을 사용하여 신뢰도 점수를 할당합니다. 로짓은 특정 클래스 또는 범주에 속하는 항목의 정규화되지 않은 확률입니다.

## 3단계: 환자 재입원 가능성 예측

병원 재입원은 의료 관리 비용이 높고 환자의 안녕에 미치는 영향으로 인해 주요 관심사입니다. 병원 재입원률 계산은 환자 진료의 품질과 의료 제공자의 성과를 측정하는 한 가지 방법입니다.

재입원율을 계산하기 위해 7일 재입원율과 같은 지표를 정의했습니다. 이 지표는 퇴원 후 7일 이내에 계획되지 않은 방문을 위해 입원한 입원 환자의 비율입니다. 환자의 재입원 가능성을 예측하기 위해 미세 조정된 LLM은에서 생성한 의료 지식 그래프의 시간 데이터를 사용할 수 있습니다 [1단계: 의료 지식 그래프를 사용하여 환자 결과 예측](#). 이 지식 그래프는 환자 진료, 절차, 약물 및 증세에 대한 시간별 기록을 유지합니다. 이러한 데이터 레코드에는 다음이 포함됩니다.

- 환자의 마지막 퇴원 이후 경과된 기간
- 과거 치료 및 의약품에 대한 환자의 응답
- 시간 경과에 따른 증상 또는 상태의 진행

이러한 시계열 이벤트를 처리하여 선별된 시스템 프롬프트를 통해 환자의 재입원 가능성을 예측할 수 있습니다. 프롬프트는 미세 조정된 LLM에 예측 로직을 부여합니다.

1. 다음과 같이 입력 데이터를 준비합니다.

- 준수 이력 - 의료 지식 그래프에서 약물 픽업 날짜, 약물 리필 빈도, 진단 및 약물 세부 정보, 시간별 의료 기록 및 기타 정보를 추출합니다.
- 행동 지표 - 약속 누락 및 환자 보고 부작용에 대한 임상 기록을 검색하고 포함합니다.

2. 준수 기록 및 행동 지표가 포함된 입력 프롬프트를 구성합니다. 다음은 프롬프트의 예입니다.

```
You are a highly specialized AI model trained in healthcare predictive analytics.
Your task is to analyze a patient's historical medical records, clinical events, and
adherence patterns to predict the likelihood of hospital readmission within the
next few days.
```

```
Patient Details
```

- **\*\*Patient ID:\*\*** {patient\_id}
- **\*\*Age:\*\*** {age}
- **\*\*Gender:\*\*** {gender}
- **\*\*Primary Diagnoses:\*\*** {diagnoses}
- **\*\*Current Medications:\*\*** {current\_medications}
- **\*\*Prescribed Treatments:\*\*** {prescribed\_treatments}

```
Chronological Medical History
```

- **\*\*Recent Hospital Encounters:\*\*** {encounters}

```

- Time Since Last Discharge: {time_since_last_discharge}
- Previous Readmissions: {past_readmissions}
- Recent Lab Results & Vital Signs: {recent_lab_results}
- Procedures Performed: {procedures_performed}
- Prescribed Medications & Treatments: {medications_treatments}
- Past Adherence Patterns: {historical_adherence}
- Instances of Non-Adherence: {past_non_adherence}

Patient-Specific Insights
- Clinical Notes & Discharge Summaries: {clinical_notes}
- Missed Appointments & Non-Compliance Patterns: {missed_appointments}
- Patient-Reported Side Effects & Complications: {side_effects}

Reasoning Process - You have to analyze this use case step-by-step.
1. First assess time since last discharge and whether recent hospital encounters suggest a pattern of frequent readmissions.
2. Second examine recent lab results, vital signs, and procedures performed to identify clinical deterioration.
3. Third analyze adherence history, checking if past non-adherence to medications or treatments correlates with readmissions.
4. Then identify missed appointments, self-reported side effects, or symptoms worsening from clinical notes.
5. Finally predict the likelihood of readmission based on these contextual insights.

Output Format (JSON)
Return the prediction in the following structured format:
```json
{
  "patient_id": "{patient_id}",
  "likelihood_of_readmission": "{low | moderate | high}",
  "reasoning": "{detailed_explanation_based_on_patient_history}"
}

```

3. 프롬프트를 미세 조정된 LLM에 전달합니다. LLM은 프롬프트를 처리하고 재입원 가능성과 이유를 예측합니다. 다음은 LLM의 응답 예제입니다.

```

{
  "patient_id": "P67890",
  "likelihood_of_readmission": "high",
  "reasoning": "The patient was discharged only 5 days ago, has a history of more than two readmissions to hospitals where the patient received treatment. Recent

```

```
lab results indicate abnormal kidney function and high liver enzymes. These factors
suggest a medium risk of readmission."
}
```

4. 예측을 낮음, 중간 또는 높음과 같은 표준화된 척도로 분류합니다.
5. LLM에서 제공하는 추론을 검토하고 예측에 기여하는 주요 요소를 식별합니다.
6. 정성적 출력을 정량적 점수에 매핑합니다. 예를 들어 매우 높으면 확률이 0.9일 수 있습니다.
7. 검증 데이터 세트를 사용하여 모델 출력을 실제 재입원률과 비교하여 보정합니다.

4단계: 병원 재입원 성향 점수 계산

다음으로 환자당 병원 재입원 성향 점수를 계산합니다. 이 점수는 이전 단계에서 수행된 세 가지 분석, 즉 잠재적 환자 결과, 약물 및 치료에 대한 환자 행동, 환자 재입원 가능성의 순 영향을 반영합니다. 환자 수준 재입원 성향 점수를 특수 수준으로 집계한 다음 병원 수준에서 임상의, 의료 관리자 및 관리자에게 인사이트를 얻을 수 있습니다. 병원 재입원 성향 점수는 시설, 전문 분야 또는 조건별로 전반적인 성과를 평가하는 데 도움이 됩니다. 그런 다음이 점수를 사용하여 선제적 개입을 구현할 수 있습니다.

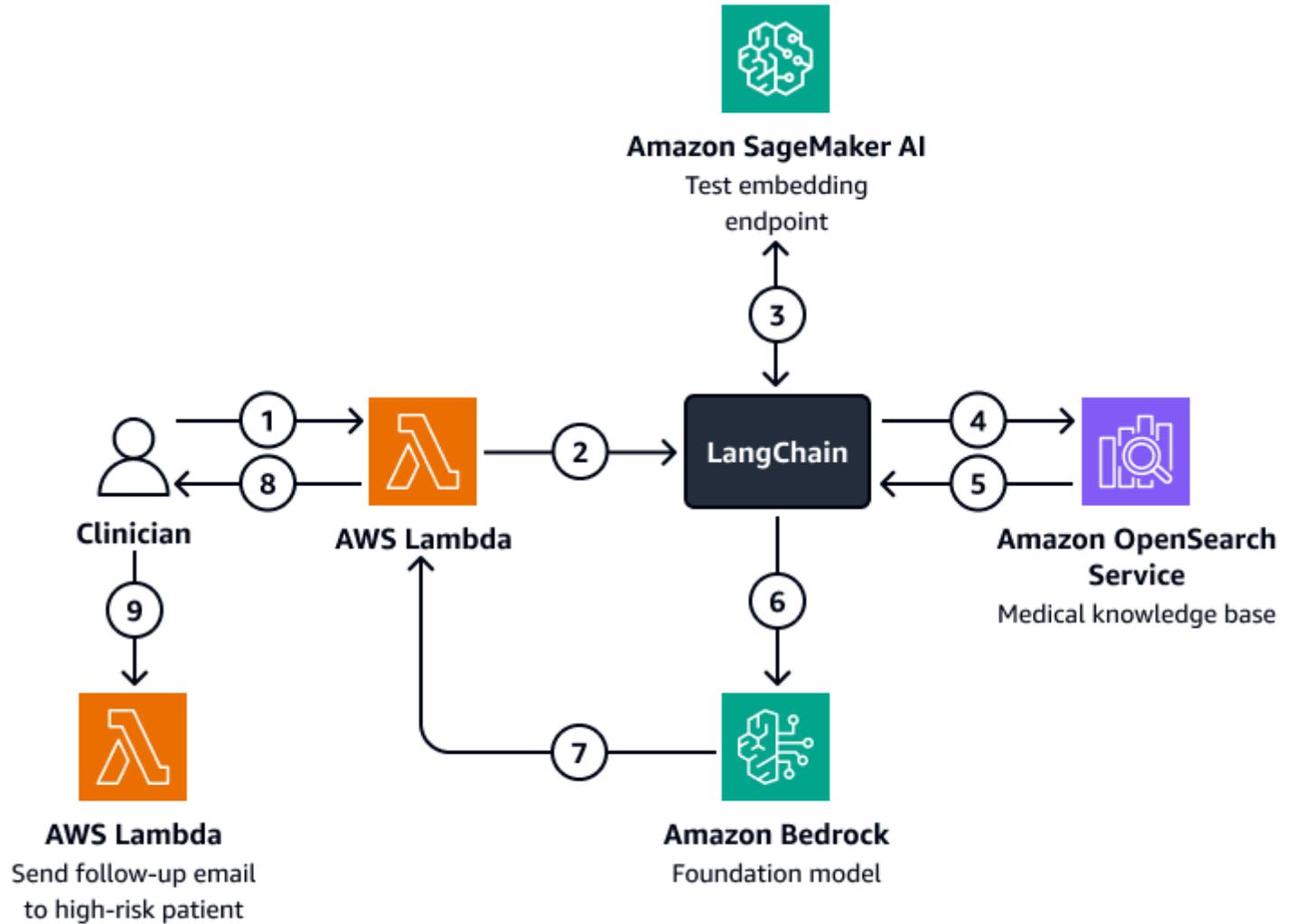
1. 각 요인(결과 예측, 준수 가능성, 재입원)에 가중치를 할당합니다. 다음은 가중치의 예입니다.
 - 결과 예측 가중치: 0.4
 - 준수 예측 가중치: 0.3
 - 재입원 가능성 가중치: 0.3
2. 다음 계산을 사용하여 복합 점수를 계산합니다.

$$\text{ReadmissionPropensityScore} = (\text{OutcomeScore} \times \text{OutcomeWeight}) + (\text{AdherenceScore} \times \text{AdherenceWeight}) + (\text{ReadmissionLikelihoodScore} \times \text{ReadmissionLikelihoodWeight})$$

3. 모든 개별 점수가 0~1과 같은 동일한 척도에 있는지 확인합니다.
4. 작업의 임계값을 정의합니다. 예를 들어 점수가 0.7을 초과하면 알림이 시작됩니다.

위의 분석 및 환자의 재입원 성향 점수를 기반으로 임상의 또는 의료 관리자는 계산된 점수를 기반으로 개별 환자를 모니터링하도록 알림을 설정할 수 있습니다. 사전 정의된 임계값을 초과하면 해당 임계값에 도달하면 알림을 받습니다. 이를 통해 의료 관리자는 환자를 위한 퇴원 의료 계획을 생성할 때 사후 대응보다는 사전 예방적으로 대처할 수 있습니다. 의료 관리자가 대화형 AI 에이전트를 사용하여 원활하게 검색할 수 있도록 환자 결과, 동작 및 재입원 성향 점수를 Amazon OpenSearch Service 벡터 데이터베이스에 인덱싱된 형식으로 저장합니다.

다음 다이어그램은 임상 의 또는 의료 관리자가 환자 결과, 예상 동작 및 재입원 성향에 대한 인사이트를 검색하는 데 사용할 수 있는 대화형 AI 에이전트의 워크플로를 보여줍니다. 사용자는 환자 수준, 부서 수준 또는 병원 수준에서 인사이트를 검색할 수 있습니다. AI 에이전트는 Amazon OpenSearch Service 벡터 데이터베이스에 인덱싱된 형식으로 저장된 이러한 인사이트를 검색합니다. 에이전트는 쿼리를 사용하여 관련 데이터를 검색하고 재입원 위험이 높은 환자를 위한 권장 조치를 포함하여 맞춤형 응답을 제공합니다. 에이전트는 위험 수준에 따라 환자 및 의료 제공자를 위한 미리 알림을 설정할 수도 있습니다.



이 다이어그램은 다음 워크플로를 보여줍니다.

1. 임상 의는 함수가 있는 AWS Lambda 대화형 AI 에이전트에게 질문을 합니다.
2. Lambda 함수는 LangChain 에이전트를 시작합니다.
3. LangChain 에이전트는 사용자의 질문을 Amazon SageMaker AI 텍스트 임베딩 엔드포인트로 보냅니다. 엔드포인트는 질문을 포함합니다.

4. LangChain 에이전트는 Amazon OpenSearch Service의 의료 지식 기반에 임베디드 질문을 전달합니다.
5. Amazon OpenSearch Service는 사용자 쿼리와 가장 관련성이 높은 특정 인사이트를 LangChain 에이전트에게 반환합니다.
6. LangChain 에이전트는 지식 기반에서 Amazon Bedrock 파운데이션 모델로 쿼리와 검색된 컨텍스트를 전송합니다.
7. Amazon Bedrock 파운데이션 모델은 응답을 생성하여 Lambda 함수로 전송합니다.
8. Lambda 함수는 임상의에게 응답을 반환합니다.
9. 임상의는 재입원 위험이 높은 환자에게 후속 이메일을 보내는 Lambda 함수를 시작합니다.

AWS Well-Architected 프레임워크에 맞게 조정

환자 행동을 추적하고 병원 재입원률을 예측하기 위한 아키텍처는 [AWS Well-Architected Framework](#)의 6가지 원칙에 부합하면서 의료 성과를 개선하기 위해 AWS 서비스, 의료 지식 그래프 및 LLMs을 통합합니다.

- 운영 우수성 - 솔루션은 실시간 알림을 AWS Lambda 위해 Amazon Bedrock 및를 사용하는 분리되고 자동화된 시스템입니다.
- 보안 -이 솔루션은 HIPAA와 같은 의료 규정을 준수하도록 설계되었습니다. 또한 암호화, 세분화된 액세스 제어 및 Amazon Bedrock 가드레일을 구현하여 환자 데이터를 보호할 수 있습니다.
- 신뢰성 - 아키텍처는 내결함성 서버리스를 사용합니다 AWS 서비스.
- 성능 효율성 - Amazon OpenSearch Service 및 미세 조정된 LLMs 빠르고 정확한 예측을 제공할 수 있습니다.
- 비용 최적화 - 서버리스 기술 및 pay-per-inference 모델은 비용을 최소화하는 데 도움이 됩니다. 미세 조정된 LLM을 사용하면 추가 요금이 발생할 수 있지만 모델은 미세 조정 프로세스에 필요한 데이터 및 계산 시간을 줄이는 RAG 접근 방식을 사용합니다.
- 지속 가능성 - 아키텍처는 서버리스 인프라를 사용하여 리소스 소비를 최소화합니다. 또한 효율적이고 확장 가능한 의료 운영을 지원합니다.

사용 사례: 의료진 관리 및 기술 향상

인재 혁신 및 기술 향상 전략을 구현하면 인력이 의료 및 의료 서비스에서 새로운 기술과 관행을 사용하는 데 능숙할 수 있습니다. 선제적 기술 향상 이니셔티브를 통해 의료 전문가는 고품질 환자 관리를 제공하고 운영 효율성을 최적화하며 규제 표준을 준수할 수 있습니다. 또한 인재 혁신은 지속적인 학습 문화를 조성합니다. 이는 변화하는 의료 환경에 적응하고 새로운 공중 보건 문제를 해결하는 데 매우 중요합니다. 강의실 기반 훈련 및 정적 학습 모듈과 같은 기존 훈련 접근 방식은 광범위한 대상에게 균일한 콘텐츠를 제공합니다. 개별 실무자의 특정 요구 사항과 숙련도 수준을 해결하는 데 중요한 개인화된 학습 경로가 없는 경우가 많습니다. 이 one-size-fits-all 인해 참여도가 떨어지고 지식 보존이 최적화되지 않을 수 있습니다.

따라서 의료 조직은 현재 상태와 잠재적인 미래 상태에서 각 직원의 격차를 확인할 수 있는 혁신적이고 확장 가능하며 기술 중심의 솔루션을 수용해야 합니다. 이러한 솔루션은 하이퍼 개인화된 학습 경로와 적절한 학습 콘텐츠 세트를 권장해야 합니다. 이를 통해 의료 산업의 미래를 위한 인력을 효과적으로 준비할 수 있습니다.

의료 산업에서는 생성형 AI를 적용하여 작업 인력을 이해하고 기술을 향상할 수 있습니다. 대규모 언어 모델(LLMs)과 고급 리트리버의 연결을 통해 조직은 현재 보유한 기술을 이해하고 향후 필요할 수 있는 주요 기술을 식별할 수 있습니다. 이 정보는 새 작업자를 고용하고 현재 작업 인력의 기술을 향상시켜 격차를 해소하는 데 도움이 됩니다. 의료 기관은 Amazon Bedrock 및 지식 그래프를 사용하여 지속적인 학습과 기술 개발을 촉진하는 도메인별 애플리케이션을 개발할 수 있습니다.

이 솔루션에서 제공하는 지식은 인재를 효과적으로 관리하고, 인력 성과를 최적화하고, 조직의 성공을 촉진하고, 기존 기술을 식별하고, 인재 전략을 수립하는 데 도움이 됩니다. 이 솔루션은 몇 달이 아닌 몇 주 만에 이러한 작업을 수행하는 데 도움이 될 수 있습니다.

솔루션 개요

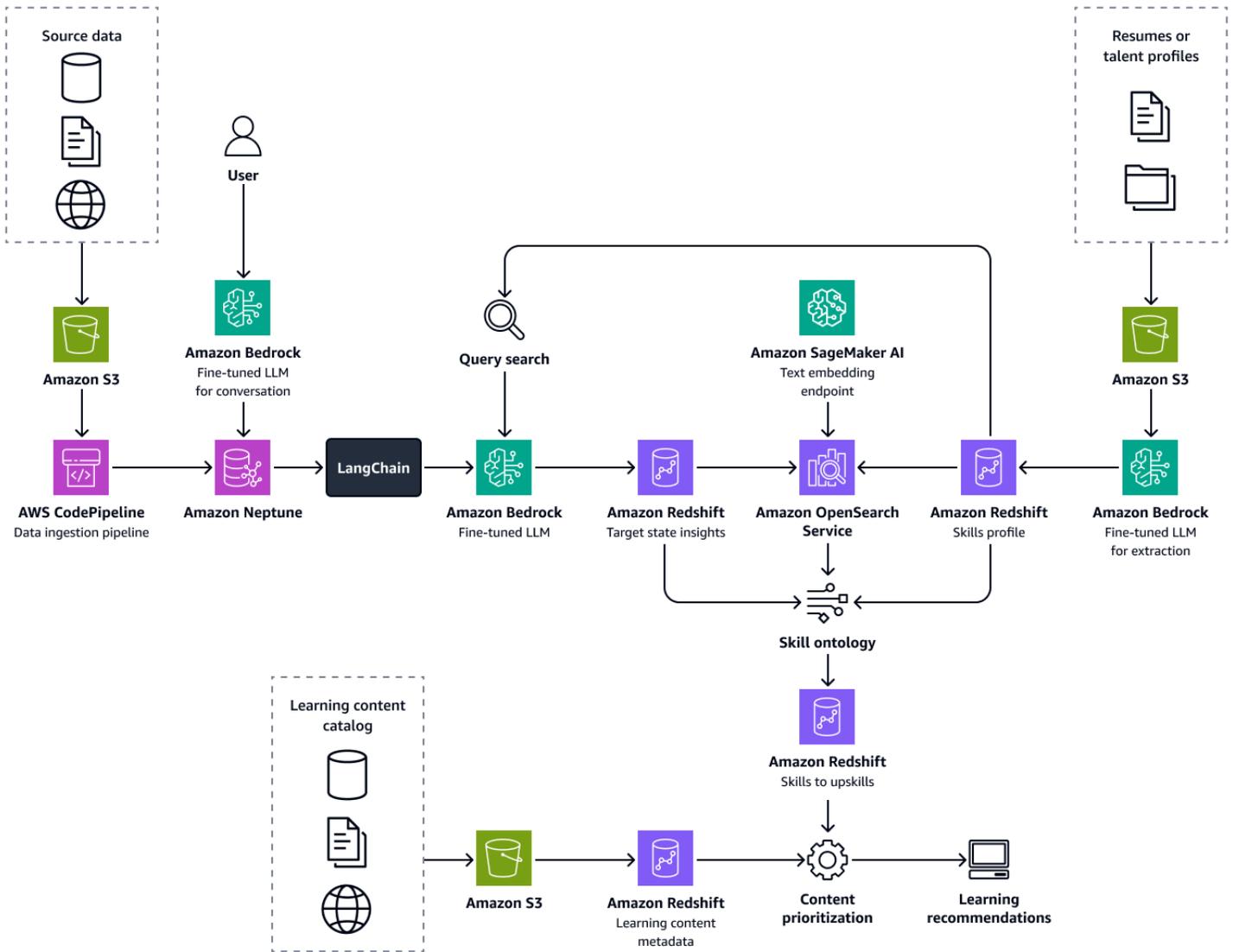
이 솔루션은 다음 구성 요소로 구성된 의료 인재 혁신 프레임워크입니다.

