



스타트업을 위한 전체 기계 학습 릴리스 가이드

프레젠테이션 대상:

개발자



기계 학습을 앱
및 서비스에 통합

데이터
사이언티스트



생산성 향상

기계 학습
전문가



최첨단 정보 얻기



AWS가 기계 학습 스타트업의 성장을 주도하는 방법



활성화

- 무료 기계 학습 교육
- 무료 기계 학습 워크숍
- 크레딧
- 근무 시간
- 전담 팀
- 참조 아키텍처



구축

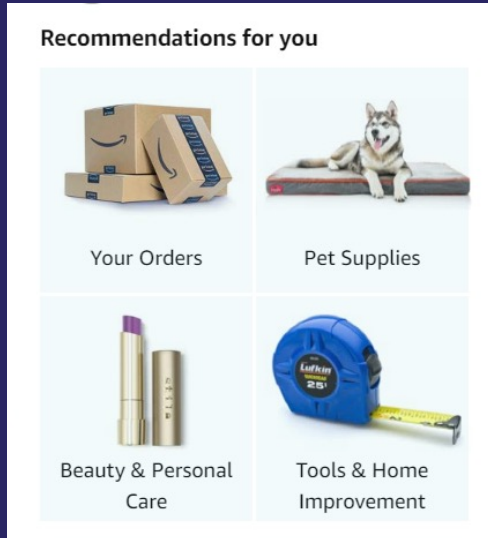
- 가장 광범위하고 세분화된 제품 세트
- 기술 지원
- 베타 제품 및 로드맵 액세스
- 비용 최적화 검토
- 상담 참여
- 공동 마케팅



연결

- 자금 조달(VC)
- 엔터프라이즈 고객
- 파트너(APN)
- AWS Marketplace
- 심층적인 산업 전문 인력

Amazon의 대규모 기계 학습 혁신



Amazon.com에서 분당
4,000개의 제품이 판매됨



매일
160만 개 패키지



매주 수십억 건의 Alexa
상호 작용



2016년 12월 7일
첫 Prime Air 배송

Amazon의 대규모 기계 학습 혁신

- Amazon은 오늘날과 같은 기계 학습 중심의 회사가 되기 위해 다년간의 기계 학습 여정을 거쳤습니다. 20년 넘게 개인화 및 공급망과 같은 분야에서 기계 학습을 적용해 왔으며, 원래의 개인화 모델을 지속적으로 대폭 개선하여 Amazon Prime과 같은 다른 제품으로 확대했습니다.
- Amazon은 주문 처리 과정 전반에서 기계 학습을 사용합니다. 예를 들어 전 세계에 판매되는 각 제품의 적정 수요를 예측할 수 있는 시스템을 개발하여 편의성, 비용 및 배송 속도에 대한 고객의 기대치를 충족합니다.
- Amazon은 Alexa를 통해 고객과 기술의 상호 작용 방식을 완전히 바꾸는 자연어 처리 기술을 개발했습니다. 또한 주문 처리 센터에서 Prime Air 드론 및 로봇을 통한 자율 비행과 같은 획기적인 기술을 개발하여 고객에게 더 빠르게 패키지를 배송했습니다. 이는 커다란 문화적 변화였지만 모든 조직이 달성할 수 있는 변화이기도 합니다.

스타트업이 기계 학습을 위해 **AWS**를 선택한 이유



스타트업이 기계 학습을 위해 **AWS**를 선택한 이유



가장 광범위하고 세분화된 AI 및 기계 학습 서비스 세트

2020년 한 해에만 200개 이상의
새로운 기능 및 서비스 출시

모두를 위한 솔루션

3가지 주요 기계 학습 프레임워크를
모두 지원



SageMaker를 활용하여 기계 학습 도입 가속화

전체 기계 학습 워크플로를 위한
단일 IDE

54% 이상의 TCO 절감

데이터 레이블 지정 비용 최대
70% 절감

Managed Spot Training을 통해
최대 90%의 비용 절감



가장 포괄적인 클라우드 플랫폼에 구축됨

탁월한 보안, 안정성 및 완벽한
기능을 갖춘 데이터 스토어

가장 강력한 컴퓨팅, 스토리지,
보안, 데이터베이스 및 분석
기능 세트 제공

AWS Machine Learning을 기반으로 하는 수천 개의 스타트업

coinbase



Aurora

common
edits

STUDIO71

stripe



CONVOY

Root
Insurance Co

DELIVERY



TransferWise

affirm



DEEPMAP

cravelabs

TECTON



ABACUS.AI



AWS를 통한 AI 및 기계 학습

혁신, 선택 및 유연성



100,000+

AWS 기반 기계 학습을 사용하고 있는 고객 수

250+

지난 12개월 동안 기계 학습 및 인공지능(AI)용으로 개발된 새로운 기능

92% AWS에서 실행되는 클라우드 기반 딥 러닝(DL) 비율

91% AWS에서 실행되는 클라우드 기반 PyTorch 비율

AWS Machine Learning 솔루션

훈련 시간 50% 단축
90%의 확장 효율성 제공
3배 더 빠른 네트워크 처리량 제공
가격 대비 성능 25% 개선

AWS를 통한 AI 및 기계 학습

혁신, 선택 및 유연성

- AWS는 100,000개 이상의 고객사가 선택한 공급 업체이자 AI 및 기계 학습 훈련과 추론을 위한 가장 광범위하고 세분화된 컴퓨팅 인프라를 제공하는 선두 업체입니다.
- 최신 CPU, GPU 및 사용자 지정 액셀러레이터와 업계 최고의 네트워킹 및 스토리지를 기반으로 하는 다양한 Amazon EC2 인스턴스 중에서 예산 요구 사항과 모델의 성능 요구 사항에 적합한 인스턴스를 선택할 수 있습니다.
- AWS는 고객이 요구 사항에 가장 적합한 옵션을 선택할 수 있도록 다양한 고성능, 저비용 인프라 옵션을 제공하기 위해 투자 및 혁신하고 있습니다.
- 예를 들어 AWS는 기계 학습 추론 성능에 최적화된 실리콘을 기초부터 구축하고 있습니다. Graviton CPU 및 Nitro 시스템 혁신에서 얻은 교훈을 사용하여 클라우드에서 가장 저렴한 추론 비용을 제공하는 맞춤형 칩인 AWS Inferentia 칩을 구축했습니다.
- AWS에서 기계 학습 워크로드를 실행하면 기계 학습 및 DL 모델을 훈련하고 배포하기 위해 사용이 간편한 고성능, 저비용 인프라 서비스에 온디맨드 방식으로 액세스할 수 있습니다.