- **지능형 재개 구문 분석기** - 이 구성 요소는 후보의 재개를 읽고 기술을 포함한 후보 정보를 정확하게 추출할 수 있습니다. Amazon Bedrock의 미세 조정된 Llama 2 모델을 사용하여 구축된 지능형 정보 추출 솔루션은 19개 이상의 산업의 재개 및 인재 프로필을 다루는 독점 훈련 데이터세트를 기반으로 합니다. 이 LLM 기반 프로세스는 재개의 수동 검토 프로세스를 자동화하고 상위 후보를 열린 역할에 매칭하여 수백 시간을 절약합니다.
- **지식 그래프** - Amazon Neptune을 기반으로 구축된 지식 그래프로, 조직의 역할 및 기술 분류와 업계를 비롯한 인재 정보의 통합 리포지토리로 기술, 역할 및 속성, 관계 및 논리적 제약 조건의 정의를 사용하여 의료 인재의 의미 체계를 캡처합니다.

- 스킬 온톨로지 - 후보 스킬과 이상적인 현재 상태 또는 미래 상태 스킬(지식 그래프를 사용하여 검색됨) 간의 스킬 근접성 검색은 후보 스킬과 대상 상태 스킬 간의 의미 유사성을 측정하는 온톨로지 알고리즘을 통해 이루어집니다.
- 학습 경로 및 콘텐츠 -이 구성 요소는 식별된 기술 격차를 기반으로 공급업체의 학습 자료 카탈로그에서 올바른 학습 콘텐츠를 추천할 수 있는 학습 권장 엔진입니다. 기술 격차를 분석하고 우선 순위가 지정된 학습 콘텐츠를 권장하여 각 후보에 가장 적합한 기술 향상 경로를 식별하여 새 역할로 전환하는 동안 각 후보에 대해 원활하고 지속적인 전문 개발을 지원합니다.

이 클라우드 기반의 자동화된 솔루션은 기계 학습 서비스, LLMs, 지식 그래프 및 검색 증강 생성(RAG)을 기반으로 합니다. 최소 시간 내에 수십 또는 수천 개의 재개를 처리하고, 즉각적인 후보 프로필을 생성하고, 현재 또는 잠재적 미래 상태의 격차를 식별한 다음 이러한 격차를 해소하기 위해 적절한 학습 콘텐츠를 효율적으로 권장하도록 규모를 조정할 수 있습니다.

다음 이미지는 프레임워크의 end-to-end 흐름을 보여줍니다. 이 솔루션은 Amazon Bedrock에서 미세 조정된 LLMs 기반으로 합니다. 이러한 LLMs Amazon Neptune의 의료 인재 지식 기반에서 데이터를 검색합니다. 데이터 기반 알고리즘은 각 후보에 대한 최적의 학습 경로를 추천합니다.



이 솔루션 구축은 다음 단계로 구성됩니다.

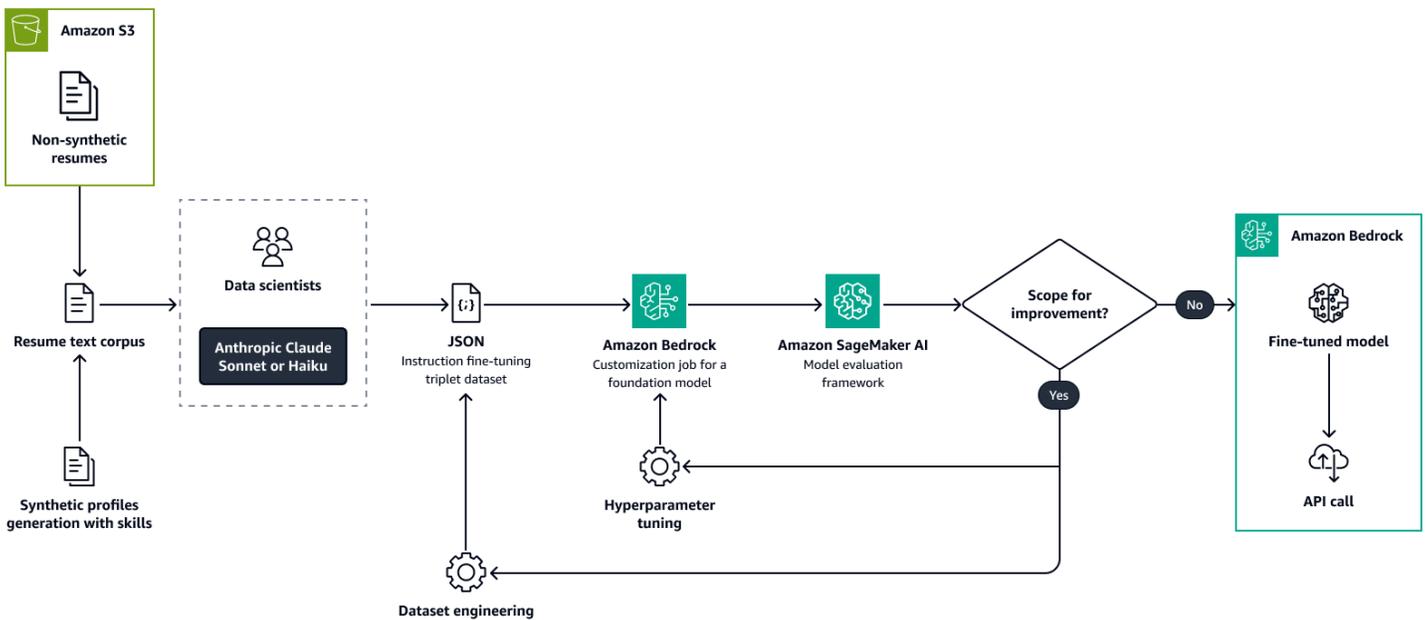
- [1단계: 인재 정보 추출 및 기술 프로파일 구축](#)
- [2단계: 지식 그래프에서 role-to-skill 관련성 발견](#)
- [3단계: 기술 격차 파악 및 교육 권장](#)

1단계: 인재 정보 추출 및 기술 프로파일 구축

먼저 사용자 지정 데이터 세트를 사용하여 Amazon Bedrock에서 Llama 2와 같은 대규모 언어 모델을 미세 조정합니다. 이렇게 하면 사용 사례에 맞게 LLM이 조정됩니다. 교육 중에 지원자 이력서 또는 유사한 인재 프로파일에서 주요 인재 속성을 정확하고 일관되게 추출합니다. 이러한 인재 속성에는 기

술, 현재 역할 제목, 날짜 범위가 있는 경험 제목, 교육 및 인증이 포함됩니다. 자세한 내용은 Amazon Bedrock 설명서의 [사용 사례에 맞게 성능을 개선하도록 모델 사용자 지정](#) 지침을 참조하세요.

다음 이미지는 Amazon Bedrock을 사용하여 재개 구문 분석 모델을 미세 조정하는 프로세스를 보여줍니다. 실제 및 합성으로 생성된 재개는 모두 LLM으로 전달되어 키 정보를 추출합니다. 데이터 과학자 그룹은 추출된 정보를 원본 원시 텍스트와 비교하여 검증합니다. 그런 다음, 추출된 정보는 [chain-of-thought](#) 프롬프트와 원본 텍스트를 사용하여 미세 조정을 위한 훈련 데이터 세트를 도출함으로써 연결됩니다. 그런 다음이 데이터 세트는 모델을 미세 조정하는 Amazon Bedrock 사용자 지정 작업으로 전달됩니다. Amazon SageMaker AI 배치 작업은 미세 조정된 모델을 평가하는 모델 평가 프레임워크를 실행합니다. 모델을 개선해야 하는 경우 더 많은 데이터 또는 다른 하이퍼파라미터로 작업이 다시 실행됩니다. 평가가 표준을 충족하면 Amazon Bedrock 프로비저닝 처리량을 통해 사용자 지정 모델을 호스팅합니다.



2단계: 지식 그래프에서 role-to-skill 관련성 발견

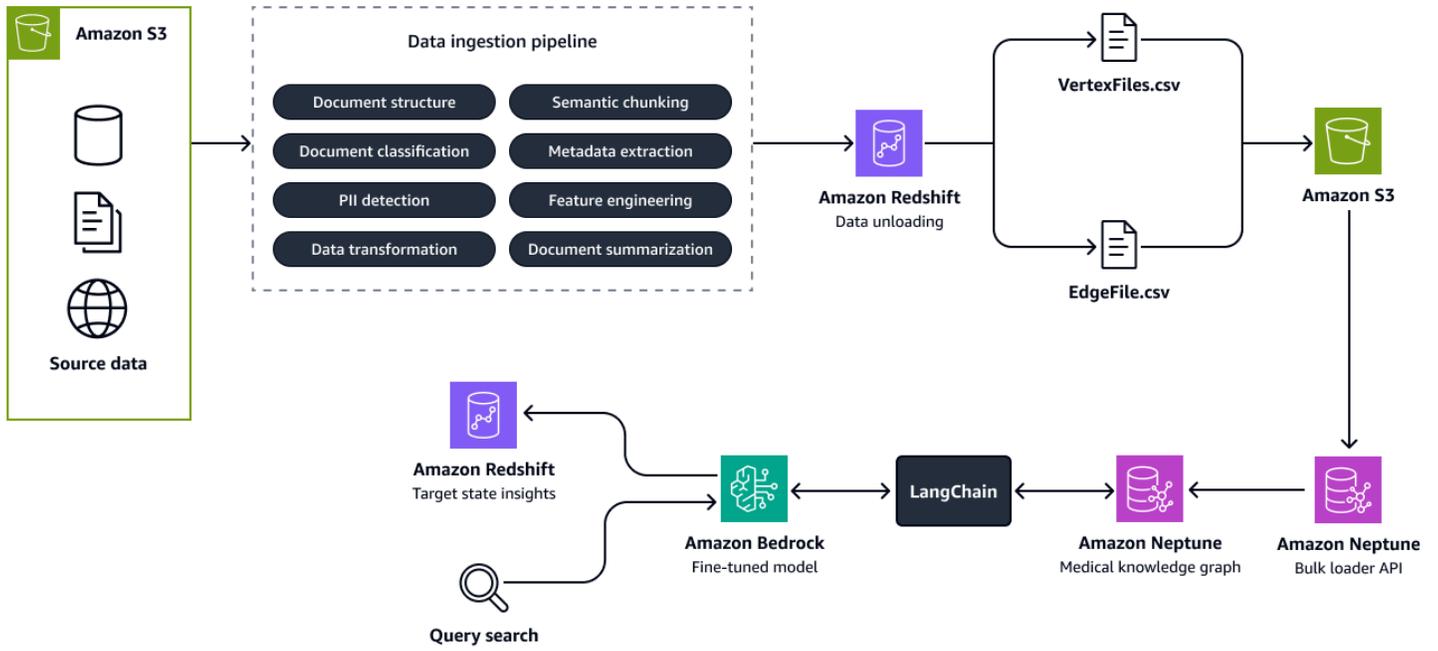
다음으로 조직과 의료 산업의 다른 조직의 기술과 역할 분류를 캡슐화하는 지식 그래프를 작성합니다. 이 향상된 지식 기반은 [Amazon Redshift](#)의 집계된 인재 및 조직 데이터에서 제공됩니다. 다양한 노동 시장 데이터 공급자와 엔터프라이즈 리소스 계획(ERP) 시스템, 인사 정보 시스템(HRIS), 직원 재개, 직무 설명, 인재 아키텍처 문서와 같은 조직별 정형 및 비정형 데이터 소스에서 인재 데이터를 수집할 수 있습니다.

[Amazon Neptune](#)에서 지식 그래프를 빌드합니다. 노드는 기술과 역할을 나타내고 엣지는 이들 간의 관계를 나타냅니다. 이 그래프를 메타데이터로 보강하여 조직 이름, 업계, 직군, 기술 유형, 역할 유형 및 업계 태그와 같은 세부 정보를 포함합니다.

다음으로 그래프 검색 증강 생성(Graph RAG) 애플리케이션을 개발합니다. 그래프 RAG는 그래프 데이터베이스에서 데이터를 검색하는 RAG 접근 방식입니다. 다음은 그래프 RAG 애플리케이션의 구성 요소입니다.

- Amazon Bedrock의 LLM과의 통합 - 애플리케이션은 자연어 이해 및 쿼리 생성을 위해 Amazon Bedrock의 LLM을 사용합니다. 사용자는 자연어를 사용하여 시스템과 상호 작용할 수 있습니다. 이를 통해 비기술 이해관계자가 액세스할 수 있습니다.
- 오케스트레이션 및 정보 검색 - [LlamaIndex](#) 또는 [LangChain](#) 오케스트레이터를 사용하여 LLM과 Neptune 지식 그래프 간의 통합을 용이하게 합니다. 자연어 쿼리를 [openCypher](#) 쿼리로 변환하는 프로세스를 관리합니다. 그런 다음 지식 그래프에서 쿼리를 실행합니다. 프롬프트 엔지니어링을 사용하여 openCypher 쿼리를 구성하는 모범 사례에 대해 LLM에 지시합니다. 이렇게 하면 쿼리를 최적화하여 쿼리된 역할 및 기술에 대한 모든 관련 엔터티 및 관계가 포함된 관련 하위 그래프를 검색할 수 있습니다.
- 인사이트 생성 - Amazon Bedrock의 LLM은 검색된 그래프 데이터를 처리합니다. 쿼리된 역할 및 관련 기술에 대한 현재 상태 및 프로젝트 미래 상태에 대한 자세한 인사이트를 생성합니다.

다음 이미지는 소스 데이터에서 지식 그래프를 작성하는 단계를 보여줍니다. 정형 및 비정형 소스 데이터를 데이터 수집 파이프라인에 전달합니다. 파이프라인은 Amazon Neptune과 호환되는 CSV 대량 로드 형식으로 정보를 추출하고 변환합니다. 대량 로더 API는 Amazon S3 버킷에 저장된 CSV 파일을 Neptune 지식 그래프에 업로드합니다. 현재 미래 상태, 관련 역할 또는 기술과 관련된 사용자 쿼리의 경우 Amazon Bedrock의 미세 조정된 LLM은 LangChain 오케스트레이터를 통해 지식 그래프와 상호 작용합니다. 오케스트레이터는 지식 그래프에서 관련 컨텍스트를 검색하고 응답을 Amazon Redshift의 인사이트 테이블에 푸시합니다. [GraphQAChain](#)과 같은 LangChain 오케스트레이터는 자연어 쿼리를 사용자의 openCypher 쿼리로 변환하여 지식 그래프를 쿼리합니다. Amazon Bedrock 미세 조정 모델은 검색된 컨텍스트를 기반으로 응답을 생성합니다.



3단계: 기술 격차 파악 및 교육 권장

이 단계에서는 의료 전문가의 현재 상태와 잠재적인 미래 상태 역할 간의 근접성을 정확하게 계산합니다. 이렇게 하려면 개인의 기술 세트를 직무 역할과 비교하여 기술 선호도 분석을 수행합니다. [Amazon OpenSearch Service](#) 벡터 데이터베이스에서는 스킬 설명, 스킬 유형 및 스킬 클러스터와 같은 스킬 분류 정보 및 스킬 메타데이터를 저장합니다. Amazon [Titan Text Embeddings 모델과 같은 Amazon Bedrock 임베딩 모델을](#) 사용하여 식별된 키 스킬을 벡터에 임베딩합니다. 벡터 검색을 통해 현재 상태 스킬 및 대상 상태 스킬에 대한 설명을 검색하고 온톨로지 분석을 수행합니다. 분석은 현재 및 대상 상태 스킬 페어 간의 근접 점수를 제공합니다. 각 페어에 대해 계산된 온톨로지 점수를 사용하여 스킬 친화도의 격차를 식별합니다. 그런 다음 역할 전환 중에 후보가 고려할 수 있는 최적의 기술 향상 경로를 권장합니다.

각 역할에 대해 기술 향상 또는 리스킬링에 올바른 학습 콘텐츠를 추천하려면 포괄적인 학습 콘텐츠 카탈로그를 만드는 것으로 시작하는 체계적인 접근 방식이 필요합니다. Amazon Redshift 데이터베이스에 저장하는 이 카탈로그는 다양한 공급자의 콘텐츠를 집계하고 콘텐츠 기간, 난이도 수준, 학습 모드와 같은 메타데이터를 포함합니다. 다음 단계는 각 콘텐츠에서 제공하는 주요 기술을 추출한 다음 대상 역할에 필요한 개별 기술에 매핑하는 것입니다. 기술 근접 분석을 통해 콘텐츠에서 제공하는 적용 범위를 분석하여 이 매핑을 달성할 수 있습니다. 이 분석은 콘텐츠가 학습한 기술이 역할에 필요한 기술과 얼마나 일치하는지 평가합니다. 메타데이터는 각 스킬에 가장 적합한 콘텐츠를 선택하는 데 중요한 역할을 하여 학습자에게 학습 요구 사항에 맞는 맞춤형 추천을 제공합니다. Amazon Bedrock의 LLMs 사용하

여 콘텐츠 메타데이터에서 기술을 추출하고, 기능 엔지니어링을 수행하고, 콘텐츠 권장 사항을 검증합니다. 이렇게 하면 업스킬링 또는 리스킬링 프로세스의 정확성과 관련성이 향상됩니다.

AWS Well-Architected 프레임워크에 맞게 조정

이 솔루션은 [AWS Well-Architected Framework](#)의 6가지 원칙에 모두 부합합니다.

- **운영 우수성** - 모듈식 자동 파이프라인은 운영 우수성을 향상시킵니다. 파이프라인의 주요 구성 요소는 분리되고 자동화되므로 모델 업데이트가 더 빨라지고 모니터링이 더 쉬워집니다. 또한 자동 훈련 파이프라인은 미세 조정된 모델의 더 빠른 릴리스를 지원합니다.
- **보안** - 이 솔루션은 재개 및 인재 프로필의 데이터와 같은 민감한 개인 식별 정보(PII)를 처리합니다. [AWS Identity and Access Management \(IAM\)](#)에서 세분화된 액세스 제어 정책을 구현하고 권한이 있는 직원만이 데이터에 액세스할 수 있도록 합니다.
- **신뢰성** - 이 솔루션은 Neptune AWS 서비스, Amazon Bedrock 및 OpenSearch Service와 같은를 사용하여 수요가 많은 경우에도 내결함성, 고가용성 및 인사이트에 대한 중단 없는 액세스를 제공합니다.
- **성능 효율성** - Amazon Bedrock 및 OpenSearch Service 벡터 데이터베이스의 미세 조정된 LLMs은 대규모 데이터 세트를 빠르고 정확하게 처리하여 적시에 개인화된 학습 권장 사항을 제공하도록 설계되었습니다.
- **비용 최적화** - 이 솔루션은 RAG 접근 방식을 사용하므로 모델의 지속적인 사전 훈련이 필요하지 않습니다. 전체 모델을 반복적으로 미세 조정하는 대신 시스템은 재개에서 정보 추출 및 출력 구조화와 같은 특정 프로세스만 미세 조정합니다. 이로 인해 비용을 크게 절감할 수 있습니다. 리소스 집약적인 모델 훈련의 빈도와 규모를 최소화하고 pay-per-use 클라우드 서비스를 사용하면 의료 기관은 높은 성능을 유지하면서 운영 비용을 최적화할 수 있습니다.
- **지속 가능성** - 이 솔루션은 컴퓨팅 리소스를 동적으로 할당하는 확장 가능한 클라우드 네이티브 서비스를 사용합니다. 이를 통해 대규모의 데이터 집약적인 인재 혁신 이니셔티브를 지원하면서 에너지 소비와 환경 영향을 줄일 수 있습니다.

의료용 생성형 AI 솔루션 개발 및 오케스트레이션

이 가이드의 솔루션을 구축하려면 미세 조정된 LLMs을 사용하여 증강된 환자 데이터, 임상 및 진단 인사이트, 예측된 환자 결과를 의료 공급자에게 제공하는 RAG 아키텍처를 구축해야 합니다. 이를 위해서는 응집력 있고 효율적인 워크플로를 생성하기 위해 여러 AWS 서비스 및 도구를 통합해야 합니다. 이 섹션에서는 다음에 대해 설명합니다.

- [Amazon Q Developer](#) - Amazon Q Developer를 사용하여 개발 프로세스 중에 엔지니어링 질문 및 코드 오류를 해결합니다.
- [다중 검색기 RAG 설계](#) - 여러 리트리버를 사용하여 사용자의 질문에 맞는 의료 컨텍스트를 가져오는 RAG 솔루션을 설계하고 구현합니다.
- [ReAct 에이전트](#) - 추론과 동적 작업을 결합하는 에이전트를 구현합니다.

Amazon Q Developer

생성형 AI 솔루션을 구축할 때 AI 에이전트와 연결 키 서비스를 생성하기 어려울 수 있습니다. 그러나 [Amazon Q Developer](#)는 고급 생성형 AI 어시스턴트에 대한 액세스를 제공하여 데이터 과학자와 AI 엔지니어를 지원합니다. Amazon Q는 사용자 질문과 코드 오류를 빠르고 정확하게 해결할 수 있으므로 LLM 개발 프로세스를 최적화하는 데 도움이 될 수 있습니다. Amazon Q는 Amazon Bedrock 파운데이션 모델을 사용하는 애플리케이션을 생성하는 개발자에게 상당한 이점을 제공합니다. 워크플로를 간소화하고 코드 품질을 향상시킬 수 있습니다. Python 스크립트 및 코드형 인프라(IaC) 구성 생성을 자동화하여 개발 시간과 노력을 크게 줄입니다. Amazon Q는 고급 리팩터링 기능을 통해 코드 성능을 개선하고, 보안 취약성을 식별하고, 개발자가 모범 사례를 준수하도록 할 수 있습니다. 또한 컨텍스트 인식 제안 및 설명을 제공하여 초보자의 학습 및 채택을 용이하게 하여 복잡한 코딩 작업을 더 쉽게 액세스하고 효율적으로 만들 수 있습니다.

다중 검색기 RAG 설계

생성형 AI 애플리케이션에서 다중 검색기 RAG 파이프라인은 여러 데이터 소스에서 정보를 효율적으로 검색하여 의료 서비스 제공자와 임상가가 의료 질문에 답변하는 데 도움이 될 수 있습니다. 이 파이프라인은 다양한 유형의 리트리버를 사용하여 다양한 지식 기반에서 관련 데이터를 가져옵니다. 각 리트리버는 환자 기록, 진단 인사이트, 임상 정보 또는 의학 연구 및 학술 텍스트의 콘텐츠와 같은 특정 유형의 정보를 가져오는 데 특화되어 있습니다.

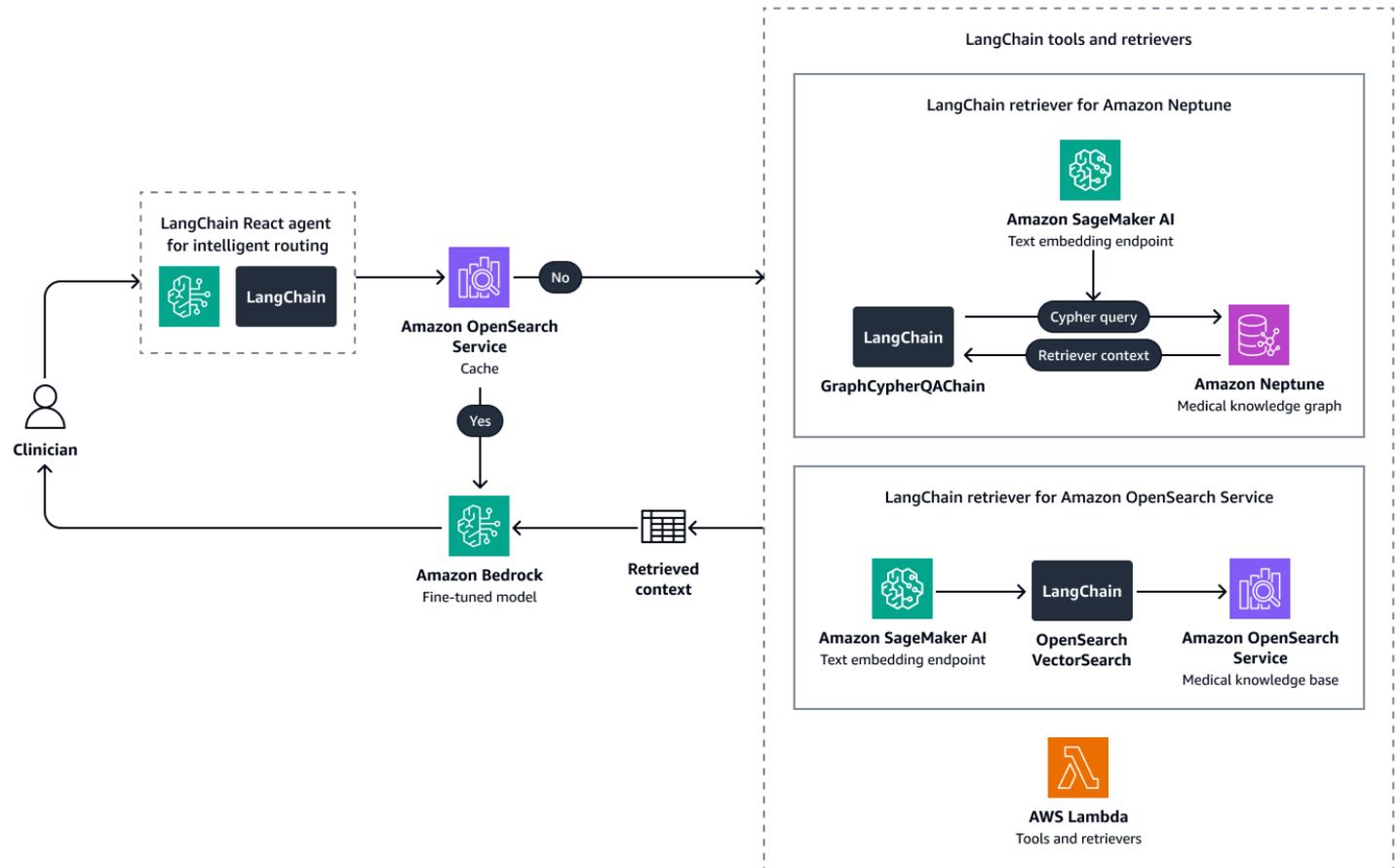
데이터의 특성과 특정 애플리케이션 요구 사항을 사용하여 사용 사례에 맞는 올바른 백엔드 지식 기반을 결정합니다. Amazon OpenSearch Service 벡터 데이터베이스는 이미지 진단 평가 요약, 퇴원 요약,

임상 보고서, 의학 연구, 학술 텍스트 콘텐츠를 포함한 대량의 비정형 또는 반정형 의료 데이터에 매우 적합합니다. 반면 Amazon Neptune과 같은 그래프 데이터베이스 서비스는 환자, 환자 기록, 의료 서비스 제공자, 약물, 증상, 치료 등 엔터티 간의 시간적 관계를 심층적으로 탐색해야 하는 의료 사용 사례에 적합할 수 있습니다.

이 파이프라인의 중요한 구성 요소는 사용자 쿼리 의도 예측입니다. 이렇게 하면 시스템이 쿼리를 올바른 리트리버 체인으로 라우팅합니다. 예를 들어 임상가가 환자의 치료 기록, 증상, 병원과의 상호 작용, 병원 재입원 가능성 또는 잠재적 환자 결과에 대해 묻는 경우 쿼리 의도 예측 모듈은 이 의도를 식별합니다. 의료 지식 그래프에서 환자 레코드 또는 시간별 치료 데이터를 가져올 수 있는 리트리버 체인으로 요청을 보냅니다. 또는 질병 발견, 특정 진단 평가 또는 학술 교과서의 특정 임상 절차 세부 정보에 대한 질문이 있는 경우 쿼리는 OpenSearch Service 벡터 데이터베이스에서 이 정보를 가져올 수 있는 리트리버 체인으로 라우팅됩니다. 의 [도구 호출](#) 기능을 사용하여 사용자 질문을 사전 정의된 의도로 분류할 수 있는 Amazon Bedrock LLM에 사용자 지정 도구를 바인딩 LangChain할 수 있습니다.

이 다중 검색 RAG 시스템에는 특정 지식 기반에 대한 액세스를 관리하도록 설계된 LangChain 에이전트가 포함되어 있습니다. LangChain를 사용하여 Amazon Bedrock LLM, 다양한 리트리버 및 도구 간의 상호 작용을 오케스트레이션할 수 있습니다. LangChain에는 의도 분류기, Neptune용 리트리버, OpenSearch Service용 리트리버 또는 사용자 의도를 분류하고 특정 지식 기반의 데이터에 구조화된 형식으로 액세스하기 위해 개발할 수 있는 기타 도구와 같은 사용자 지정 도구를 생성하는 데 도움이 되는 도구 호출 클래스가 포함되어 있습니다. 그런 다음 이러한 도구를 클래스에 공급하여 추론 및 조치(ReAct) 에이전트를 생성합니다. ReAct 에이전트는 사용자 질문을 처리하고 질문에 답변하기 위한 순차적 단계를 계획한 다음 사용 가능한 도구를 반복적으로 실행하고 도구 응답을 처리하여 사용자 쿼리에 최종적으로 답변합니다.

다음 이미지는 효율적인 지식 검색 및 지능형 쿼리 해결을 위해 설계된 다중 검색 RAG 시스템이 작동하는 방식을 보여줍니다. LangChain ReAct 에이전트는 사용자의 의도를 분석하고, 실행을 위한 구조화된 계획을 수립하고, 가장 관련성이 높은 검색 도구를 선택합니다. 시스템은 이전 질문 캐시를 쿼리하고 환자 ID, 의학적 상태, 방문 날짜와 같은 주요 속성을 기반으로 유사한 쿼리를 확인합니다. 매우 유사한 질문이 발견되면 해당 답변이 직접 검색됩니다. 그렇지 않으면 에이전트가 적절한 리트리버를 실행합니다. 시스템은 치료 기록, 증상, 병원 상호 작용 또는 재입원 가능성과 같은 환자 중심 정보를 검색하기 위해 그래프 리트리버를 사용합니다. 진단 평가, 임상 절차 및 구조화된 의료 결과를 위해 에이전트는 벡터 데이터베이스 리트리버를 사용합니다. 두 데이터 스토어의 컨텍스트 지식을 조합하여 포괄적인 응답을 생성해야 하는 시나리오에서 시스템은 지식 그래프와 벡터 데이터베이스의 결과를 통합하는 하이브리드 검색 전략을 사용합니다.



ReAct 에이전트

추론 및 조치(ReAct) 에이전트는 다면 RAG 애플리케이션을 위해 설계되었습니다. 이러한 에이전트는 추론 및 동적 작업의 강력한 조합을 제공하며, 특히 step-by-step 논리적 정보 검색 워크플로가 포함된 복잡한 애플리케이션에 적합합니다. 자세한 내용은 [ReAct: Synergizing Reasoning and Acting in Language Models](#)를 참조하세요.

의료 및 의료 맥락에서 임상 의 또는 의사의 쿼리는 종종 다면적입니다. 예를 들어, 임상 의는 "향상성 당뇨병과 제2형 당뇨병이 모두 있는 유사한 환자에게 어떤 치료를 받았나요?"라고 질문할 수 있습니다. 사용자 의도를 식별한 후, 즉 당뇨병 및 제2형 당뇨병에 대한 치료를 가져오려면 AI 에이전트가 쿼리를 하위 작업으로 나눈 다음 가장 효율적인 검색 전략을 선택해야 합니다. 이 경우 AI 에이전트는 가장 관련성이 높은 노드(예: 환자 연령, 성별, 상태, 치료 및 약물)를 식별한 다음 그래프에서 이러한 개체와 해당 속성 및 관계를 쿼리해야 합니다. ReAct 에이전트는 LLM의 추론(논리 추론) 기능을 작업(외부 리소스 또는 지식 기반 쿼리 또는 상호 작용)과 결합하기 때문에 매우 유용합니다.

사용자 쿼리 "어떤 치료를 받았습니까?"에 답하기 위해 다음 예제에서는 ReAct 에이전트의 작동 방식을 보여줍니다.

1. 에이전트 추론 - ReAct 에이전트는 질문에 조건(당뇨병 및 초조)에 대한 정보 검색이 포함되어 있다고 추론합니다. 환자 연령, 치료, 약물 및 분석할 기간을 고려합니다.
2. 에이전트 작업 - 에이전트는 openCypher를 사용하여 지식 그래프에서 제2형 당뇨병 및 당뇨병과 관련된 처리를 쿼리합니다. 또한 투여 약물, 병원 방문 날짜, 약물 부작용, 알려진 환자 결과, 유사한 환자(예: 동일한 성별 및 연령의 환자)에 대한 교차 참조 데이터도 검색합니다.
3. 에이전트 관측치 - 지식 그래프에서 에이전트는 최근 6개월간의 표 형식 데이터를 검색하여 당뇨병과 제2형 당뇨병 환자에게 제공된 치료에 대한 데이터를 검색합니다.
4. 에이전트 추론 - 검색된 레코드의 결과에 순위를 매기기 위해 에이전트는 최신성, 약물 부작용 또는 알려진 환자 결과와 같은 중요한 속성을 식별합니다.
5. 에이전트 작업 - 에이전트는 식별된 속성과 시스템 프롬프트를 통해 부여된 사전 정의된 로직을 기반으로 레코드의 순위를 다시 매깁니다.
6. 응답 생성 - Amazon Bedrock의 LLM은 ReAct 에이전트가 준비한 컨텍스트를 기반으로 응답을 생성합니다.

의료용 생성형 AI 솔루션 평가

구축한 의료 AI 솔루션을 평가하는 것은 실제 의료 환경에서 효과적이고 안정적이며 확장 가능한지 확인하는 데 매우 중요합니다. 체계적인 접근 방식을 사용하여 솔루션의 각 구성 요소의 성능을 평가합니다. 다음은 솔루션을 평가하는 데 사용할 수 있는 방법론 및 지표에 대한 요약입니다.

주제

- [정보 추출 평가](#)
- [여러 리트리버를 사용하여 RAG 솔루션 평가](#)
- [LLM을 사용하여 솔루션 평가](#)

정보 추출 평가

[지능형 재개 구문 분석기](#) 및 [사용자 지정 개체 추출기](#)와 같은 정보 추출 솔루션의 성능을 평가합니다. 테스트 데이터 세트를 사용하여 이러한 솔루션의 응답 정렬을 측정할 수 있습니다. 다양한 의료 인체 프로필 및 환자 의료 기록을 다루는 데이터 세트가 없는 경우 LLM의 추론 기능을 사용하여 사용자 지정 테스트 데이터 세트를 생성할 수 있습니다. 예를 들어 모델과 같은 대규모 파라미터 Anthropic Claude 모델을 사용하여 테스트 데이터 세트를 생성할 수 있습니다.

다음은 정보 추출 모델을 평가하는 데 사용할 수 있는 세 가지 주요 지표입니다.

- 정확성 및 완전성 - 이러한 지표는 출력이 실측 데이터에 있는 정확하고 완전한 정보를 캡처한 정도를 평가합니다. 여기에는 추출된 정보의 정확성과 추출된 정보에 모든 관련 세부 정보가 있는지 확인하는 작업이 포함됩니다.
- 유사성 및 관련성 - 이러한 지표는 출력과 실측 데이터 간의 의미, 구조적 및 컨텍스트 유사성(유사성)과 출력이 실측 데이터의 콘텐츠, 컨텍스트 및 의도와 일치하고 이를 해결하는 정도(관련성)를 평가합니다.
- 조정된 재현율 또는 캡처 속도 - 이러한 속도는 실제 데이터에서 모델에 의해 올바르게 식별된 현재 값의 수를 경험적으로 결정합니다. 비율에는 모델이 추출하는 모든 거짓 값에 대한 페널티가 포함되어야 합니다.
- 정밀도 점수 - 정밀도 점수는 참 긍정과 비교하여 예측에 존재하는 거짓 긍정 수를 결정하는 데 도움이 됩니다. 예를 들어 정밀도 지표를 사용하여 추출된 스킵 속련도의 정확성을 측정할 수 있습니다.

여러 리트리버를 사용하여 RAG 솔루션 평가

시스템이 관련 정보를 얼마나 잘 검색하고 해당 정보를 얼마나 효과적으로 사용하여 정확하고 상황에 맞는 응답을 생성하는지 평가하려면 다음 지표를 사용할 수 있습니다.

- **응답 관련성** - 검색된 컨텍스트를 사용하는 생성된 응답이 원래 쿼리와 얼마나 관련이 있는지 측정합니다.