스타트업의 일반적인 과제



개발자

데이터 사일로, 준비 및 정리, 전체 데이터 워크플로



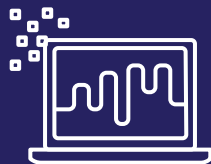
기술 역량 격차

인력 및 경험 부족



비즈니스 사례 사용

기계 학습을 활용할 수 있는 올바른 비즈니스 사용 사례 찾기



기계 학습 운영

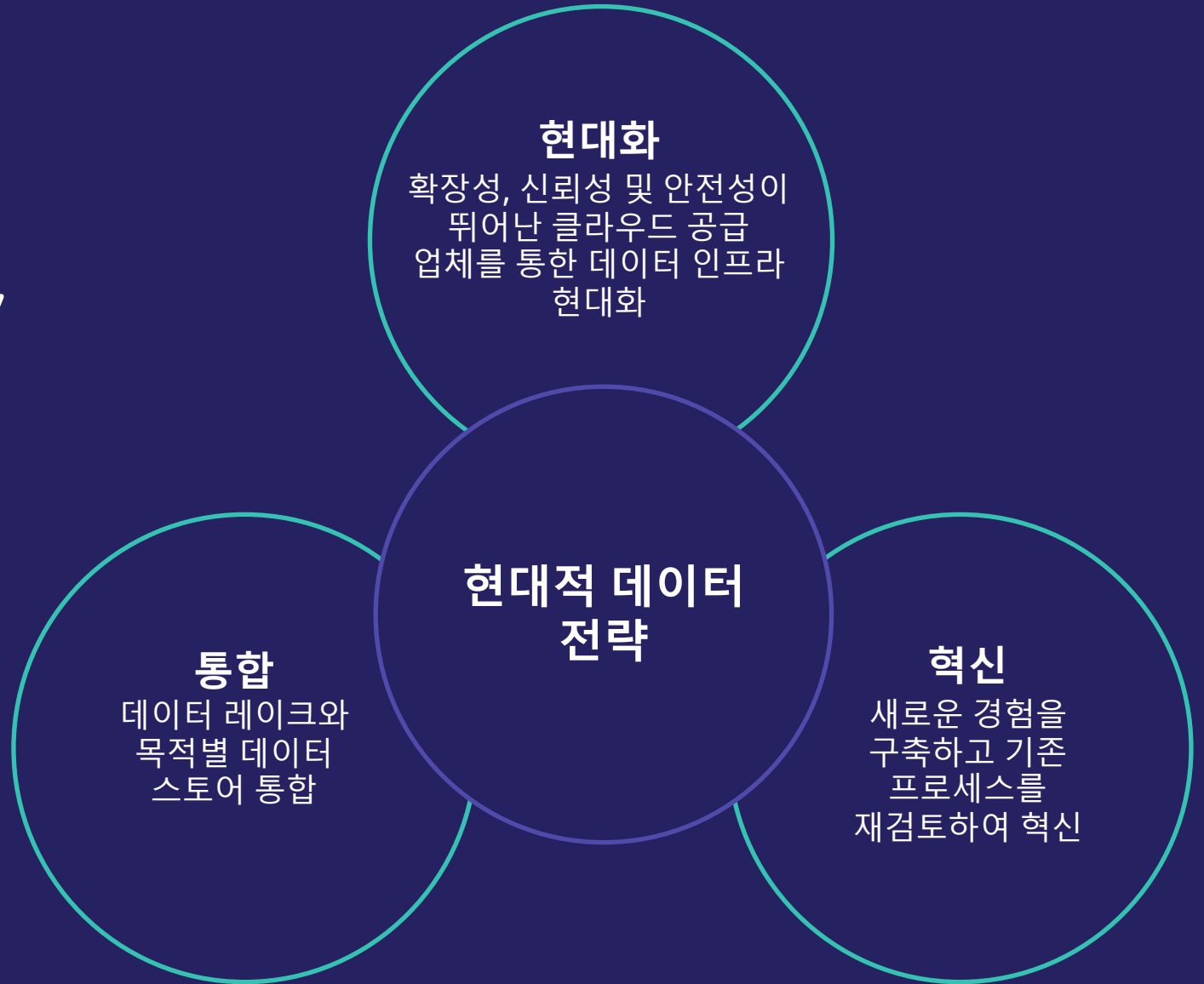
시간이 많이 걸리고 복잡한 기계 학습 워크플로 생성 및 관리

AWS를 통한 현대적 데이터 전략



AWS를 통한 현대적 데이터 전략

현대적 데이터 전략 수립 방법 현대화,
통합 및 혁신



AWS를 통한 현대적 데이터 전략

디지털 트랜스포메이션의 새로운 물결인 데이터 활용

1

AWS는 스타트업이 다음을 수행할 수 있도록 지원합니다.

확장성, 신뢰성, 안전성이 뛰어난 클라우드 공급 업체를 통해 데이터 인프라를 현대화합니다. 레거시 온프레미스 데이터 스토어를 실행하거나 클라우드에서 자체 관리하는 조직은 여전히 데이터베이스 프로비저닝, 패치 적용, 구성 또는 백업과 같은 관리 작업을 수행해야 합니다. AWS가 이러한 문제를 처리하는 데 도움을 드릴 수 있습니다. 가장 중요한 애플리케이션에 AWS의 탁월한 경험, 성숙도, 안정성, 보안 및 성능을 활용해 보세요.

AWS를 통한 현대적 데이터 전략

디지털 트랜스포메이션의 새로운 물결인 데이터 활용

2

AWS는 스타트업이 다음을 수행할 수 있도록 지원합니다.

통합: 안전하고 잘 통제된 데이터 액세스를 통해 데이터 작업을 수행합니다. 신속하게 결정을 내리려면 비즈니스 요구 사항의 변화에 따라 크기를 조정하고 확장할 수 있는 새로운 데이터 스토어가 필요합니다. 또한 데이터 레이크, 데이터 웨어하우스, 목적별 데이터 스토어를 비롯한 모든 저장소를 안전하고 잘 관리되는 일관된 시스템으로 연결해야 합니다. AWS는 데이터 레이크와 목적별 데이터 스토어를 통해 이러한 작업을 수행할 수 있도록 지원합니다.

AWS를 통한 현대적 데이터 전략

디지털 트랜스포메이션의 새로운 물결인 데이터 활용

3

AWS는 스타트업이 다음을 수행할 수 있도록 지원합니다.

AI 및 기계 학습을 통해 새로운 경험을 창조하고 기존 프로세스를 혁신합니다. 기계 학습은 지난 25년 동안 가장 혁신적인 기술 중 하나입니다. 이를 통해 완전히 새로운 수익 기회를 창출하고, 더 나은 의사 결정을 더 빠르게 내리며, 운영 효율성을 개선할 수 있습니다. AWS는 모든 수준의 전문 지식을 갖춘 빌더를 위한 가장 광범위하고 포괄적인 기계 학습 및 AI 서비스 세트를 제공하여 여정 중 어느 단계에 있든 고객을 지원합니다.



기계 학습 여정 시작



기계 학습 여정 시작



데이터 전략 수립

변화와 혁신의
기반 마련



비즈니스 과제 시작

비즈니스 요구 사항에 따라
올바른 사용 사례 찾기



성공을 위해 협업

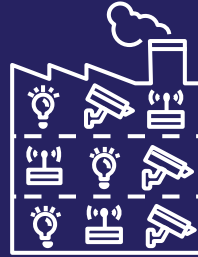
다양한 프로그램과 훈련 옵션을
활용하여 작업 진행 지원

기계 학습 여정 시작



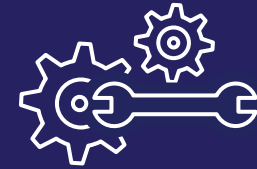
수평적 사용 사례 해결

개발자가 모든
애플리케이션에
인텔리전스를 쉽게
추가할 수 있도록 지원



분야별 솔루션 찾기

요구 사항에 더 적합한 산업별
서비스 검색



자체 모델 구축

Amazon SageMaker를 사용하여
기계 학습을 더 빠르고 쉽게 수행

시작하기: 일반적인 기계 학습 사용 사례

기계 학습을 통해 실제 문제 해결

고객 경험 개선

비즈니스 최적화

혁신 가속화



개인화



고객 센터
인텔리전스



미디어
인텔리전스



지능형 검색



지능형 문서
처리



사기 탐지



비즈니스
지표 분석



기계 학습
현대화



차세대
DevOps

운영 비용을 절감하면서 고객 만족

생산성 향상 및 비즈니스 프로세스 최적화

기계 학습을 통해 혁신 가속화 및 확장

시작하기: 일반적인 기계 학습 사용 사례

기계 학습을 통해 실제 문제 해결

- 기계 학습은 다양한 비즈니스 부문 및 산업 분야 전반에 걸쳐 많은 일반적인 과제에 적용될 수 있습니다. 고객 경험 개선에서 기술 팀의 생산성 향상에 이르기까지, 이러한 기계 학습 사용 사례는 복잡하고 비용이 많이 들며 인적 오버헤드가 필요한 작업을 최적화하는 데 도움이 됩니다. AWS는 사용이 간편한 여러 AI 서비스를 제공하여 이러한 일반적인 사용 사례를 처리할 수 있도록 지원합니다. 사전 기계 학습 경험은 필요하지 않습니다. 클라우드에서 기계 학습 워크플로를 실행하고 자체 기계 학습 모델 및 알고리즘을 사용하려는 경우 데이터 사이언티스트와 기계 학습 개발자가 기계 학습 모델을 신속하게 구축, 훈련 및 배포할 수 있도록 지원하는 완전관리형 서비스인 Amazon SageMaker를 사용할 수 있습니다.

AWS Machine Learning 스택

가장 방대하고 완전한 기계 학습 기능 세트

AI 서비스



기계 학습 서비스



기계 학습 프레임워크 및 인프라

PyTorch, Apache MXNet, TensorFlow

Amazon EC2

CPU

GPU

AWS Inferentia

AWS Trainium

Habana Gaudi

FPGA

Elastic Inference



AI 서비스: 애플리케이션에 인텔리전스를 손쉽게 추가

기계 학습 기술 역량 필요 없음



비전



챗봇



비즈니스 도구



검색



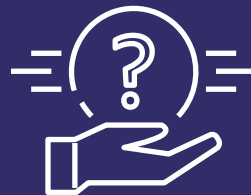
의료



음성



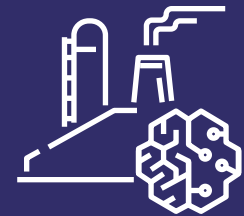
텍스트



고객 센터



코드 및 DevOps



산업

AI 서비스: 애플리케이션에 인텔리전스를 손쉽게 추가

기계 학습 기술 역량 필요 없음

- AWS의 사전에 훈련된 AI 서비스는 애플리케이션 및 워크플로에 바로 사용할 수 있는 인텔리전스를 제공합니다. AI 서비스는 애플리케이션을 통합하여 개인 맞춤 추천, 고객 센터 현대화, 안전 및 보안 개선, 고객 참여 개선과 같은 일반적인 사용 사례를 다룹니다.
- Amazon.com 및 Amazon Machine Learning 서비스를 구동하는 것과 동일한 딥 러닝 기술을 사용하므로 API를 지속적으로 훈련하여 품질과 정확성을 얻을 수 있습니다. 무엇보다도 AWS의 AI 서비스는 기계 학습 경험이 없어도 시작할 수 있습니다.

Amazon Fraud Detector

신속한 사기 탐지



기계 학습으로 사기 탐지 개선



기계 학습 전문 지식 수준에 관계없이 기계 학습 사기 탐지 모델 구축 가능



Amazon의 경험 및 풍부한 기능을 통한 기계 학습 향상



오탐지 및 수작업 검토 감소



사기 방지 직원 셀프 서비스를 통해 신속하게 위협 요소 해결



TCO 절감 및 TTV 가속화

Amazon Fraud Detector

신속한 사기 탐지

고객은 Amazon Fraud Detector를 사용할 경우 다음을 포함한 다양한 이점을 얻을 수 있다고 보고하고 있습니다.

- **기계 학습을 통한 사기 탐지 작업 개선.** Amazon Fraud Detector를 사용하여 기존의 규칙 기반 사기 탐지 솔루션을 보강하거나 교체합니다.
- **기계 학습 전문 지식의 수준에 관계없이 모든 사용자가 기계 학습 모델 구축 가능.** Amazon Fraud Detector를 사용하면 기계 학습 전문가가 아닌 사기 방지 담당 직원이 기계 학습 기반 사기 탐지 모델을 구축하고 사용할 수 있습니다. 결과적으로 사기 탐지 작업에 전담 데이터 사이언티스트가 관여할 필요가 없거나 다른 조직의 요구 사항을 해결하도록 용도를 변경할 수 있습니다.
- **Amazon의 풍부한 기능 및 사기 탐지 경험으로 기계 학습 모델의 결과 개선.** 대부분의 고객은 Amazon Fraud Detector로 생성된 모델이 자신이 직접 구축한 모델보다 낫다고 보고했습니다.
- **오탐지 알림 및 전반적인 수작업 검토 처리 감소.** Amazon Fraud Detector로 모델을 생성하면 고객 서비스 및 사기 방지 팀에서 처리해야 하는 작업량이 크게 줄어듭니다.
- **사기 방지 팀이 위협을 해결하기 위한 새로운 조치를 구현할 수 있도록 허용.** 사기 방지 팀은 Amazon Fraud Detector를 사용하여 새로운 공격을 처리하고 IT와의 협력이 필요 없는 새로운 모델 로직을 즉시 정의 및 배포합니다.
- **총 소유 비용(TOC) 절감 및 가치 창출 시간(TTV) 단축.** AWS 콘솔에 로그인하고 선불 비용이나 약정 없이 또는 소프트웨어를 설치, 구성 또는 유지 관리할 필요 없이 Amazon Fraud Detector를 사용하여 사기 탐지 솔루션 개발 및 배포를 즉시 시작합니다.