- **컨텍스트 정밀도** - 검색된 총 결과 중에서 쿼리와 관련된 검색된 문서 또는 코드 조각의 비율을 평가합니다. 컨텍스트 정밀도가 높을수록 검색 메커니즘이 관련 정보를 선택하는 데 효과적임을 나타냅니다.
- **충실도** - 생성된 응답이 검색된 컨텍스트의 정보를 얼마나 정확하게 반영하는지 평가합니다. 즉, 소스 정보에 대한 응답이 true로 유지되는지 측정합니다.

LLM을 사용하여 솔루션 평가

LLM-as-a-judge라는 기법을 사용하여 생성형 AI 솔루션의 텍스트 응답을 평가할 수 있습니다. 여기에는 LLMs 사용하여 모델 출력의 성능을 평가하고 평가하는 작업이 포함됩니다. 이 기법은 Amazon Bedrock의 기능을 사용하여 사람의 선호도 또는 실측 데이터에 대한 응답 품질, 일관성, 준수, 정확성 및 완전성과 같은 다양한 속성에 대한 판단을 제공합니다. 포괄적인 평가를 위해 사고 [chain-of-thought\(CoT\)](#) 및 [스크린샷이 거의 없는](#) 프롬프트 기법을 사용합니다. 프롬프트는 LLM에 점수 마찰로 생성된 응답을 평가하도록 지시하고 프롬프트의 몇 번의 샘플은 실제 평가 프로세스를 보여줍니다. 프롬프트에는 LLM 평가자가 따라야 할 지침도 포함되어 있습니다. 예를 들어, LLM을 사용하여 생성된 응답을 판단하는 다음 평가 기법 중 하나 이상을 사용하는 것을 고려할 수 있습니다.

- **쌍 비교** - LLM 평가자에게 생성한 RAG 시스템의 다양한 반복 버전에서 생성된 의료 질문과 여러 응답을 제공합니다. LLM 평가자에게 응답 품질, 일관성 및 원래 질문 준수를 기반으로 최상의 응답을 결정하도록 유도합니다.
- **단일 답변 그레이딩** - 이 기법은 환자 결과 분류, 환자 행동 분류, 환자 재입원 가능성 및 위험 분류와 같은 분류의 정확도를 평가해야 하는 사용 사례에 적합합니다. LLM 평가자를 사용하여 개별 분류 또는 분류를 개별적으로 분석하고 실제 데이터와 비교하여 제공한 추론을 평가합니다.
- **참조 기반 등급 지정** - LLM 평가자에게 설명 답변이 필요한 일련의 의료 질문을 제공합니다. 참조 답변 또는 이상적인 응답과 같이 이러한 질문에 대한 샘플 응답을 생성합니다. LLM 평가자에게 LLM 생성 응답을 참조 답변 또는 이상적인 응답과 비교하도록 유도하고, LLM 평가자에게 정확성, 완전성, 유사성, 관련성 또는 기타 속성에 대해 생성된 응답의 등급을 매기도록 유도합니다. 이 기법은 생성된 응답이 잘 정의된 표준 또는 예시적 답변과 일치하는지 평가하는 데 도움이 됩니다.

리소스

AWS 설명서

- [Amazon Bedrock 설명서](#)
- [Amazon Neptune 설명서](#)
- [Amazon OpenSearch Service 설명서](#)
- [Amazon Neptune용 AWS Well-Architected 프레임워크 적용\(AWS 권장 가이드\)](#)
- [Amazon OpenSearch Service 운영 모범 사례\(OpenSearch Service 설명서\)](#)
- [의료 및 생명과학에 Amazon Comprehend Medical 및 LLMs 사용\(AWS 권장 가이드\)](#)

AWS 블로그 게시물

- [Amazon Bedrock에서 사용할 수 있는 새로운 Amazon Titan Text Premier 모델로 RAG 및 에이전트 기반 생성형 AI 애플리케이션 구축](#)
- [Amazon Neptune을 사용하여 데이터 웨어하우스에서 지식 그래프를 구축하여 상용 인텔리전스 보완](#)
- [지식 그래프를 사용하여 Amazon Bedrock 및 Amazon Neptune으로 GraphRAG 애플리케이션 구축](#)

기타 리소스

- [검색 증강 세대를 신장학의 대규모 언어 모델과 통합: 실용적인 애플리케이션 개선\(PubMed Central, National Library of Medicine\)](#)
- [소개LangChain\(LangChain 설명서\)](#)

기여자

작성

- Nitu Nivedita, 상무이사 - Accenture의 데이터 및 AI, 인공 지능 책임자
- Manoj Appully, Cadiem의 설립자 및 CTO
- Conor Folan, 컨설턴트 - 데이터 및 AI, Accenture
- Deepak Krishna AR, 컨설턴트 - 데이터 및 AI, Accenture
- Almore Cato, 관리자 - 데이터 및 AI, Accenture
- Soonam Kurian, Principal Solutions Architect AWS

검토

- Sally Lin, 데이터 과학 선임 관리자 - Accenture의 데이터 및 AI
- Terry Huang, 데이터 과학 관리자 - Accenture의 데이터 및 AI
- William Lorenz, 파트너 솔루션 아키텍트, AWS

기술 작성

- GxP AbouHarb, Senior Technical Writer, AWS

문서 기록

아래 표에 이 가이드의 주요 변경 사항이 설명되어 있습니다. 향후 업데이트에 대한 알림을 받으려면 [RSS 피드](#)를 구독하십시오.

변경 사항	설명	날짜
최초 게시	—	2025년 3월 14일

AWS 권장 가이드 용어집

다음은 AWS 권장 가이드에서 제공하는 전략, 가이드 및 패턴에서 일반적으로 사용되는 용어입니다. 용어집 항목을 제안하려면 용어집 끝에 있는 피드백 제공 링크를 사용하십시오.

숫자

7가지 전략

애플리케이션을 클라우드로 이전하기 위한 7가지 일반적인 마이그레이션 전략 이러한 전략은 Gartner가 2011년에 파악한 5가지 전략을 기반으로 하며 다음으로 구성됩니다.

- 리팩터링/리아키텍트 - 클라우드 네이티브 기능을 최대한 활용하여 애플리케이션을 이동하고 해당 아키텍처를 수정함으로써 민첩성, 성능 및 확장성을 개선합니다. 여기에는 일반적으로 운영 체제와 데이터베이스 이식이 포함됩니다. 예: 온프레미스 Oracle 데이터베이스를 Amazon Aurora PostgreSQL 호환 버전으로 마이그레이션합니다.
- 리플랫폼(리프트 앤드 리세이프) - 애플리케이션을 클라우드로 이동하고 일정 수준의 최적화를 도입하여 클라우드 기능을 활용합니다. 예:에서 온프레미스 Oracle 데이터베이스를 Oracle용 Amazon Relational Database Service(RDS)로 마이그레이션합니다 AWS 클라우드.
- 재구매(드롭 앤드 쇼) - 일반적으로 기존 라이선스에서 SaaS 모델로 전환하여 다른 제품으로 전환합니다. 예: 고객 관계 관리(CRM) 시스템을 Salesforce.com 마이그레이션합니다.
- 리호스팅(리프트 앤드 시프트) - 애플리케이션을 변경하지 않고 클라우드로 이동하여 클라우드 기능을 활용합니다. 예:의 EC2 인스턴스에서 온프레미스 Oracle 데이터베이스를 Oracle로 마이그레이션합니다 AWS 클라우드.
- 재배포(하이퍼바이저 수준의 리프트 앤 시프트) - 새 하드웨어를 구매하거나, 애플리케이션을 다시 작성하거나, 기존 운영을 수정하지 않고도 인프라를 클라우드로 이동합니다. 온프레미스 플랫폼에서 동일한 플랫폼의 클라우드 서비스로 서버를 마이그레이션합니다. 예: Microsoft Hyper-V 애플리케이션을 로 마이그레이션합니다 AWS.
- 유지(보관) - 소스 환경에 애플리케이션을 유지합니다. 대규모 리팩터링이 필요하고 해당 작업을 나중에 연기하려는 애플리케이션과 비즈니스 차원에서 마이그레이션할 이유가 없어 유지하려는 레거시 애플리케이션이 여기에 포함될 수 있습니다.
- 사용 중지 - 소스 환경에서 더 이상 필요하지 않은 애플리케이션을 폐기하거나 제거합니다.

A

ABAC

[속성 기반 액세스 제어를](#) 참조하세요.

추상화된 서비스

[관리형 서비스를](#) 참조하세요.

ACID

[원자성, 일관성, 격리, 내구성](#)을 참조하세요.

능동-능동 마이그레이션

양방향 복제 도구 또는 이중 쓰기 작업을 사용하여 소스 데이터베이스와 대상 데이터베이스가 동기화된 상태로 유지되고, 두 데이터베이스 모두 마이그레이션 중 연결 애플리케이션의 트랜잭션을 처리하는 데이터베이스 마이그레이션 방법입니다. 이 방법은 일회성 전환이 필요한 대신 소규모의 제어된 배치로 마이그레이션을 지원합니다. 더 유연하지만 [액티브-패시브 마이그레이션](#)보다 더 많은 작업이 필요합니다.

능동-수동 마이그레이션

소스 데이터베이스와 대상 데이터베이스가 동기화된 상태로 유지되지만 소스 데이터베이스만 연결 애플리케이션의 트랜잭션을 처리하고 데이터는 대상 데이터베이스로 복제되는 데이터베이스 마이그레이션 방법입니다. 대상 데이터베이스는 마이그레이션 중 어떤 트랜잭션도 허용하지 않습니다.

집계 함수

행 그룹에서 작동하고 그룹의 단일 반환 값을 계산하는 SQL 함수입니다. 집계 함수의 예로는 SUM 및 MAX가 있습니다.

AI

[인공 지능](#)을 참조하세요.

AIOps

[인공 지능 작업을](#) 참조하세요.

익명화

데이터세트에서 개인 정보를 영구적으로 삭제하는 프로세스입니다. 익명화는 개인 정보 보호에 도움이 될 수 있습니다. 익명화된 데이터는 더 이상 개인 데이터로 간주되지 않습니다.

안티 패턴

솔루션이 다른 솔루션보다 비생산적이거나 비효율적이거나 덜 효과적이어서 반복되는 문제에 자주 사용되는 솔루션입니다.

애플리케이션 제어

맬웨어로부터 시스템을 보호하기 위해 승인된 애플리케이션만 사용할 수 있는 보안 접근 방식입니다.

애플리케이션 포트폴리오

애플리케이션 구축 및 유지 관리 비용과 애플리케이션의 비즈니스 가치를 비롯하여 조직에서 사용하는 각 애플리케이션에 대한 세부 정보 모음입니다. 이 정보는 [포트폴리오 검색 및 분석 프로세스](#)의 핵심이며 마이그레이션, 현대화 및 최적화할 애플리케이션을 식별하고 우선순위를 정하는 데 도움이 됩니다.

인공 지능

컴퓨터 기술을 사용하여 학습, 문제 해결, 패턴 인식 등 일반적으로 인간과 관련된 인지 기능을 수행하는 것을 전문으로 하는 컴퓨터 과학 분야입니다. 자세한 내용은 [What is Artificial Intelligence?](#)를 참조하십시오.

인공 지능 운영(AIOps)

기계 학습 기법을 사용하여 운영 문제를 해결하고, 운영 인시던트 및 사용자 개입을 줄이고, 서비스 품질을 높이는 프로세스입니다. AWS 마이그레이션 전략에서 AIOps가 사용되는 방법에 대한 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

비대칭 암호화

한 쌍의 키, 즉 암호화를 위한 퍼블릭 키와 복호화를 위한 프라이빗 키를 사용하는 암호화 알고리즘입니다. 퍼블릭 키는 복호화에 사용되지 않으므로 공유할 수 있지만 프라이빗 키에 대한 액세스는 엄격히 제한되어야 합니다.

원자성, 일관성, 격리성, 내구성(ACID)

오류, 정전 또는 기타 문제가 발생한 경우에도 데이터베이스의 데이터 유효성과 운영 신뢰성을 보장하는 소프트웨어 속성 세트입니다.

ABAC(속성 기반 액세스 제어)

부서, 직무, 팀 이름 등의 사용자 속성을 기반으로 세분화된 권한을 생성하는 방식입니다. 자세한 내용은 AWS Identity and Access Management (IAM) 설명서의 [용 ABAC AWS](#)를 참조하세요.

신뢰할 수 있는 데이터 소스

가장 신뢰할 수 있는 정보 소스로 간주되는 기본 버전의 데이터를 저장하는 위치입니다. 익명화, 편집 또는 가명화와 같은 데이터 처리 또는 수정의 목적으로 신뢰할 수 있는 데이터 소스의 데이터를 다른 위치로 복사할 수 있습니다.

가용 영역

다른 가용 영역의 장애로부터 격리 AWS 리전 되고 동일한 리전의 다른 가용 영역에 저렴하고 지연 시간이 짧은 네트워크 연결을 제공하는 내의 고유한 위치입니다.

AWS 클라우드 채택 프레임워크(AWS CAF)

조직이 클라우드로 성공적으로 전환 AWS 하기 위한 효율적이고 효과적인 계획을 개발하는 데 도움이 되는 지침 및 모범 사례 프레임워크입니다. AWS CAF는 지침을 비즈니스, 사람, 거버넌스, 플랫폼, 보안 및 운영이라는 6가지 중점 영역으로 구성합니다. 비즈니스, 사람 및 거버넌스 관점은 비즈니스 기술과 프로세스에 초점을 맞추고, 플랫폼, 보안 및 운영 관점은 전문 기술과 프로세스에 중점을 둡니다. 예를 들어, 사람 관점은 인사(HR), 직원 배치 기능 및 인력 관리를 담당하는 이해관계자를 대상으로 합니다. 이러한 관점에서 AWS CAF는 성공적인 클라우드 채택을 위해 조직을 준비하는 데 도움이 되는 인력 개발, 교육 및 커뮤니케이션에 대한 지침을 제공합니다. 자세한 내용은 [AWS CAF 웹 사이트](#)와 [AWS CAF 백서](#)를 참조하십시오.

AWS 워크로드 검증 프레임워크(AWS WQF)

데이터베이스 마이그레이션 워크로드를 평가하고, 마이그레이션 전략을 권장하고, 작업 견적을 제공하는 도구입니다. AWS WQF는 AWS Schema Conversion Tool (AWS SCT)에 포함되어 있습니다. 데이터베이스 스키마 및 코드 객체, 애플리케이션 코드, 종속성 및 성능 특성을 분석하고 평가 보고서를 제공합니다.

B

잘못된 봇

개인 또는 조직을 방해하거나 해를 입히기 위한 [봇](#)입니다.

BCP

[비즈니스 연속성 계획을](#) 참조하세요.

동작 그래프

리소스 동작과 시간 경과에 따른 상호 작용에 대한 통합된 대화형 뷰입니다. Amazon Detective에서 동작 그래프를 사용하여 실패한 로그인 시도, 의심스러운 API 호출 및 유사한 작업을 검사할 수 있습니다. 자세한 내용은 Detective 설명서의 [Data in a behavior graph](#)를 참조하십시오.

빅 엔디안 시스템

가장 중요한 바이트를 먼저 저장하는 시스템입니다. [Endianness](#)도 참조하세요.

바이너리 분류

바이너리 결과(가능한 두 클래스 중 하나)를 예측하는 프로세스입니다. 예를 들어, ML 모델이 “이 이메일이 스팸인가요, 스팸이 아닌가요?”, ‘이 제품은 책인가요, 자동차인가요?’ 등의 문제를 예측해야 할 수 있습니다.

블룸 필터

요소가 세트의 멤버인지 여부를 테스트하는 데 사용되는 메모리 효율성이 높은 확률론적 데이터 구조입니다.

블루/그린(Blue/Green) 배포

별개의 동일한 두 환경을 생성하는 배포 전략입니다. 현재 애플리케이션 버전은 한 환경(파란색)에서 실행하고 새 애플리케이션 버전은 다른 환경(녹색)에서 실행합니다. 이 전략을 사용하면 영향을 최소화하면서 빠르게 롤백할 수 있습니다.

bot

인터넷을 통해 자동화된 작업을 실행하고 인적 활동 또는 상호 작용을 시뮬레이션하는 소프트웨어 애플리케이션입니다. 인터넷에서 정보를 인덱싱하는 웹 크롤러와 같은 일부 봇은 유용하거나 유용합니다. 잘못된 봇이라고 하는 일부 다른 봇은 개인 또는 조직을 방해하거나 해를 입히기 위한 것입니다.

봇넷

[맬웨어](#)에 감염되어 [있고 봇](#) 셰이더 또는 봇 운영자라고 하는 단일 당사자가 제어하는 봇 네트워크입니다. Botnet은 봇과 봇의 영향을 확장하는 가장 잘 알려진 메커니즘입니다.

브랜치

코드 리포지토리의 포함된 영역입니다. 리포지토리에 생성되는 첫 번째 브랜치가 기본 브랜치입니다. 기존 브랜치에서 새 브랜치를 생성한 다음 새 브랜치에서 기능을 개발하거나 버그를 수정할 수 있습니다. 기능을 구축하기 위해 생성하는 브랜치를 일반적으로 기능 브랜치라고 합니다. 기능을 출시할 준비가 되면 기능 브랜치를 기본 브랜치에 다시 병합합니다. 자세한 내용은 [About branches](#)(GitHub 설명서)를 참조하십시오.

브레이크 글래스 액세스

예외적인 상황에서 승인된 프로세스를 통해 사용자가 일반적으로 액세스할 권한이 없는데 액세스할 수 있는 빠른 방법입니다. 자세한 내용은 Well-Architected 지침의 [깨진 절차 구현](#) 표 시기를 AWS 참조하세요.

브라운필드 전략

사용자 환경의 기존 인프라 시스템 아키텍처에 브라운필드 전략을 채택할 때는 현재 시스템 및 인프라의 제약 조건을 중심으로 아키텍처를 설계합니다. 기존 인프라를 확장하는 경우 브라운필드 전략과 [그린필드](#) 전략을 혼합할 수 있습니다.

버퍼 캐시

가장 자주 액세스하는 데이터가 저장되는 메모리 영역입니다.

사업 역량

기업이 가치를 창출하기 위해 하는 일(예: 영업, 고객 서비스 또는 마케팅)입니다. 마이크로서비스 아키텍처 및 개발 결정은 비즈니스 역량에 따라 이루어질 수 있습니다. 자세한 내용은 백서의 [AWS에서 컨테이너화된 마이크로서비스 실행의 비즈니스 역량 중심의 구성화](#) 섹션을 참조하십시오.

비즈니스 연속성 계획(BCP)

대규모 마이그레이션과 같은 중단 이벤트가 운영에 미치는 잠재적 영향을 해결하고 비즈니스가 신속하게 운영을 재개할 수 있도록 지원하는 계획입니다.

C

CAF

[AWS 클라우드 채택 프레임워크](#)를 참조하세요.

canary 배포

최종 사용자에게 버전의 느린 증분 릴리스입니다. 확신이 드는 경우 새 버전을 배포하고 현재 버전을 완전히 교체합니다.

CCoE

[Cloud Center of Excellence](#)를 참조하세요.

CDC

[변경 데이터 캡처](#)를 참조하세요.