Amazon Personalize

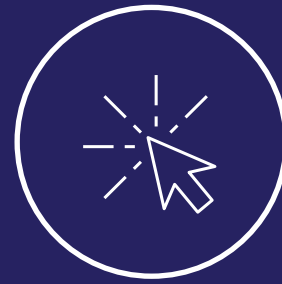
고객 만족 및 고객 경험 개선



고품질
추천 제공



실시간으로 고객의
의도 변화 반영



몇 번의 클릭으로 추천
모델 훈련



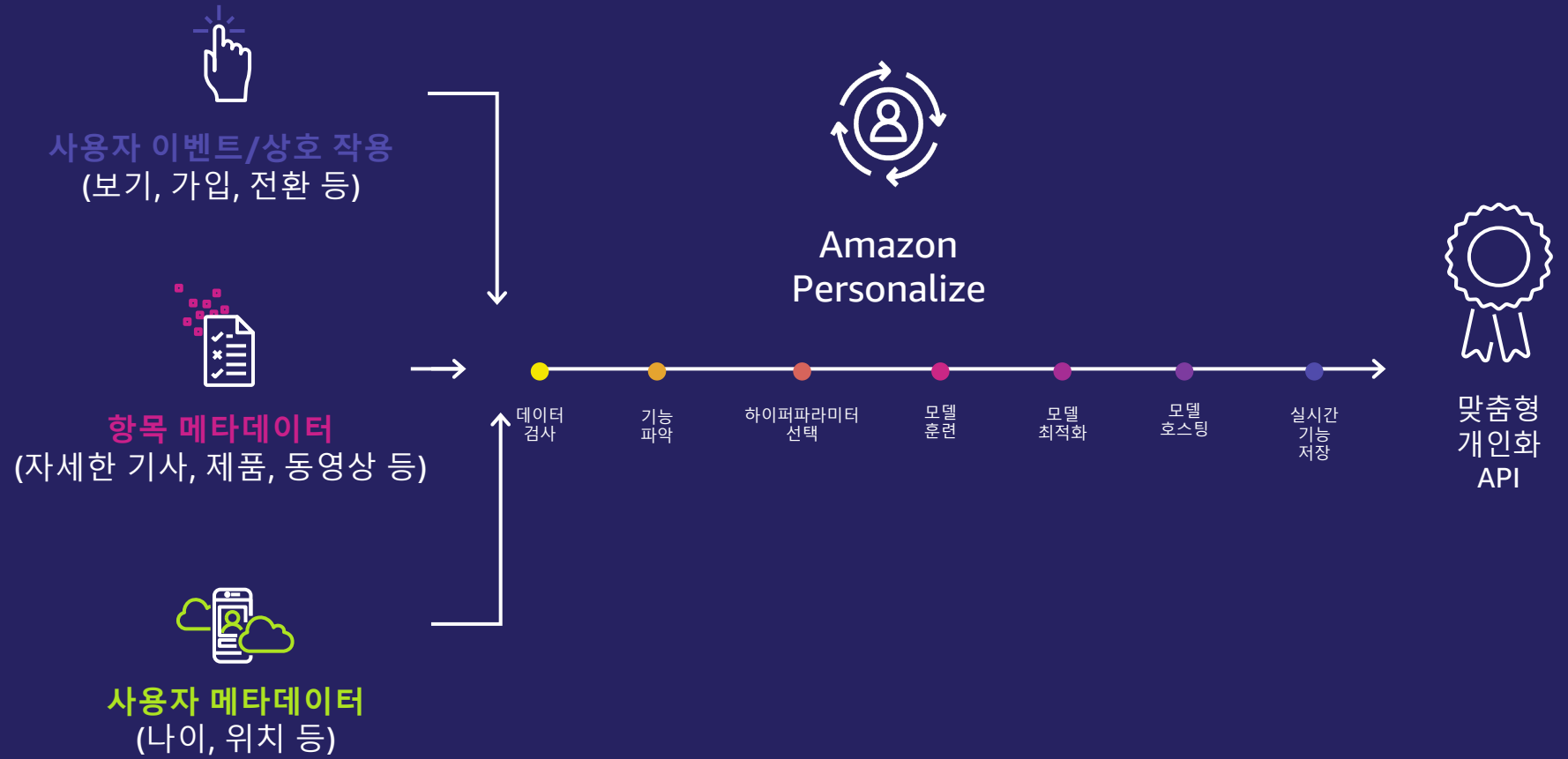
거의 모든 제품 또는
콘텐츠에 대한 추천 생성

Amazon Personalize는 기본 사용 사례에 즉시 사용할 수 있는 솔루션을 제공합니다.
이러한 기능을 사용하여 다양한 맞춤형 사용자 경험을 생성할 수 있습니다.