변경 데이터 캡처(CDC)

데이터베이스 테이블과 같은 데이터 소스의 변경 내용을 추적하고 변경 사항에 대한 메타데이터를 기록하는 프로세스입니다. 대상 시스템의 변경 내용을 감사하거나 복제하여 동기화를 유지하는 등의 다양한 용도로 CDC를 사용할 수 있습니다.

카오스 엔지니어링

시스템의 복원력을 테스트하기 위해 의도적으로 장애 또는 중단 이벤트를 도입합니다. [AWS Fault Injection Service \(AWS FIS\)](#)를 사용하여 AWS 워크로드에 스트레스를 주고 응답을 평가하는 실험을 수행할 수 있습니다.

CI/CD

[지속적 통합 및 지속적 전달](#)을 참조하세요.

분류

예측을 생성하는 데 도움이 되는 분류 프로세스입니다. 분류 문제에 대한 ML 모델은 이산 값을 예측합니다. 이산 값은 항상 서로 다릅니다. 예를 들어, 모델이 이미지에 자동차가 있는지 여부를 평가해야 할 수 있습니다.

클라이언트측 암호화

대상이 데이터를 AWS 서비스 수신하기 전에 로컬에서 데이터를 암호화합니다.

클라우드 혁신 센터(CCoE)

클라우드 모범 사례 개발, 리소스 동원, 마이그레이션 타임라인 설정, 대규모 혁신을 통한 조직 선도 등 조직 전체에서 클라우드 채택 노력을 추진하는 다분야 팀입니다. 자세한 내용은 AWS 클라우드 엔터프라이즈 전략 블로그의 [CCoE 게시물](#)을 참조하세요.

클라우드 컴퓨팅

원격 데이터 스토리지와 IoT 디바이스 관리에 일반적으로 사용되는 클라우드 기술 클라우드 컴퓨팅은 일반적으로 [엣지 컴퓨팅](#) 기술과 연결됩니다.

클라우드 운영 모델

IT 조직에서 하나 이상의 클라우드 환경을 구축, 성숙화 및 최적화하는 데 사용되는 운영 모델입니다. 자세한 내용은 [클라우드 운영 모델 구축](#)을 참조하십시오.

클라우드 채택 단계

조직이 로 마이그레이션할 때 일반적으로 거치는 4단계 AWS 클라우드:

- 프로젝트 - 개념 증명 및 학습 목적으로 몇 가지 클라우드 관련 프로젝트 실행
- 기반 - 클라우드 채택 확장을 위한 기초 투자(예: 랜딩 존 생성, CCoE 정의, 운영 모델 구축)
- 마이그레이션 - 개별 애플리케이션 마이그레이션
- Re-invention - 제품 및 서비스 최적화와 클라우드 혁신

이러한 단계는 Stephen Orban이 블로그 게시물 [The Journey Toward Cloud-First and the Stages of Adoption](#) on the AWS 클라우드 Enterprise Strategy 블로그에서 정의했습니다. AWS 마이그레이션 전략과 어떤 관련이 있는지에 대한 자세한 내용은 [마이그레이션 준비 가이드](#)를 참조하세요.

CMDB

[구성 관리 데이터베이스](#)를 참조하세요.

코드 리포지토리

소스 코드와 설명서, 샘플, 스크립트 등의 기타 자산이 버전 관리 프로세스를 통해 저장되고 업데이트되는 위치입니다. 일반적인 클라우드 리포지토리에는 GitHub 또는가 포함됩니다Bitbucket Cloud. 코드의 각 버전을 브랜치라고 합니다. 마이크로서비스 구조에서 각 리포지토리는 단일 기능 전용입니다. 단일 CI/CD 파이프라인은 여러 리포지토리를 사용할 수 있습니다.

콜드 캐시

비어 있거나, 제대로 채워지지 않았거나, 오래되었거나 관련 없는 데이터를 포함하는 버퍼 캐시입니다. 주 메모리나 디스크에서 데이터베이스 인스턴스를 읽어야 하기 때문에 성능에 영향을 미치며, 이는 버퍼 캐시에서 읽는 것보다 느립니다.

콜드 데이터

거의 액세스되지 않고 일반적으로 과거 데이터인 데이터. 이런 종류의 데이터를 쿼리할 때는 일반적으로 느린 쿼리가 허용됩니다. 이 데이터를 성능이 낮고 비용이 저렴한 스토리지 계층 또는 클래스로 옮기면 비용을 절감할 수 있습니다.

컴퓨터 비전(CV)

기계 학습을 사용하여 디지털 이미지 및 비디오와 같은 시각적 형식에서 정보를 분석하고 추출하는 [AI](#) 필드입니다. 예를 들어 Amazon SageMaker AI는 CV에 대한 이미지 처리 알고리즘을 제공합니다.

구성 드리프트

워크로드의 경우 구성이 예상 상태에서 변경됩니다. 이로 인해 워크로드가 규정 미준수가 될 수 있으며 일반적으로 점진적이고 의도하지 않습니다.

구성 관리 데이터베이스(CMDB)

하드웨어 및 소프트웨어 구성 요소와 해당 구성을 포함하여 데이터베이스와 해당 IT 환경에 대한 정보를 저장하고 관리하는 리포지토리입니다. 일반적으로 마이그레이션의 포트폴리오 검색 및 분석 단계에서 CMDB의 데이터를 사용합니다.

규정 준수 팩

규정 준수 및 보안 검사를 사용자 지정하기 위해 조합할 수 있는 AWS Config 규칙 및 수정 작업 모음입니다. YAML 템플릿을 사용하여 적합성 팩을 AWS 계정 및 리전 또는 조직 전체에 단일 엔터티로 배포할 수 있습니다. 자세한 내용은 AWS Config 설명서의 [적합성 팩](#)을 참조하세요.

지속적 통합 및 지속적 전달(CI/CD)

소프트웨어 릴리스 프로세스의 소스, 빌드, 테스트, 스테이징 및 프로덕션 단계를 자동화하는 프로세스입니다. CI/CD는 일반적으로 파이프라인으로 설명됩니다. CI/CD를 통해 프로세스를 자동화하고, 생산성을 높이고, 코드 품질을 개선하고, 더 빠르게 제공할 수 있습니다. 자세한 내용은 [지속적 전달의 이점](#)을 참조하십시오. CD는 지속적 배포를 의미하기도 합니다. 자세한 내용은 [지속적 전달\(Continuous Delivery\)](#)과 [지속적인 개발](#)을 참조하십시오.

CV

[컴퓨터 비전을](#) 참조하세요.

D

저장 데이터

스토리지에 있는 데이터와 같이 네트워크에 고정되어 있는 데이터입니다.

데이터 분류

중요도와 민감도를 기준으로 네트워크의 데이터를 식별하고 분류하는 프로세스입니다. 이 프로세스는 데이터에 대한 적절한 보호 및 보존 제어를 결정하는 데 도움이 되므로 사이버 보안 위험 관리 전략의 중요한 구성 요소입니다. 데이터 분류는 AWS Well-Architected Framework의 보안 원칙 구성 요소입니다. 자세한 내용은 [데이터 분류](#)를 참조하십시오.

데이터 드리프트

프로덕션 데이터와 ML 모델 학습에 사용된 데이터 간의 상당한 차이 또는 시간 경과에 따른 입력 데이터의 의미 있는 변화. 데이터 드리프트는 ML 모델 예측의 전반적인 품질, 정확성 및 공정성을 저하시킬 수 있습니다.

전송 중 데이터

네트워크를 통과하고 있는 데이터입니다. 네트워크 리소스 사이를 이동 중인 데이터를 예로 들 수 있습니다.

데이터 메시

중앙 집중식 관리 및 거버넌스를 통해 분산되고 분산된 데이터 소유권을 제공하는 아키텍처 프레임워크입니다.

데이터 최소화

꼭 필요한 데이터만 수집하고 처리하는 원칙입니다. 에서 데이터를 최소화하면 개인 정보 보호 위험, 비용 및 분석 탄소 발자국을 줄일 AWS 클라우드 수 있습니다.

데이터 경계

신뢰할 수 있는 자격 증명만 예상 네트워크에서 신뢰할 수 있는 리소스에 액세스하도록 하는 데 도움이 되는 AWS 환경의 예방 가드레일 세트입니다. 자세한 내용은 [데이터 경계 구축을 참조하세요 AWS](#).

데이터 사전 처리

원시 데이터를 ML 모델이 쉽게 구문 분석할 수 있는 형식으로 변환하는 것입니다. 데이터를 사전 처리한다는 것은 특정 열이나 행을 제거하고 누락된 값, 일관성이 없는 값 또는 중복 값을 처리함을 의미할 수 있습니다.

데이터 출처

라이프사이클 전반에 걸쳐 데이터의 출처와 기록을 추적하는 프로세스(예: 데이터 생성, 전송, 저장 방법).

데이터 주체

데이터를 수집 및 처리하는 개인입니다.

데이터 웨어하우스

분석과 같은 비즈니스 인텔리전스를 지원하는 데이터 관리 시스템입니다. 데이터 웨어하우스에는 일반적으로 많은 양의 기록 데이터가 포함되며 일반적으로 쿼리 및 분석에 사용됩니다.

데이터 정의 언어(DDL)

데이터베이스에서 테이블 및 객체의 구조를 만들거나 수정하기 위한 명령문 또는 명령입니다.

데이터베이스 조작 언어(DML)

데이터베이스에서 정보를 수정(삽입, 업데이트 및 삭제)하기 위한 명령문 또는 명령입니다.

DDL

[데이터베이스 정의 언어](#)를 참조하세요.

딥 앙상블

예측을 위해 여러 딥 러닝 모델을 결합하는 것입니다. 딥 앙상블을 사용하여 더 정확한 예측을 얻거나 예측의 불확실성을 추정할 수 있습니다.

딥 러닝

여러 계층의 인공 신경망을 사용하여 입력 데이터와 관심 대상 변수 간의 매핑을 식별하는 ML 하위 분야입니다.

심층 방어

네트워크와 그 안의 데이터 기밀성, 무결성 및 가용성을 보호하기 위해 컴퓨터 네트워크 전체에 일련의 보안 메커니즘과 제어를 신중하게 계층화하는 정보 보안 접근 방식입니다. 이 전략을 채택하면 AWS Organizations 구조의 여러 계층에 여러 컨트롤을 AWS 추가하여 리소스를 보호할 수 있습니다. 예를 들어, 심층 방어 접근 방식은 다단계 인증, 네트워크 세분화 및 암호화를 결합할 수 있습니다.

위임된 관리자

에서 AWS Organizations 호환되는 서비스는 AWS 멤버 계정을 등록하여 조직의 계정을 관리하고 해당 서비스에 대한 권한을 관리할 수 있습니다. 이러한 계정을 해당 서비스의 위임된 관리자라고 합니다. 자세한 내용과 호환되는 서비스 목록은 AWS Organizations 설명서의 [AWS Organizations 와 함께 사용할 수 있는 AWS 서비스](#)를 참조하십시오.

배포

대상 환경에서 애플리케이션, 새 기능 또는 코드 수정 사항을 사용할 수 있도록 하는 프로세스입니다. 배포에는 코드 베이스의 변경 사항을 구현한 다음 애플리케이션 환경에서 해당 코드베이스를 구축하고 실행하는 작업이 포함됩니다.

개발 환경

[환경](#)을 참조하세요.

탐지 제어

이벤트 발생 후 탐지, 기록 및 알림을 수행하도록 설계된 보안 제어입니다. 이러한 제어는 기존의 예방적 제어를 우회한 보안 이벤트를 알리는 2차 방어선입니다. 자세한 내용은 Implementing security controls on AWS의 [Detective controls](#)를 참조하십시오.

개발 가치 흐름 매핑 (DVSM)

소프트웨어 개발 라이프사이클에서 속도와 품질에 부정적인 영향을 미치는 제약 조건을 식별하고 우선 순위를 지정하는 데 사용되는 프로세스입니다. DVSM은 원래 린 제조 방식을 위해 설계된 가치 흐름 매핑 프로세스를 확장합니다. 소프트웨어 개발 프로세스를 통해 가치를 창출하고 이동하는 데 필요한 단계와 팀에 중점을 둡니다.

디지털 트윈

건물, 공장, 산업 장비 또는 생산 라인과 같은 실제 시스템을 가상으로 표현한 것입니다. 디지털 트윈은 예측 유지 보수, 원격 모니터링, 생산 최적화를 지원합니다.

차원 테이블

[스타 스키마](#)에서는 팩트 테이블의 정량적 데이터에 대한 데이터 속성을 포함하는 더 작은 테이블입니다. 차원 테이블 속성은 일반적으로 텍스트 필드 또는 텍스트처럼 동작하는 개별 숫자입니다. 이러한 속성은 일반적으로 쿼리 제약, 필터링 및 결과 집합 레이블 지정에 사용됩니다.

재해

워크로드 또는 시스템이 기본 배포 위치에서 비즈니스 목표를 달성하지 못하게 방해하는 이벤트입니다. 이러한 이벤트는 자연재해, 기술적 오류, 의도하지 않은 구성 오류 또는 멀웨어 공격과 같은 사람의 행동으로 인한 결과일 수 있습니다.

재해 복구(DR)

[재해](#)로 인한 가동 중지 시간과 데이터 손실을 최소화하는 데 사용하는 전략 및 프로세스입니다. 자세한 내용은 AWS Well-Architected Framework의 [Disaster Recovery of Workloads on AWS: Recovery in the Cloud](#)를 참조하세요.

DML

[데이터베이스 조작 언어](#)를 참조하세요.

도메인 기반 설계

구성 요소를 각 구성 요소가 제공하는 진화하는 도메인 또는 핵심 비즈니스 목표에 연결하여 복잡한 소프트웨어 시스템을 개발하는 접근 방식입니다. 이 개념은 에릭 에반스에 의해 그의 저서인 도메인 기반 디자인: 소프트웨어 중심의 복잡성 해결(Boston: Addison-Wesley Professional, 2003)에서 소개되었습니다. Strangler Fig 패턴과 함께 도메인 기반 설계를 사용하는 방법에 대한 자세한 내용은 [컨테이너 및 Amazon API Gateway를 사용하여 기존의 Microsoft ASP.NET\(ASMX\) 웹 서비스를 점진적으로 현대화하는 방법](#)을 참조하십시오.

DR

[재해 복구](#)를 참조하세요.

드리프트 감지

기존 구성과의 편차 추적. 예를 들어 AWS CloudFormation 를 사용하여 [시스템 리소스의 드리프트를 감지](#)하거나 사용하여 AWS Control Tower 거버넌스 요구 사항 준수에 영향을 미칠 수 있는 [랜딩 존의 변경 사항을 감지](#)할 수 있습니다.

DVSM

[개발 값 스트림 매핑](#)을 참조하세요.

E

EDA

[탐색 데이터 분석](#)을 참조하세요.

EDI

[전자 데이터 교환](#)을 참조하세요.

엣지 컴퓨팅

IoT 네트워크의 엣지에서 스마트 디바이스의 컴퓨팅 성능을 개선하는 기술 [클라우드 컴퓨팅](#)과 비교할 때 엣지 컴퓨팅은 통신 지연 시간을 줄이고 응답 시간을 개선할 수 있습니다.

전자 데이터 교환(EDI)

조직 간의 비즈니스 문서 자동 교환. 자세한 내용은 [전자 데이터 교환이란 무엇입니까?](#)를 참조하세요.

암호화

사람이 읽을 수 있는 일반 텍스트 데이터를 사이버텍스트로 변환하는 컴퓨팅 프로세스입니다.

암호화 키

암호화 알고리즘에 의해 생성되는 무작위 비트의 암호화 문자열입니다. 키의 길이는 다양할 수 있으며 각 키는 예측할 수 없고 고유하게 설계되었습니다.

엔디안

컴퓨터 메모리에 바이트가 저장되는 순서입니다. 빅 엔디안 시스템은 가장 중요한 바이트를 먼저 저장합니다. 리틀 엔디안 시스템은 가장 덜 중요한 바이트를 먼저 저장합니다.

엔드포인트

[서비스 엔드포인트](#)를 참조하세요.

엔드포인트 서비스

Virtual Private Cloud(VPC)에서 호스팅하여 다른 사용자와 공유할 수 있는 서비스입니다. 를 사용하여 엔드포인트 서비스를 생성하고 다른 AWS 계정 또는 AWS Identity and Access Management (IAM) 보안 주체에 권한을 AWS PrivateLink 부여할 수 있습니다. 이러한 계정 또는 보안 주체는 인터페이스 VPC 엔드포인트를 생성하여 엔드포인트 서비스에 비공개로 연결할 수 있습니다. 자세한 내용은 Amazon Virtual Private Cloud(VPC) 설명서의 [엔드포인트 서비스 생성](#)을 참조하십시오.

엔터프라이즈 리소스 계획(ERP)

엔터프라이즈의 주요 비즈니스 프로세스(예: 회계, [MES](#), 프로젝트 관리)를 자동화하고 관리하는 시스템입니다.

봉투 암호화

암호화 키를 다른 암호화 키로 암호화하는 프로세스입니다. 자세한 내용은 AWS Key Management Service (AWS KMS) 설명서의 [봉투 암호화](#)를 참조하세요.

환경

실행 중인 애플리케이션의 인스턴스입니다. 다음은 클라우드 컴퓨팅의 일반적인 환경 유형입니다.

- 개발 환경 - 애플리케이션 유지 관리를 담당하는 핵심 팀만 사용할 수 있는 실행 중인 애플리케이션의 인스턴스입니다. 개발 환경은 변경 사항을 상위 환경으로 승격하기 전에 테스트하는 데 사용됩니다. 이러한 유형의 환경을 테스트 환경이라고도 합니다.
- 하위 환경 - 초기 빌드 및 테스트에 사용되는 환경을 비롯한 애플리케이션의 모든 개발 환경입니다.
- 프로덕션 환경 - 최종 사용자가 액세스할 수 있는 실행 중인 애플리케이션의 인스턴스입니다. CI/CD 파이프라인에서 프로덕션 환경이 마지막 배포 환경입니다.
- 상위 환경 - 핵심 개발 팀 이외의 사용자가 액세스할 수 있는 모든 환경입니다. 프로덕션 환경, 프로덕션 이전 환경 및 사용자 수용 테스트를 위한 환경이 여기에 포함될 수 있습니다.

에픽

애자일 방법론에서 작업을 구성하고 우선순위를 정하는 데 도움이 되는 기능적 범주입니다. 에픽은 요구 사항 및 구현 작업에 대한 개괄적인 설명을 제공합니다. 예를 들어, AWS CAF 보안 에픽에는 ID 및 액세스 관리, 탐지 제어, 인프라 보안, 데이터 보호 및 인시던트 대응이 포함됩니다. AWS 마 이그레이션 전략의 에픽에 대한 자세한 내용은 [프로그램 구현 가이드](#)를 참조하십시오.

ERP

[엔터프라이즈 리소스 계획을](#) 참조하세요.

탐색 데이터 분석(EDA)

데이터 세트를 분석하여 주요 특성을 파악하는 프로세스입니다. 데이터를 수집 또는 집계한 다음 초기 조사를 수행하여 패턴을 찾고, 이상을 탐지하고, 가정을 확인합니다. EDA는 요약 통계를 계산하고 데이터 시각화를 생성하여 수행됩니다.

F

팩트 테이블

[스타 스키마](#)의 중앙 테이블입니다. 비즈니스 운영에 대한 정량적 데이터를 저장합니다. 일반적으로 팩트 테이블에는 측정값이 포함된 열과 차원 테이블에 대한 외래 키가 포함된 열의 두 가지 유형이 포함됩니다.

빠른 실패

개발 수명 주기를 줄이기 위해 자주 증분 테스트를 사용하는 철학입니다. 애자일 접근 방식의 중요한 부분입니다.

장애 격리 경계

에서 장애의 영향을 제한하고 워크로드의 복원력을 개선하는 데 도움이 되는 가용 영역, AWS 리전 컨트롤 플레인 또는 데이터 플레인과 같은 AWS 클라우드경계입니다. 자세한 내용은 [AWS 장애 격리 경계를 참조하세요](#).

기능 브랜치

[브랜치를 참조하세요](#).

기능

예측에 사용하는 입력 데이터입니다. 예를 들어, 제조 환경에서 기능은 제조 라인에서 주기적으로 캡처되는 이미지일 수 있습니다.

기능 중요도

모델의 예측에 특성이 얼마나 중요한지를 나타냅니다. 이는 일반적으로 SHAP(Shapley Additive Descriptions) 및 통합 그래디언트와 같은 다양한 기법을 통해 계산할 수 있는 수치 점수로 표현됩니다. 자세한 내용은 [기계 학습 모델 해석 가능성을 참조하세요 AWS](#).

기능 변환

추가 소스로 데이터를 보강하거나, 값을 조정하거나, 단일 데이터 필드에서 여러 정보 세트를 추출하는 등 ML 프로세스를 위해 데이터를 최적화하는 것입니다. 이를 통해 ML 모델이 데이터를 활용

할 수 있습니다. 예를 들어, 날짜 '2021-05-27 00:15:37'을 '2021년', '5월', '목', '15일'로 분류하면 학습 알고리즘이 다양한 데이터 구성 요소와 관련된 미묘한 패턴을 학습하는 데 도움이 됩니다.

몇 장의 샷 프롬프트

유사한 작업을 수행하도록 요청하기 전에 작업과 원하는 출력을 보여주는 몇 가지 예제를 [LLM](#)에 제공합니다. 이 기법은 컨텍스트 내 학습을 적용하여 모델이 프롬프트에 포함된 예제(샷)에서 학습합니다. 퓨샷 프롬프트는 특정 형식 지정, 추론 또는 도메인 지식이 필요한 작업에 효과적일 수 있습니다. [제로샷 프롬프트도 참조하세요.](#)

FGAC

[세분화된 액세스 제어를 참조하세요.](#)

세분화된 액세스 제어(FGAC)

여러 조건을 사용하여 액세스 요청을 허용하거나 거부합니다.

플래시컷 마이그레이션

단계적 접근 방식을 사용하는 대신 [변경 데이터 캡처](#)를 통해 연속 데이터 복제를 사용하여 최대한 짧은 시간 내에 데이터를 마이그레이션하는 데이터베이스 마이그레이션 방법입니다. 목표는 가동 중지 시간을 최소화하는 것입니다.

FM

[파운데이션 모델을 참조하세요.](#)

파운데이션 모델(FM)

일반화 및 레이블 지정되지 않은 데이터의 대규모 데이터 세트에 대해 훈련된 대규모 딥 러닝 신경망입니다. FMs은 언어 이해, 텍스트 및 이미지 생성, 자연어 대화와 같은 다양한 일반 작업을 수행할 수 있습니다. 자세한 내용은 [파운데이션 모델이란 무엇입니까?](#)를 참조하세요.

G

생성형 AI

대량의 데이터에 대해 훈련되었으며 간단한 텍스트 프롬프트를 사용하여 이미지, 비디오, 텍스트 및 오디오와 같은 새 콘텐츠 및 아티팩트를 생성할 수 있는 [AI](#) 모델의 하위 집합입니다. 자세한 내용은 [생성형 AI란 무엇입니까?](#)를 참조하세요.

지리적 차단

[지리적 제한을 참조하세요.](#)

지리적 제한(지리적 차단)

Amazon CloudFront에서 특정 국가의 사용자가 콘텐츠 배포에 액세스하지 못하도록 하는 옵션입니다. 허용 목록 또는 차단 목록을 사용하여 승인된 국가와 차단된 국가를 지정할 수 있습니다. 자세한 내용은 CloudFront 설명서의 [콘텐츠의 지리적 배포 제한](#)을 참조하십시오.

Gitflow 워크플로

하위 환경과 상위 환경이 소스 코드 리포지토리의 서로 다른 브랜치를 사용하는 방식입니다. Gitflow 워크플로는 레거시로 간주되며 [트렁크 기반 워크플로](#)는 현대적이고 선호하는 접근 방식입니다.

골든 이미지

시스템 또는 소프트웨어의 새 인스턴스를 배포하기 위한 템플릿으로 사용되는 시스템 또는 소프트웨어의 스냅샷입니다. 예를 들어 제조업에서는 골든 이미지를 사용하여 여러 디바이스에 소프트웨어를 프로비저닝할 수 있으며 디바이스 제조 작업의 속도, 확장성 및 생산성을 개선하는 데 도움이 됩니다.

브라운필드 전략

새로운 환경에서 기존 인프라의 부재 시스템 아키텍처에 대한 그린필드 전략을 채택할 때 [브라운필드](#)라고도 하는 기존 인프라와의 호환성 제한 없이 모든 새로운 기술을 선택할 수 있습니다. 기존 인프라를 확장하는 경우 브라운필드 전략과 그린필드 전략을 혼합할 수 있습니다.

가드레일

조직 단위(OU) 전체에서 리소스, 정책 및 규정 준수를 관리하는 데 도움이 되는 중요 규칙입니다. 예방 가드레일은 규정 준수 표준에 부합하도록 정책을 시행하며, 서비스 제어 정책과 IAM 권한 경계를 사용하여 구현됩니다. 탐지 가드레일은 정책 위반 및 규정 준수 문제를 감지하고 해결을 위한 알림을 생성하며, 이는 AWS Config Amazon GuardDuty AWS Security Hub, , AWS Trusted Advisor Amazon Inspector 및 사용자 지정 AWS Lambda 검사를 사용하여 구현됩니다.

H

HA

[고가용성을](#) 참조하세요.

이기종 데이터베이스 마이그레이션

다른 데이터베이스 엔진을 사용하는 대상 데이터베이스로 소스 데이터베이스 마이그레이션(예: Oracle에서 Amazon Aurora로) 이기종 마이그레이션은 일반적으로 리아키텍트 작업의 일부이며 스

키마를 변환하는 것은 복잡한 작업일 수 있습니다. AWS 는 스키마 변환에 도움이 되는 [AWS SCT를](#) [제공](#)합니다.

높은 가용성(HA)

문제나 재해 발생 시 개입 없이 지속적으로 운영할 수 있는 워크로드의 능력. HA 시스템은 자동으로 장애 조치되고, 지속적으로 고품질 성능을 제공하고, 성능에 미치는 영향을 최소화하면서 다양한 부하와 장애를 처리하도록 설계되었습니다.

히스토리언 현대화

제조 산업의 요구 사항을 더 잘 충족하도록 운영 기술(OT) 시스템을 현대화하고 업그레이드하는 데 사용되는 접근 방식입니다. 히스토리언은 공장의 다양한 출처에서 데이터를 수집하고 저장하는 데 사용되는 일종의 데이터베이스입니다.

홀드아웃 데이터

[기계 학습](#) 모델을 훈련하는 데 사용되는 데이터 세트에서 보류된 레이블이 지정된 기록 데이터의 일부입니다. 홀드아웃 데이터를 사용하여 모델 예측을 홀드아웃 데이터와 비교하여 모델 성능을 평가할 수 있습니다.

동종 데이터베이스 마이그레이션

동일한 데이터베이스 엔진을 공유하는 대상 데이터베이스로 소스 데이터베이스 마이그레이션(예: Microsoft SQL Server에서 Amazon RDS for SQL Server로) 동종 마이그레이션은 일반적으로 리호스팅 또는 리플랫폼 작업의 일부입니다. 네이티브 데이터베이스 유틸리티를 사용하여 스키마를 마이그레이션할 수 있습니다.

핫 데이터

자주 액세스하는 데이터(예: 실시간 데이터 또는 최근 번역 데이터). 일반적으로 이 데이터에는 빠른 쿼리 응답을 제공하기 위한 고성능 스토리지 계층 또는 클래스가 필요합니다.

핫픽스

프로덕션 환경의 중요한 문제를 해결하기 위한 긴급 수정입니다. 핫픽스는 긴급하기 때문에 일반적인 DevOps 릴리스 워크플로 외부에서 실행됩니다.

하이퍼케어 기간

전환 직후 마이그레이션 팀이 문제를 해결하기 위해 클라우드에서 마이그레이션된 애플리케이션을 관리하고 모니터링하는 기간입니다. 일반적으로 이 기간은 1~4일입니다. 하이퍼케어 기간이 끝나면 마이그레이션 팀은 일반적으로 애플리케이션에 대한 책임을 클라우드 운영 팀에 넘깁니다.

정보

laC

[코드형 인프라를 참조하세요.](#)

자격 증명 기반 정책

AWS 클라우드 환경 내에서 권한을 정의하는 하나 이상의 IAM 보안 주체에 연결된 정책입니다.

유휴 애플리케이션

90일 동안 평균 CPU 및 메모리 사용량이 5~20%인 애플리케이션입니다. 마이그레이션 프로젝트에서는 이러한 애플리케이션을 사용 중지하거나 온프레미스에 유지하는 것이 일반적입니다.

IIoT

[산업용 사물 인터넷을 참조하십시오.](#)

변경 불가능한 인프라

기존 인프라를 업데이트, 패치 적용 또는 수정하는 대신 프로덕션 워크로드를 위한 새 인프라를 배포하는 모델입니다. 변경 불가능한 인프라는 [변경 가능한 인프라](#)보다 본질적으로 더 일관되고 안정적이며 예측 가능합니다. 자세한 내용은 AWS Well-Architected Framework의 [변경 불가능한 인프라를 사용한 배포](#) 모범 사례를 참조하세요.

인바운드(수신) VPC

AWS 다중 계정 아키텍처에서 애플리케이션 외부에서 네트워크 연결을 수락, 검사 및 라우팅하는 VPC입니다. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

중분 마이그레이션

한 번에 전체 전환을 수행하는 대신 애플리케이션을 조금씩 마이그레이션하는 전환 전략입니다. 예를 들어, 처음에는 소수의 마이크로서비스나 사용자만 새 시스템으로 이동할 수 있습니다. 모든 것이 제대로 작동하는지 확인한 후에는 레거시 시스템을 폐기할 수 있을 때까지 추가 마이크로서비스 또는 사용자를 점진적으로 이동할 수 있습니다. 이 전략을 사용하면 대규모 마이그레이션과 관련된 위험을 줄일 수 있습니다.

Industry 4.0

연결성, 실시간 데이터, 자동화, 분석 및 AI/ML의 발전을 통해 제조 프로세스의 현대화를 참조하기 위해 2016년에 [Klaus Schwab](#)에서 도입한 용어입니다.

인프라

애플리케이션의 환경 내에 포함된 모든 리소스와 자산입니다.

코드형 인프라(IaC)

구성 파일 세트를 통해 애플리케이션의 인프라를 프로비저닝하고 관리하는 프로세스입니다. IaC는 새로운 환경의 반복 가능성, 신뢰성 및 일관성을 위해 인프라 관리를 중앙 집중화하고, 리소스를 표준화하고, 빠르게 확장할 수 있도록 설계되었습니다.

산업용 사물 인터넷(IIoT)

제조, 에너지, 자동차, 의료, 생명과학, 농업 등의 산업 부문에서 인터넷에 연결된 센서 및 디바이스의 사용 자세한 내용은 [산업용 사물 인터넷\(IoT\) 디지털 트랜스포메이션 전략 구축](#)을 참조하십시오.

검사 VPC

AWS 다중 계정 아키텍처에서는 VPC(동일하거나 다른 AWS 리전), 인터넷 및 온프레미스 네트워크 간의 네트워크 트래픽 검사를 관리하는 중앙 집중식 VPCs. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

사물 인터넷(IoT)

인터넷이나 로컬 통신 네트워크를 통해 다른 디바이스 및 시스템과 통신하는 센서 또는 프로세서가 내장된 연결된 물리적 객체의 네트워크 자세한 내용은 [IoT란?](#)을 참조하십시오.

해석력

모델의 예측이 입력에 따라 어떻게 달라지는지를 사람이 이해할 수 있는 정도를 설명하는 기계 학습 모델의 특성입니다. 자세한 내용은 [기계 학습 모델 해석 가능성을 참조하세요 AWS](#).

IoT

[사물 인터넷](#)을 참조하세요.

IT 정보 라이브러리(TIL)

IT 서비스를 제공하고 이러한 서비스를 비즈니스 요구 사항에 맞게 조정하기 위한 일련의 모범 사례 ITIL은 ITSM의 기반을 제공합니다.

IT 서비스 관리(TSM)

조직의 IT 서비스 설계, 구현, 관리 및 지원과 관련된 활동 클라우드 운영을 ITSM 도구와 통합하는 방법에 대한 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

ITIL

[IT 정보 라이브러리](#)를 참조하세요.

ITSM

[IT 서비스 관리](#)를 참조하세요.

L

레이블 기반 액세스 제어(LBAC)

사용자 및 데이터 자체에 각각 보안 레이블 값을 명시적으로 할당하는 필수 액세스 제어(MAC)를 구현한 것입니다. 사용자 보안 레이블과 데이터 보안 레이블 간의 교차 부분에 따라 사용자가 볼 수 있는 행과 열이 결정됩니다.

랜딩 존

랜딩 존은 확장 가능하고 안전한 잘 설계된 다중 계정 AWS 환경입니다. 조직은 여기에서부터 보안 및 인프라 환경에 대한 확신을 가지고 워크로드와 애플리케이션을 신속하게 시작하고 배포할 수 있습니다. 랜딩 존에 대한 자세한 내용은 [안전하고 확장 가능한 다중 계정 AWS 환경 설정](#)을 참조하십시오.

대규모 언어 모델(LLM)

방대한 양의 데이터를 기반으로 사전 훈련된 딥 러닝 [AI](#) 모델입니다. LLM은 질문 답변, 문서 요약, 텍스트를 다른 언어로 변환, 문장 완성과 같은 여러 작업을 수행할 수 있습니다. 자세한 내용은 [LLMs](#) 참조하십시오.

대규모 마이그레이션

300대 이상의 서버 마이그레이션입니다.

LBAC

[레이블 기반 액세스 제어를](#) 참조하세요.

최소 권한

작업을 수행하는 데 필요한 최소 권한을 부여하는 보안 모범 사례입니다. 자세한 내용은 IAM 설명서의 [최소 권한 적용](#)을 참조하십시오.

리프트 앤드 시프트

[7R](#)을 참조하세요.

리틀 엔디안 시스템

가장 덜 중요한 바이트를 먼저 저장하는 시스템입니다. [Endianness](#)도 참조하세요.

LLM

[대규모 언어 모델을](#) 참조하세요.

하위 환경

[환경을](#) 참조하세요.

M

기계 학습(ML)

패턴 인식 및 학습에 알고리즘과 기법을 사용하는 인공지능의 한 유형입니다. ML은 사물 인터넷 (IoT) 데이터와 같은 기록된 데이터를 분석하고 학습하여 패턴을 기반으로 통계 모델을 생성합니다. 자세한 내용은 [기계 학습](#)을 참조하십시오.

기본 브랜치

[브랜치를](#) 참조하세요.

맬웨어

컴퓨터 보안 또는 개인 정보 보호를 손상하도록 설계된 소프트웨어입니다. 맬웨어는 컴퓨터 시스템을 중단하거나, 민감한 정보를 유출하거나, 무단 액세스를 가져올 수 있습니다. 맬웨어의 예로는 바이러스, 웜, 랜섬웨어, 트로이 목마, 스파이웨어, 키로거 등이 있습니다.

관리형 서비스

AWS 서비스는 인프라 계층, 운영 체제 및 플랫폼을 AWS 운영하며 사용자는 엔드포인트에 액세스하여 데이터를 저장하고 검색합니다. Amazon Simple Storage Service(Amazon S3) 및 Amazon DynamoDB는 관리형 서비스의 예입니다. 이를 추상화된 서비스라고도 합니다.

제조 실행 시스템(MES)

원재료를 작업 현장의 완성된 제품으로 변환하는 생산 프로세스를 추적, 모니터링, 문서화 및 제어하기 위한 소프트웨어 시스템입니다.

MAP

[마이그레이션 가속화 프로그램을](#) 참조하세요.

메커니즘

도구를 생성하고 도구 채택을 유도한 다음 결과를 검사하여 조정하는 전체 프로세스입니다. 메커니즘은 작동 시 자체를 강화하고 개선하는 주기입니다. 자세한 내용은 AWS Well-Architected Framework의 [메커니즘 구축](#)을 참조하세요.

멤버 계정

조직의 일부인 관리 계정을 AWS 계정 제외한 모든 계정. AWS Organizations 하나의 계정은 한 번에 하나의 조직 멤버만 될 수 있습니다.

MES

[제조 실행 시스템을](#) 참조하세요.

메시지 대기열 원격 측정 전송(MQTT)

리소스가 제한된 [IoT](#) 디바이스에 대한 [게시/구독](#) 패턴을 기반으로 하는 경량 M2M(machine-to-machine) 통신 프로토콜입니다.

마이크로서비스

잘 정의된 API를 통해 통신하고 일반적으로 소규모 자체 팀이 소유하는 소규모 독립 서비스입니다. 예를 들어, 보험 시스템에는 영업, 마케팅 등의 비즈니스 역량이나 구매, 청구, 분석 등의 하위 영역에 매핑되는 마이크로 서비스가 포함될 수 있습니다. 마이크로서비스의 이점으로 민첩성, 유연한 확장, 손쉬운 배포, 재사용 가능한 코드, 복원력 등이 있습니다. 자세한 내용은 [AWS 서버리스 서비스를 사용하여 마이크로서비스 통합을 참조하세요](#).

마이크로서비스 아키텍처

각 애플리케이션 프로세스를 마이크로서비스로 실행하는 독립 구성 요소를 사용하여 애플리케이션을 구축하는 접근 방식입니다. 이러한 마이크로서비스는 경량 API를 사용하여 잘 정의된 인터페이스를 통해 통신합니다. 애플리케이션의 특정 기능에 대한 수요에 맞게 이 아키텍처의 각 마이크로 서비스를 업데이트, 배포 및 조정할 수 있습니다. 자세한 내용은 [에서 마이크로서비스 구현을 참조하세요 AWS](#).

Migration Acceleration Program(MAP)

조직이 클라우드로 전환하기 위한 강력한 운영 기반을 구축하고 초기 마이그레이션 비용을 상쇄하는 데 도움이 되는 컨설팅 지원, 교육 및 서비스를 제공하는 AWS 프로그램입니다. MAP에는 레거시 마이그레이션을 체계적인 방식으로 실행하기 위한 마이그레이션 방법론과 일반적인 마이그레이션 시나리오를 자동화하고 가속화하는 도구 세트가 포함되어 있습니다.