Amazon Personalize

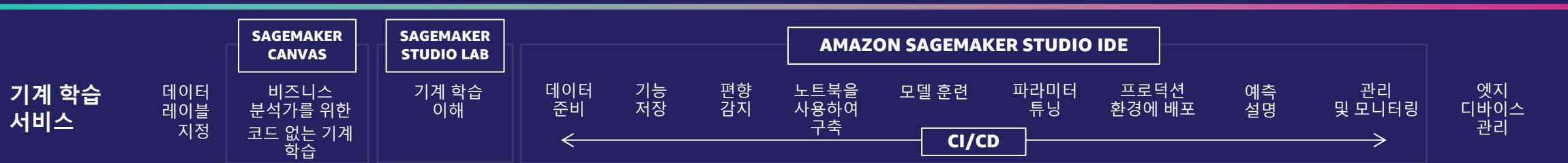
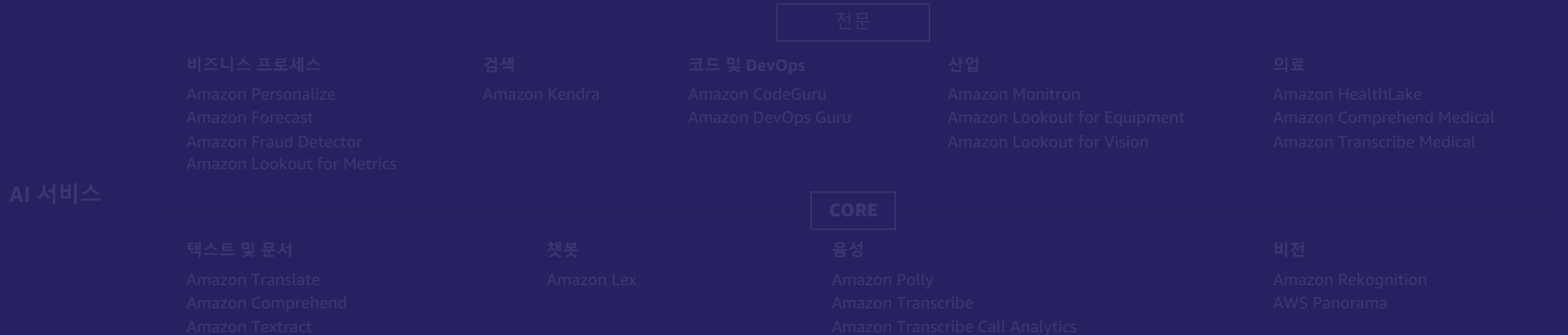
작동 방식

1. 데이터 처리 및 검사
2. 의미 있는 내용 식별
3. 올바른 알고리즘 선택
4. 데이터에 맞게 맞춤화된 모델 훈련 및 최적화



AWS Machine Learning 스택

가장 방대하고 완전한 기계 학습 기능 세트



Amazon SageMaker



통합 워크벤치

기계 학습, 데이터 준비, 실험 관리 및 파이프라인을 위해 특별히 설계된 IDE

관리형 인프라

매우 짧은 대기 시간 및 높은 처리량, 자동 크기 조정 및 분산 훈련을 위해 설계됨

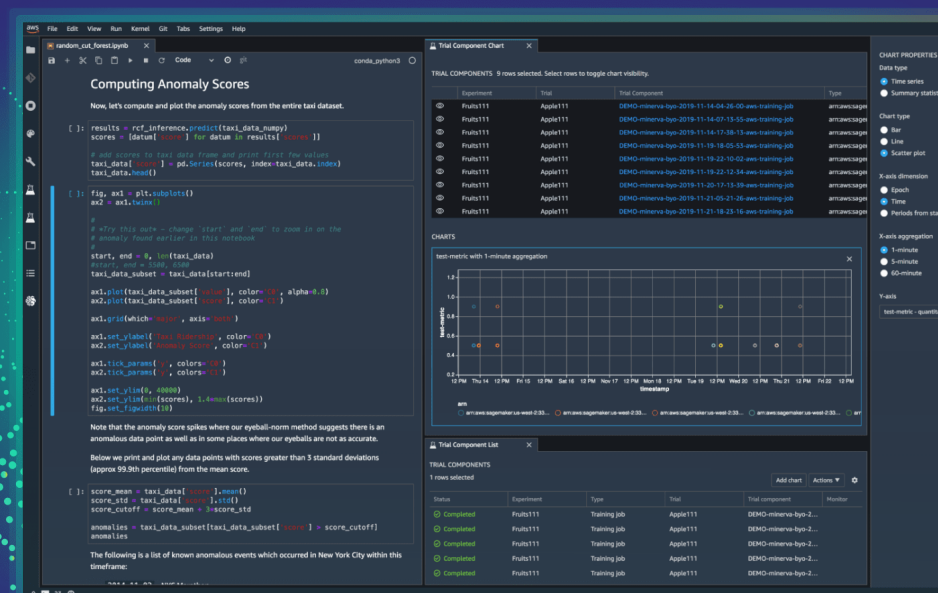
관리형 도구

처음부터 함께 작동하도록 특별히 제작됨:
SageMaker Autopilot, 협업, Jupyter 노트북, Experiments, Debugger, Model Monitor 등