대규모 마이그레이션

애플리케이션 포트폴리오의 대다수를 웨이브를 통해 클라우드로 이동하는 프로세스로, 각 웨이브에서 더 많은 애플리케이션이 더 빠른 속도로 이동합니다. 이 단계에서는 이전 단계에서 배운 모범 사례와 교훈을 사용하여 팀, 도구 및 프로세스의 마이그레이션 팩토리를 구현하여 자동화 및 민첩한 제공을 통해 워크로드 마이그레이션을 간소화합니다. 이것은 [AWS 마이그레이션 전략](#)의 세 번째 단계입니다.

마이그레이션 팩토리

자동화되고 민첩한 접근 방식을 통해 워크로드 마이그레이션을 간소화하는 다기능 팀입니다. 마이그레이션 팩토리 팀에는 일반적으로 스프린트에서 일하는 운영, 비즈니스 분석가 및 소유자, 마이그레이션 엔지니어, 개발자, DevOps 전문가가 포함됩니다. 엔터프라이즈 애플리케이션 포트폴리오의 20~50%는 공장 접근 방식으로 최적화할 수 있는 반복되는 패턴으로 구성되어 있습니다. 자세한 내용은 이 콘텐츠 세트의 [클라우드 마이그레이션 팩토리 가이드](#)와 [마이그레이션 팩토리에 대한 설명](#)을 참조하십시오.

마이그레이션 메타데이터

마이그레이션을 완료하는 데 필요한 애플리케이션 및 서버에 대한 정보 각 마이그레이션 패턴에는 서로 다른 마이그레이션 메타데이터 세트가 필요합니다. 마이그레이션 메타데이터의 예로는 대상 서브넷, 보안 그룹 및 AWS 계정이 있습니다.

마이그레이션 패턴

사용되는 마이그레이션 전략, 마이그레이션 대상, 마이그레이션 애플리케이션 또는 서비스를 자세히 설명하는 반복 가능한 마이그레이션 작업입니다. 예: AWS Application Migration Service를 사용하여 Amazon EC2로 마이그레이션을 리호스팅합니다.

Migration Portfolio Assessment(MPA)

로 마이그레이션하기 위한 비즈니스 사례를 검증하기 위한 정보를 제공하는 온라인 도구입니다 AWS 클라우드. MPA는 상세한 포트폴리오 평가(서버 적정 규모 조정, 가격 책정, TCO 비교, 마이그레이션 비용 분석)와 마이그레이션 계획(애플리케이션 데이터 분석 및 데이터 수집, 애플리케이션 그룹화, 마이그레이션 우선순위 지정, 웨이브 계획)을 제공합니다. [MPA 도구](#)(로그인 필요)는 모든 AWS 컨설턴트와 APN 파트너 컨설턴트가 무료로 사용할 수 있습니다.

마이그레이션 준비 상태 평가(MRA)

AWS CAF를 사용하여 조직의 클라우드 준비 상태에 대한 인사이트를 얻고, 강점과 약점을 식별하고, 식별된 격차를 해소하기 위한 실행 계획을 수립하는 프로세스입니다. 자세한 내용은 [마이그레이션 준비 가이드](#)를 참조하십시오. MRA는 [AWS 마이그레이션 전략](#)의 첫 번째 단계입니다.

마이그레이션 전략

워크로드를 로 마이그레이션하는 데 사용되는 접근 방식입니다 AWS 클라우드. 자세한 내용은 이 용어집의 [7R 항목](#)을 참조하고 [대규모 마이그레이션을 가속화하기 위해 조직 동원을 참조하세요](#).

ML

[기계 학습](#)을 참조하세요.

현대화

비용을 절감하고 효율성을 높이고 혁신을 활용하기 위해 구식(레거시 또는 모놀리식) 애플리케이션과 해당 인프라를 클라우드의 민첩하고 탄력적이고 가용성이 높은 시스템으로 전환하는 것입니다. 자세한 내용은 [의 애플리케이션 현대화 전략을 참조하세요 AWS 클라우드](#).

현대화 준비 상태 평가

조직 애플리케이션의 현대화 준비 상태를 파악하고, 이점, 위험 및 종속성을 식별하고, 조직이 해당 애플리케이션의 향후 상태를 얼마나 잘 지원할 수 있는지를 확인하는 데 도움이 되는 평가입니다. 평가 결과는 대상 아키텍처의 청사진, 현대화 프로세스의 개발 단계와 마일스톤을 자세히 설명하는 로드맵 및 파악된 격차를 해소하기 위한 실행 계획입니다. 자세한 내용은 [의 애플리케이션에 대한 현대화 준비 상태 평가를 참조하세요 AWS 클라우드](#).

모놀리식 애플리케이션(모놀리식 유형)

긴밀하게 연결된 프로세스를 사용하여 단일 서비스로 실행되는 애플리케이션입니다. 모놀리식 애플리케이션에는 몇 가지 단점이 있습니다. 한 애플리케이션 기능에 대한 수요가 급증하면 전체 아키텍처 규모를 조정해야 합니다. 코드 베이스가 커지면 모놀리식 애플리케이션의 기능을 추가하거나 개선하는 것도 더 복잡해집니다. 이러한 문제를 해결하기 위해 마이크로서비스 아키텍처를 사용할 수 있습니다. 자세한 내용은 [마이크로서비스로 모놀리식 유형 분해](#)를 참조하십시오.

MPA

[마이그레이션 포트폴리오 평가](#)를 참조하세요.

MQTT

[메시지 대기열 원격 측정 전송](#)을 참조하세요.

멀티클래스 분류

여러 클래스에 대한 예측(2개 이상의 결과 중 하나 예측)을 생성하는 데 도움이 되는 프로세스입니다. 예를 들어, ML 모델이 '이 제품은 책인가요, 자동차인가요, 휴대폰인가요?' 또는 '이 고객이 가장 관심을 갖는 제품 범주는 무엇인가요?'라고 물을 수 있습니다.

변경 가능한 인프라

프로덕션 워크로드를 위해 기존 인프라를 업데이트하고 수정하는 모델입니다. 일관성, 신뢰성 및 예측 가능성을 높이기 위해 AWS Well-Architected Framework에서는 [변경 불가능한 인프라](#)를 모범 사례로 사용할 것을 권장합니다.

O

OAC

[오리진 액세스 제어를](#) 참조하세요.

OAI

[오리진 액세스 ID](#)를 참조하세요.

OCM

[조직 변경 관리를](#) 참조하세요.

오프라인 마이그레이션

마이그레이션 프로세스 중 소스 워크로드가 중단되는 마이그레이션 방법입니다. 이 방법은 가동 중지 증가를 수반하며 일반적으로 작고 중요하지 않은 워크로드에 사용됩니다.

OI

[작업 통합](#)을 참조하세요.

OLA

[운영 수준 계약을](#) 참조하세요.

온라인 마이그레이션

소스 워크로드를 오프라인 상태로 전환하지 않고 대상 시스템에 복사하는 마이그레이션 방법입니다. 워크로드에 연결된 애플리케이션은 마이그레이션 중에도 계속 작동할 수 있습니다. 이 방법은 가동 중지 차단 또는 최소화를 수반하며 일반적으로 중요한 프로덕션 워크로드에 사용됩니다.

OPC-UA

[Open Process Communications - Unified Architecture](#)를 참조하세요.

Open Process Communications - 통합 아키텍처(OPC-UA)

산업 자동화를 위한 M2M(Machinemachine-to-machine) 통신 프로토콜입니다. OPC-UA는 데이터 암호화, 인증 및 권한 부여 체계와 상호 운용성 표준을 제공합니다.

운영 수준 협약(OLA)

서비스 수준에 관한 계약(SLA)을 지원하기 위해 직무 IT 그룹이 서로에게 제공하기로 약속한 내용을 명확히 하는 계약입니다.

운영 준비 상태 검토(ORR)

인시던트 및 가능한 장애의 범위를 이해, 평가, 예방 또는 줄이는 데 도움이 되는 질문 체크리스트 및 관련 모범 사례입니다. 자세한 내용은 AWS Well-Architected Framework의 [운영 준비 검토\(ORR\)](#)를 참조하세요.

운영 기술(OT)

물리적 환경과 함께 작동하여 산업 운영, 장비 및 인프라를 제어하는 하드웨어 및 소프트웨어 시스템입니다. 제조에서 OT 및 정보 기술(IT) 시스템의 통합은 [Industry 4.0](#) 혁신의 핵심 초점입니다.

운영 통합(OI)

클라우드에서 운영을 현대화하는 프로세스로 준비 계획, 자동화 및 통합을 수반합니다. 자세한 내용은 [운영 통합 가이드](#)를 참조하십시오.

조직 트레일

조직 내 모든에 대한 모든 이벤트를 로깅 AWS CloudTrail 하는 AWS 계정에서 생성된 추적입니다 AWS Organizations. 이 트레일은 조직에 속한 각 AWS 계정에 생성되고 각 계정의 활동을 추적합니다. 자세한 내용은 CloudTrail 설명서의 [Creating a trail for an organization](#)을 참조하십시오.

조직 변경 관리(OCM)

사람, 문화 및 리더십 관점에서 중대하고 파괴적인 비즈니스 혁신을 관리하기 위한 프레임워크입니다. OCM은 변화 채택을 가속화하고, 과도기적 문제를 해결하고, 문화 및 조직적 변화를 주도함으로써 조직이 새로운 시스템 및 전략을 준비하고 전환할 수 있도록 지원합니다. AWS 마이그레이션 전략에서는 클라우드 채택 프로젝트에 필요한 변경 속도 때문에이 프레임워크를 인력 가속화라고 합니다. 자세한 내용은 [사용 가이드](#)를 참조하십시오.

오리진 액세스 제어(OAC)

CloudFront에서 Amazon Simple Storage Service(S3) 콘텐츠를 보호하기 위해 액세스를 제한하는 고급 옵션입니다. OAC는 AWS KMS (SSE-KMS)를 사용한 모든 서버 측 암호화 AWS 리전와 S3 버킷에 대한 동적 PUT 및 DELETE 요청에서 모든 S3 버킷을 지원합니다.

오리진 액세스 ID(OAI)

CloudFront에서 Amazon S3 콘텐츠를 보호하기 위해 액세스를 제한하는 옵션입니다. OAI를 사용하면 CloudFront는 Amazon S3가 인증할 수 있는 보안 주체를 생성합니다. 인증된 보안 주체는 특

정 CloudFront 배포를 통해서만 S3 버킷의 콘텐츠에 액세스할 수 있습니다. 더 세분화되고 향상된 액세스 제어를 제공하는 [OAC](#)도 참조하십시오.

ORR

[운영 준비 상태 검토](#)를 참조하세요.

OT

[운영 기술을](#) 참조하세요.

아웃바운드(송신) VPC

AWS 다중 계정 아키텍처에서 애플리케이션 내에서 시작된 네트워크 연결을 처리하는 VPC입니다. [AWS Security Reference Architecture](#)에서는 애플리케이션과 더 넓은 인터넷 간의 양방향 인터페이스를 보호하기 위해 인바운드, 아웃바운드 및 검사 VPC로 네트워크 계정을 설정할 것을 권장합니다.

P

권한 경계

사용자나 역할이 가질 수 있는 최대 권한을 설정하기 위해 IAM 보안 주체에 연결되는 IAM 관리 정책입니다. 자세한 내용은 IAM 설명서의 [권한 경계](#)를 참조하십시오.

개인 식별 정보(PII)

직접 보거나 다른 관련 데이터와 함께 짝을 지을 때 개인의 신원을 합리적으로 추론하는 데 사용할 수 있는 정보입니다. PII의 예로는 이름, 주소, 연락처 정보 등이 있습니다.

PII

[개인 식별 정보를](#) 참조하세요.

플레이북

클라우드에서 핵심 운영 기능을 제공하는 등 마이그레이션과 관련된 작업을 캡처하는 일련의 사전 정의된 단계입니다. 플레이북은 스크립트, 자동화된 런북 또는 현대화된 환경을 운영하는 데 필요한 프로세스나 단계 요약의 형태를 취할 수 있습니다.

PLC

[프로그래밍 가능한 로직 컨트롤러](#)를 참조하세요.

PLM

[제품 수명 주기 관리](#)를 참조하세요.

정책

권한을 정의하거나(자격 [증명 기반 정책](#) 참조), 액세스 조건을 지정하거나([리소스 기반 정책](#) 참조), 조직의 모든 계정에 대한 최대 권한을 정의할 수 있는 객체 AWS Organizations 입니다([서비스 제어 정책](#) 참조).

다국어 지속성

데이터 액세스 패턴 및 기타 요구 사항을 기반으로 독립적으로 마이크로서비스의 데이터 스토리지 기술 선택. 마이크로서비스가 동일한 데이터 스토리지 기술을 사용하는 경우 구현 문제가 발생하거나 성능이 저하될 수 있습니다. 요구 사항에 가장 적합한 데이터 스토어를 사용하면 마이크로서비스를 더 쉽게 구현하고 성능과 확장성을 높일 수 있습니다. 자세한 내용은 [마이크로서비스에서 데이터 지속성 활성화](#)를 참조하십시오.

포트폴리오 평가

마이그레이션을 계획하기 위해 애플리케이션 포트폴리오를 검색 및 분석하고 우선순위를 정하는 프로세스입니다. 자세한 내용은 [마이그레이션 준비 상태 평가](#)를 참조하십시오.

조건자

WHERE 절에서 false 일반적으로 위치한 true 또는를 반환하는 쿼리 조건입니다.

조건자 푸시다운

전송 전에 쿼리의 데이터를 필터링하는 데이터베이스 쿼리 최적화 기법입니다. 이렇게 하면 관계형 데이터베이스에서 검색하고 처리해야 하는 데이터의 양이 줄어들고 쿼리 성능이 향상됩니다.

예방적 제어

이벤트 발생을 방지하도록 설계된 보안 제어입니다. 이 제어는 네트워크에 대한 무단 액세스나 원치 않는 변경을 방지하는 데 도움이 되는 1차 방어선입니다. 자세한 내용은 Implementing security controls on AWS의 [Preventative controls](#)를 참조하십시오.

보안 주체

작업을 수행하고 리소스에 액세스할 수 있는 AWS 있는의 엔티티입니다. 이 엔티티는 일반적으로 , AWS 계정 IAM 역할 또는 사용자의 루트 사용자입니다. 자세한 내용은 IAM 설명서의 [역할 용어 및 개념](#)의 보안 주체를 참조하십시오.

설계에 따른 개인 정보 보호

전체 개발 프로세스를 통해 개인 정보를 고려하는 시스템 엔지니어링 접근 방식입니다.

프라이빗 호스팅 영역

Amazon Route 53에서 하나 이상의 VPC 내 도메인과 하위 도메인에 대한 DNS 쿼리에 응답하는 방법에 대한 정보가 담긴 컨테이너입니다. 자세한 내용은 Route 53 설명서의 [프라이빗 호스팅 영역 작업을 참조하십시오](#).

사전 예방적 제어

규정 미준수 리소스의 배포를 방지하도록 설계된 [보안 제어](#)입니다. 이러한 제어는 리소스가 프로비저닝되기 전에 리소스를 스캔합니다. 리소스가 컨트롤을 준수하지 않으면 프로비저닝되지 않습니다. 자세한 내용은 AWS Control Tower 설명서의 [제어 참조 가이드](#)를 참조하고 보안 [제어 구현의 사전](#) 예방적 제어를 참조하세요. AWS

제품 수명 주기 관리(PLM)

설계, 개발 및 출시부터 성장 및 성숙도, 거부 및 제거에 이르기까지 전체 수명 주기 동안 제품의 데이터 및 프로세스 관리.

프로덕션 환경

[환경](#)을 참조하세요.

프로그래밍 가능한 로직 컨트롤러(PLC)

제조에서 기계를 모니터링하고 제조 프로세스를 자동화하는 매우 안정적이고 적응력이 뛰어난 컴퓨터입니다.

프롬프트 체인

한 [LLM](#) 프롬프트의 출력을 다음 프롬프트의 입력으로 사용하여 더 나은 응답을 생성합니다. 이 기법은 복잡한 작업을 하위 작업으로 나누거나 예비 응답을 반복적으로 구체화하거나 확장하는 데 사용됩니다. 이를 통해 모델 응답의 정확성과 관련성을 개선하고 보다 세분화되고 개인화된 결과를 얻을 수 있습니다.

가명화

데이터세트의 개인 식별자를 자리 표시자 값으로 바꾸는 프로세스입니다. 가명화는 개인 정보를 보호하는 데 도움이 될 수 있습니다. 가명화된 데이터는 여전히 개인 데이터로 간주됩니다.

게시/구독(pub/sub)

마이크로서비스 간의 비동기 통신을 지원하여 확장성과 응답성을 개선하는 패턴입니다. 예를 들어 마이크로서비스 기반 [MES](#)에서 마이크로서비스는 다른 마이크로서비스가 구독할 수 있는 채널에 이벤트 메시지를 게시할 수 있습니다. 시스템은 게시 서비스를 변경하지 않고도 새 마이크로서비스를 추가할 수 있습니다.

Q

쿼리 계획

SQL 관계형 데이터베이스 시스템의 데이터에 액세스하는 데 사용되는 지침과 같은 일련의 단계입니다.

쿼리 계획 회귀

데이터베이스 서비스 최적화 프로그램이 데이터베이스 환경을 변경하기 전보다 덜 최적의 계획을 선택하는 경우입니다. 통계, 제한 사항, 환경 설정, 쿼리 파라미터 바인딩 및 데이터베이스 엔진 업데이트의 변경으로 인해 발생할 수 있습니다.

R

RACI 매트릭스

[책임, 책임, 상담, 정보 제공\(RACI\)을 참조하세요.](#)

RAG

[Retrieval Augmented Generation](#)을 참조하세요.

랜섬웨어

결제가 완료될 때까지 컴퓨터 시스템이나 데이터에 대한 액세스를 차단하도록 설계된 악성 소프트웨어입니다.

RASCI 매트릭스

[책임, 책임, 상담, 정보 제공\(RACI\)을 참조하세요.](#)

RCAC

[행 및 열 액세스 제어를 참조하세요.](#)

읽기 전용 복제본

읽기 전용 용도로 사용되는 데이터베이스의 사본입니다. 쿼리를 읽기 전용 복제본으로 라우팅하여 기본 데이터베이스의 로드를 줄일 수 있습니다.

재설계

[7R을 참조하세요.](#)

Recovery Point Objective(RPO)

마지막 데이터 복구 시점 이후 허용되는 최대 시간입니다. 이에 따라 마지막 복구 시점과 서비스 중단 사이에 허용되는 데이터 손실로 간주되는 범위가 결정됩니다.

Recovery Time Objective(RTO)

서비스 중단과 서비스 복원 사이의 허용 가능한 지연 시간입니다.

리팩터링

[7R을 참조하세요.](#)

리전

지리적 영역의 AWS 리소스 모음입니다. 각 AWS 리전은 내결함성, 안정성 및 복원력을 제공하기 위해 서로 격리되고 독립적입니다. 자세한 내용은 [계정에서 사용할 수 있는 지정을 참조 AWS 리전 하세요.](#)

회귀

숫자 값을 예측하는 ML 기법입니다. 예를 들어, '이 집은 얼마에 팔릴까?'라는 문제를 풀기 위해 ML 모델은 선형 회귀 모델을 사용하여 주택에 대해 알려진 사실(예: 면적)을 기반으로 주택의 매매 가격을 예측할 수 있습니다.

리호스팅

[7R을 참조하세요.](#)

release

배포 프로세스에서 변경 사항을 프로덕션 환경으로 승격시키는 행위입니다.

재배치

[7R을 참조하세요.](#)

리플랫폼

[7R을 참조하세요.](#)

재구매

[7R을 참조하세요.](#)

복원력

중단에 저항하거나 복구할 수 있는 애플리케이션의 기능입니다. 에서 복원력을 계획할 때 [고가용성](#) 및 [재해 복구](#)가 일반적인 고려 사항입니다 AWS 클라우드. 자세한 내용은 [AWS 클라우드 복원력을 참조하세요.](#)

리소스 기반 정책

Amazon S3 버킷, 엔드포인트, 암호화 키 등의 리소스에 연결된 정책입니다. 이 유형의 정책은 액세스가 허용된 보안 주체, 지원되는 작업 및 충족해야 하는 기타 조건을 지정합니다.

RACI(Responsible, Accountable, Consulted, Informed) 매트릭스

마이그레이션 활동 및 클라우드 운영에 참여하는 모든 당사자의 역할과 책임을 정의하는 매트릭스입니다. 매트릭스 이름은 매트릭스에 정의된 책임 유형에서 파생됩니다. 실무 담당자 (R), 의사 결정권자 (A), 업무 수행 조연자 (C), 결과 통보 대상자 (I). 지원자는 (S) 선택사항입니다. 지원자를 포함하면 매트릭스를 RASCI 매트릭스라고 하고, 지원자를 제외하면 RACI 매트릭스라고 합니다.

대응 제어

보안 기준에서 벗어나거나 부정적인 이벤트를 해결하도록 설계된 보안 제어입니다. 자세한 내용은 Implementing security controls on AWS의 [Responsive controls](#)를 참조하십시오.

retain

[7R을 참조하세요.](#)

사용 중지

[7R을 참조하세요.](#)

검색 증강 세대(RAG)

응답을 생성하기 전에 [LLM](#)이 훈련 데이터 소스 외부에 있는 신뢰할 수 있는 데이터 소스를 참조하는 [생성형 AI](#) 기술입니다. 예를 들어 RAG 모델은 조직의 지식 기반 또는 사용자 지정 데이터에 대한 의미 검색을 수행할 수 있습니다. 자세한 내용은 [RAG란 무엇입니까?](#)를 참조하십시오.

교체

공격자가 보안 인증 정보에 액세스하는 것을 더 어렵게 만들기 위해 [보안 암호](#)를 주기적으로 업데이트하는 프로세스입니다.

행 및 열 액세스 제어(RCAC)

액세스 규칙이 정의된 기본적이고 유연한 SQL 표현식을 사용합니다. RCAC는 행 권한과 열 마스크로 구성됩니다.

RPO

[복구 시점 목표를](#) 참조하십시오.

RTO

[복구 시간 목표를](#) 참조하십시오.

런북

특정 작업을 수행하는 데 필요한 일련의 수동 또는 자동 절차입니다. 일반적으로 오류율이 높은 반복 작업이나 절차를 간소화하기 위해 런북을 만듭니다.

S

SAML 2.0

많은 ID 제공업체(idP)에서 사용하는 개방형 표준입니다. 이 기능을 사용하면 연동 SSO(Single Sign-On)를 AWS Management Console 사용할 수 있으므로 사용자는 조직의 모든 사용자에게 대해 IAM에서 사용자를 생성하지 않고도 로그인하거나 AWS API 작업을 호출할 수 있습니다. SAML 2.0 기반 페더레이션에 대한 자세한 내용은 IAM 설명서의 [SAML 2.0 기반 페더레이션 정보](#)를 참조하십시오.

SCADA

[감독 제어 및 데이터 획득](#)을 참조하세요.

SCP

[서비스 제어 정책](#)을 참조하세요.

secret

에는 암호 또는 사용자 자격 증명과 같이 암호화된 형식으로 저장하는 AWS Secrets Manager 기밀 또는 제한된 정보가 있습니다. 보안 암호 값과 메타데이터로 구성됩니다. 보안 암호 값은 바이너리, 단일 문자열 또는 여러 문자열일 수 있습니다. 자세한 내용은 [Secrets Manager 설명서의 Secrets Manager 보안 암호에 무엇이 있습니까?](#)를 참조하세요.

설계를 통한 보안

전체 개발 프로세스를 통해 보안을 고려하는 시스템 엔지니어링 접근 방식입니다.

보안 제어

위협 행위자가 보안 취약성을 악용하는 능력을 방지, 탐지 또는 감소시키는 기술적 또는 관리적 가이드라인입니다. 보안 제어에는 [예방](#), [탐지](#), [대응](#) 및 [사전](#) 예방의 네 가지 주요 유형이 있습니다.

보안 강화

공격 표면을 줄여 공격에 대한 저항력을 높이는 프로세스입니다. 더 이상 필요하지 않은 리소스 제거, 최소 권한 부여의 보안 모범 사례 구현, 구성 파일의 불필요한 기능 비활성화 등의 작업이 여기에 포함될 수 있습니다.

보안 정보 및 이벤트 관리(SIEM) 시스템

보안 정보 관리(SIM)와 보안 이벤트 관리(SEM) 시스템을 결합하는 도구 및 서비스입니다. SIEM 시스템은 서버, 네트워크, 디바이스 및 기타 소스에서 데이터를 수집, 모니터링 및 분석하여 위협과 보안 침해를 탐지하고 알림을 생성합니다.

보안 응답 자동화

보안 이벤트에 자동으로 응답하거나 해결하도록 설계된 사전 정의되고 프로그래밍된 작업입니다. 이러한 자동화는 보안 모범 사례를 구현하는 데 도움이 되는 [탐지](#) 또는 [대응](#) AWS 보안 제어 역할을 합니다. 자동 응답 작업의 예로는 VPC 보안 그룹 수정, Amazon EC2 인스턴스 패치 적용 또는 자격 증명 교체 등이 있습니다.

서버 측 암호화

데이터를 AWS 서비스 수신하는가 대상에서 데이터를 암호화합니다.

서비스 제어 정책(SCP)

AWS Organizations에 속한 조직의 모든 계정에 대한 권한을 중앙 집중식으로 제어하는 정책입니다. SCP는 관리자가 사용자 또는 역할에 위임할 수 있는 작업에 대해 제한을 설정하거나 가드레일을 정의합니다. SCP를 허용 목록 또는 거부 목록으로 사용하여 허용하거나 금지할 서비스 또는 작업을 지정할 수 있습니다. 자세한 내용은 AWS Organizations 설명서의 [서비스 제어 정책을](#) 참조하세요.

서비스 엔드포인트

에 대한 진입점의 URL입니다 AWS 서비스. 엔드포인트를 사용하여 대상 서비스에 프로그래밍 방식으로 연결할 수 있습니다. 자세한 내용은 AWS 일반 참조의 [AWS 서비스 엔드포인트](#)를 참조하십시오.

서비스 수준에 관한 계약(SLA)

IT 팀이 고객에게 제공하기로 약속한 내용(예: 서비스 가동 시간 및 성능)을 명시한 계약입니다.

서비스 수준 지표(SLI)

오류율, 가용성 또는 처리량과 같은 서비스의 성능 측면에 대한 측정입니다.

서비스 수준 목표(SLO)

서비스 [수준 지표](#)로 측정되는 서비스의 상태를 나타내는 대상 지표입니다.

공동 책임 모델

클라우드 보안 및 규정 준수를 AWS 위해와 공유하는 책임을 설명하는 모델입니다. AWS 는 클라우드의 보안을 담당하는 반면, 사용자는 클라우드의 보안을 담당합니다. 자세한 내용은 [공동 책임 모델](#)을 참조하십시오.

SIEM

[보안 정보 및 이벤트 관리 시스템을 참조하세요.](#)

단일 장애 지점(SPOF)

애플리케이션의 중요한 단일 구성 요소에 장애가 발생하여 시스템이 중단될 수 있습니다.

SLA

[서비스 수준 계약을 참조하세요.](#)

SLI

[서비스 수준 표시기를 참조하세요.](#)

SLO

[서비스 수준 목표를 참조하세요.](#)

분할 앤 시드 모델

현대화 프로젝트를 확장하고 가속화하기 위한 패턴입니다. 새로운 기능과 제품 릴리스가 정의되면 핵심 팀이 분할되어 새로운 제품 팀이 만들어집니다. 이를 통해 조직의 역량과 서비스 규모를 조정하고, 개발자 생산성을 개선하고, 신속한 혁신을 지원할 수 있습니다. 자세한 내용은 [에서 애플리케이션 현대화에 대한 단계별 접근 방식을 참조하세요 AWS 클라우드.](#)

SPOF

[단일 장애 지점을 참조하세요.](#)

스타 스키마

하나의 큰 팩트 테이블을 사용하여 트랜잭션 또는 측정된 데이터를 저장하고 하나 이상의 작은 차원 테이블을 사용하여 데이터 속성을 저장하는 데이터베이스 조직 구조입니다. 이 구조는 [데이터 웨어하우스](#) 또는 비즈니스 인텔리전스용으로 설계되었습니다.

Strangler Fig 패턴

레거시 시스템을 폐기할 수 있을 때까지 시스템 기능을 점진적으로 다시 작성하고 교체하여 모놀리식 시스템을 현대화하기 위한 접근 방식. 이 패턴은 무화과 덩굴이 나무로 자라 결국 숙주를 압도

하고 대체하는 것과 비슷합니다. [Martin Fowler](#)가 모놀리식 시스템을 다시 작성할 때 위험을 관리하는 방법으로 이 패턴을 도입했습니다. 이 패턴을 적용하는 방법의 예는 [컨테이너 및 Amazon API Gateway를 사용하여 기존의 Microsoft ASP.NET\(ASMX\) 웹 서비스를 점진적으로 현대화하는 방법](#)을 참조하십시오.

서브넷

VPC의 IP 주소 범위입니다. 서브넷은 단일 가용 영역에 상주해야 합니다.

감독 제어 및 데이터 획득(SCADA)

제조에서 하드웨어와 소프트웨어를 사용하여 물리적 자산과 프로덕션 작업을 모니터링하는 시스템입니다.

대칭 암호화

동일한 키를 사용하여 데이터를 암호화하고 복호화하는 암호화 알고리즘입니다.

합성 테스트

사용자 상호 작용을 시뮬레이션하여 잠재적 문제를 감지하거나 성능을 모니터링하는 방식으로 시스템을 테스트합니다. [Amazon CloudWatch Synthetics](#)를 사용하여 이러한 테스트를 생성할 수 있습니다.

시스템 프롬프트

[LLM](#)에 컨텍스트, 지침 또는 지침을 제공하여 동작을 지시하는 기법입니다. 시스템 프롬프트는 컨텍스트를 설정하고 사용자와의 상호 작용을 위한 규칙을 설정하는 데 도움이 됩니다.

T

tags

AWS 리소스를 구성하기 위한 메타데이터 역할을 하는 키-값 페어입니다. 태그를 사용하면 리소스를 손쉽게 관리, 식별, 정리, 검색 및 필터링할 수 있습니다. 자세한 내용은 [AWS 리소스에 태그 지정](#)을 참조하십시오.

대상 변수

지도 ML에서 예측하려는 값으로, 결과 변수라고도 합니다. 예를 들어, 제조 설정에서 대상 변수는 제품 결함일 수 있습니다.

작업 목록

런북을 통해 진행 상황을 추적하는 데 사용되는 도구입니다. 작업 목록에는 런북의 개요와 완료해야 할 일반 작업 목록이 포함되어 있습니다. 각 일반 작업에 대한 예상 소요 시간, 소유자 및 진행 상황이 작업 목록에 포함됩니다.

테스트 환경

[환경을](#) 참조하세요.

훈련

ML 모델이 학습할 수 있는 데이터를 제공하는 것입니다. 훈련 데이터에는 정답이 포함되어야 합니다. 학습 알고리즘은 훈련 데이터에서 대상(예측하려는 답)에 입력 데이터 속성을 매핑하는 패턴을 찾고, 이러한 패턴을 캡처하는 ML 모델을 출력합니다. 그런 다음 ML 모델을 사용하여 대상을 모르는 새 데이터에 대한 예측을 할 수 있습니다.

전송 게이트웨이

VPC와 온프레미스 네트워크를 상호 연결하는 데 사용할 수 있는 네트워크 전송 허브입니다. 자세한 내용은 AWS Transit Gateway 설명서의 [전송 게이트웨이란 무엇입니까?](#)를 참조하세요.

트렁크 기반 워크플로

개발자가 기능 브랜치에서 로컬로 기능을 구축하고 테스트한 다음 해당 변경 사항을 기본 브랜치에 병합하는 접근 방식입니다. 이후 기본 브랜치는 개발, 프로덕션 이전 및 프로덕션 환경에 순차적으로 구축됩니다.

신뢰할 수 있는 액세스

사용자를 대신하여 AWS Organizations 및 해당 계정에서 조직에서 작업을 수행하도록 지정하는 서비스에 권한 부여. 신뢰할 수 있는 서비스는 필요할 때 각 계정에 서비스 연결 역할을 생성하여 관리 작업을 수행합니다. 자세한 내용은 설명서의 [다른 AWS 서비스와 AWS Organizations 함께 사용을](#) 참조하세요 AWS Organizations .

튜닝

ML 모델의 정확도를 높이기 위해 훈련 프로세스의 측면을 여러 변경하는 것입니다. 예를 들어, 레이블링 세트를 생성하고 레이블을 추가한 다음 다양한 설정에서 이러한 단계를 여러 번 반복하여 모델을 최적화하는 방식으로 ML 모델을 훈련할 수 있습니다.

피자 두 판 팀

피자 두 판이면 충분한 소규모 DevOps 팀. 피자 두 판 팀 규모는 소프트웨어 개발에 있어 가능한 최상의 공동 작업 기회를 보장합니다.

U

불확실성

예측 ML 모델의 신뢰성을 저해할 수 있는 부정확하거나 불완전하거나 알려지지 않은 정보를 나타내는 개념입니다. 불확실성에는 두 가지 유형이 있습니다. 인식론적 불확실성은 제한적이고 불완전한 데이터에 의해 발생하는 반면, 우연한 불확실성은 데이터에 내재된 노이즈와 무작위성에 의해 발생합니다. 자세한 내용은 [Quantifying uncertainty in deep learning systems](#) 가이드를 참조하십시오.

차별화되지 않은 작업

애플리케이션을 만들고 운영하는 데 필요하지만 최종 사용자에게 직접적인 가치를 제공하거나 경쟁 우위를 제공하지 못하는 작업을 헤비 리프팅이라고도 합니다. 차별화되지 않은 작업의 예로는 조달, 유지보수, 용량 계획 등이 있습니다.

상위 환경

[환경](#)을 참조하세요.

V

정리

스토리지를 회수하고 성능을 향상시키기 위해 증분 업데이트 후 정리 작업을 수행하는 데이터베이스 유지 관리 작업입니다.

버전 제어

리포지토리의 소스 코드 변경과 같은 변경 사항을 추적하는 프로세스 및 도구입니다.

VPC 피어링

프라이빗 IP 주소를 사용하여 트래픽을 라우팅할 수 있게 하는 두 VPC 간의 연결입니다. 자세한 내용은 Amazon VPC 설명서의 [VPC 피어링이란?](#)을 참조하십시오.

취약성

시스템 보안을 손상시키는 소프트웨어 또는 하드웨어 결함입니다.

W

웜 캐시

자주 액세스하는 최신 관련 데이터를 포함하는 버퍼 캐시입니다. 버퍼 캐시에서 데이터베이스 인스턴스를 읽을 수 있기 때문에 주 메모리나 디스크에서 읽는 것보다 빠릅니다.

웜 데이터

자주 액세스하지 않는 데이터입니다. 이런 종류의 데이터를 쿼리할 때는 일반적으로 적절히 느린 쿼리가 허용됩니다.

창 함수

현재 레코드와 어떤 식으로든 관련된 행 그룹에 대해 계산을 수행하는 SQL 함수입니다. 창 함수는 이동 평균을 계산하거나 현재 행의 상대 위치를 기반으로 행 값에 액세스하는 등의 작업을 처리하는 데 유용합니다.

워크로드

고객 대면 애플리케이션이나 백엔드 프로세스 같이 비즈니스 가치를 창출하는 리소스 및 코드 모음입니다.

워크스트림

마이그레이션 프로젝트에서 특정 작업 세트를 담당하는 직무 그룹입니다. 각 워크스트림은 독립적이지만 프로젝트의 다른 워크스트림을 지원합니다. 예를 들어, 포트폴리오 워크스트림은 애플리케이션 우선순위 지정, 웨이브 계획, 마이그레이션 메타데이터 수집을 담당합니다. 포트폴리오 워크스트림은 이러한 자산을 마이그레이션 워크스트림에 전달하고, 마이그레이션 워크스트림은 서버와 애플리케이션을 마이그레이션합니다.

WORM

[쓰기를 한 번, 많이 읽기를 참조하세요.](#)

WQF

[AWS 워크로드 검증 프레임워크](#)를 참조하세요.

한 번 쓰기, 많이 읽기(WORM)

데이터를 한 번 쓰고 데이터가 삭제되거나 수정되지 않도록 하는 스토리지 모델입니다. 권한 있는 사용자는 필요한 만큼 데이터를 읽을 수 있지만 변경할 수는 없습니다. 이 데이터 스토리지 인프라는 [변경할 수 없는](#) 것으로 간주됩니다.

Z

제로데이 익스플로잇

[제로데이 취약성](#)을 활용하는 공격, 일반적으로 맬웨어입니다.

제로데이 취약성

프로덕션 시스템의 명백한 결함 또는 취약성입니다. 위협 행위자는 이러한 유형의 취약성을 사용하여 시스템을 공격할 수 있습니다. 개발자는 공격의 결과로 취약성을 인지하는 경우가 많습니다.

제로샷 프롬프트

[LLM](#)에 작업을 수행하기 위한 지침을 제공하지만 작업에 도움이 될 수 있는 예제(샷)는 제공하지 않습니다. LLM은 사전 훈련된 지식을 사용하여 작업을 처리해야 합니다. 제로샷 프롬프트의 효과는 작업의 복잡성과 프롬프트의 품질에 따라 달라집니다. [스크린샷이 거의 없는 프롬프트도 참조하세요.](#)

좀비 애플리케이션

평균 CPU 및 메모리 사용량이 5% 미만인 애플리케이션입니다. 마이그레이션 프로젝트에서는 이러한 애플리케이션을 사용 중지하는 것이 일반적입니다.

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.