<https://aws.amazon.com/sagemaker>

Amazon SageMaker

가장 완벽한 엔드 투 엔드 기계 학습 서비스



Amazon SageMaker

가장 완벽한 엔드 투 엔드 기계 학습 서비스

- Amazon SageMaker는 기계 학습을 위한 가장 완벽한 엔드 투 엔드 서비스입니다. 데이터 사이언티스트 및 기계 학습 운영 팀을 위한 관리형 서비스로, 기계 학습과 관련된 획일적인 업무 부담을 제거하여 더 많은 시간, 리소스 및 에너지를 비즈니스에 집중할 수 있도록 지원합니다. SageMaker는 많은 기능과 완전한 기능 워크벤치를 갖추고 있으며 SageMaker가 무엇이며 어떻게 사용되는지 설명하는 3가지 주요 요소가 있습니다.
- 먼저 SageMaker Studio는 통합된 도구 워크벤치를 제공합니다. 예를 들어 SageMaker Studio를 통해 Jupyter 노트북 및 JupyterLab 환경을 즉시 시작할 수 있습니다. SageMaker는 또한 데이터 사이언티스트의 생산성을 높일 수 있도록 완벽한 실험 관리, 데이터 준비, 파이프라인 자동화 및 오케스트레이션 기능을 제공합니다.

Amazon SageMaker

가장 완벽한 엔드 투 엔드 기계 학습 서비스

- Jupyter 노트북을 사용한 경험이 있다면 컴퓨팅 환경에서 실행해야 한다는 것을 알고 있을 것입니다. SageMaker는 데이터 사이언티스트와 개발자가 이를 쉽게 수행할 수 있도록 클라우드에서 완전관리형 서버를 제공합니다. 그러나 노트북 외에도 SageMaker는 다른 관리형 인프라 기능도 제공합니다. SageMaker는 분산 훈련 작업, 데이터 처리 작업 및 모델 호스팅에서 모델 구축, 훈련 및 호스팅과 관련된 확장, 패치 적용, 고가용성 등을 모두 처리합니다.
- 마지막으로, 관리형 인프라에 있는 이 통합 워크벤치는 데이터 사이언티스트의 생산성을 높이기 위해 기계 학습용으로 특별히 제작되고 처음부터 함께 작동하도록 설계된 거대한 도구 에코시스템으로 강화되었습니다.

Amazon SageMaker: 기계 학습에 더 쉽게 액세스할 수 있도록 구축됨

데이터 레이블 지정



데이터 수집 및 준비



기능 저장



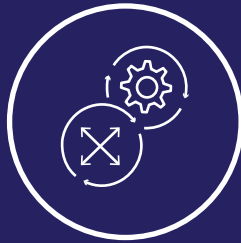
데이터 검사



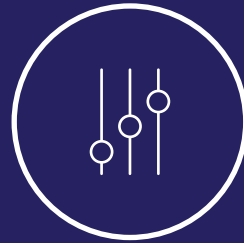
노트북에서 시각화



알고리즘 선택



모델 훈련



파라미터 튜닝



프로덕션 환경에 배포



관리 및 모니터링



CI/CD

SageMaker Studio IDE

Amazon SageMaker:

기계 학습에 더 쉽게 액세스할 수 있도록 구축됨

- Amazon SageMaker는 가장 완벽한 엔드 투 엔드 기계 학습 서비스로 민첩성, 생산성 및 비용 효율성을 개선하도록 지원합니다.
- AWS는 모든 개발자와 데이터 사이언티스트가 기계 학습 모델을 더 저렴한 비용으로 신속하게 구축, 훈련 및 배포할 수 있도록 SageMaker를 기초부터 구축했습니다. 이를 위해 기계 학습 개발 수명 주기의 모든 단계에 필요한 도구를 하나로 통합된 완전관리형 서비스로 제공합니다. 실제로 고객이 이 프로세스를 더 쉽게 수행할 수 있도록 하기 위해 지난 해에만 50개 이상의 기능을 출시했습니다.
- 그리고 작년에는 한 곳에서 모든 도구에 액세스할 수 있는 Amazon SageMaker Studio를 출시했습니다.

Amazon SageMaker 개요

Amazon SageMaker

준비

SageMaker Ground Truth

기계 학습용 훈련 데이터 레이블 지정

SageMaker Data Wrangler

기계 학습용 데이터 집계 및 준비

SageMaker Processing

기본 제공된 Python 또는 기존 보유(BYO) R/Spark 사용

SageMaker Feature Store

기능 저장, 업데이트, 가져오기 및 공유

구축

SageMaker Studio 노트북

탄력적 컴퓨팅 및 공유 기능이 있는 Jupyter 노트북 사용

기본 제공 및 기존 보유 알고리즘

수십 개의 최적화된 알고리즘 또는 기존 보유 알고리즘 사용

로컬 모드

로컬 시스템에서 테스트 및 프로토타입 생성

SageMaker Autopilot

완벽한 가시성이 보장되는 기계 학습 모델 자동 생성

훈련 및 튜닝

원클릭 훈련

분산 인프라 관리

SageMaker Experiments

모든 단계 캡처, 구성 및 비교

자동 모델 튜닝

하이퍼파라미터 최적화

SageMaker Debugger

훈련 실행 디버깅

Managed Spot Training

훈련 비용 90% 절감

배포 및 관리

원클릭 배포

완전관리형, 매우 짧은 대기 시간, 높은 처리량

Kubernetes 및 Kubeflow 통합

Kubernetes 기반 기계 학습 간소화

다중 모델 엔드포인트

인스턴스당 여러 모델을 호스팅하여 비용 절감

SageMaker Model Monitor

배포한 모델의 정확성 유지

SageMaker Pipelines

워크플로 오케스트레이션 및 자동화 구현

SageMaker Studio

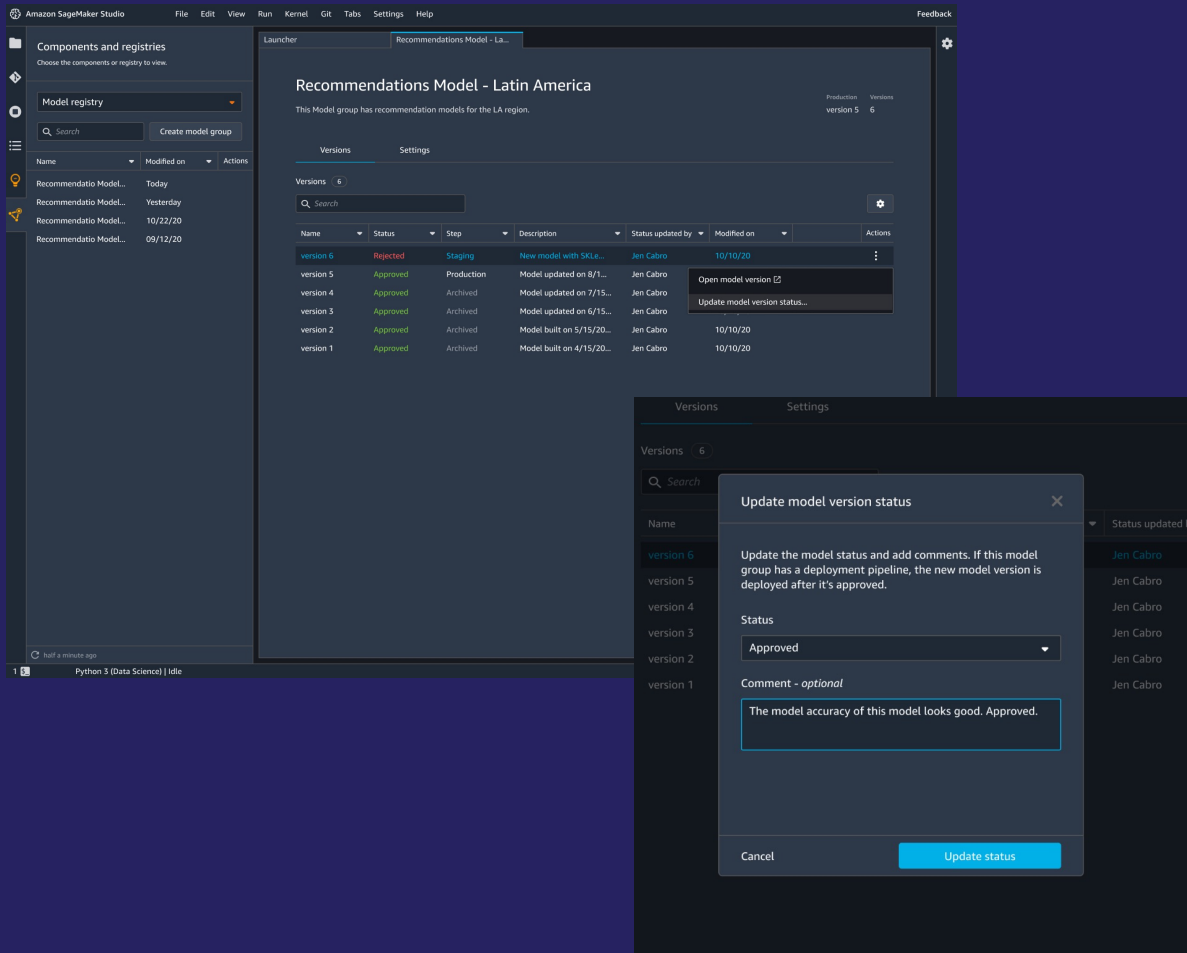
기계 학습용 통합 개발 환경(IDE)

조직의 기계 학습 활용을 지원하는 **Amazon SageMaker**

- Amazon SageMaker는 비즈니스 분석가, 데이터 사이언티스트 및 MLOps 엔지니어가 기계 학습 전문 지식에 관계없이 모든 사용 사례에 대해 기계 학습 모델을 구축, 훈련 및 배포할 수 있도록 지원하는 포괄적인 기계 학습 서비스입니다.
- 비즈니스 분석가는 SageMaker Canvas와 시각적 인터페이스를 사용하여 기계 학습 예측을 수행할 수 있습니다.
- 데이터 사이언티스트는 SageMaker Studio를 통해 데이터를 쉽게 준비하고 기계 학습 모델을 구축, 훈련 및 배포할 수 있습니다.
- MLOps 엔지니어는 SageMaker MLOps를 사용하여 대규모로 모델을 배포하고 관리할 수 있습니다.

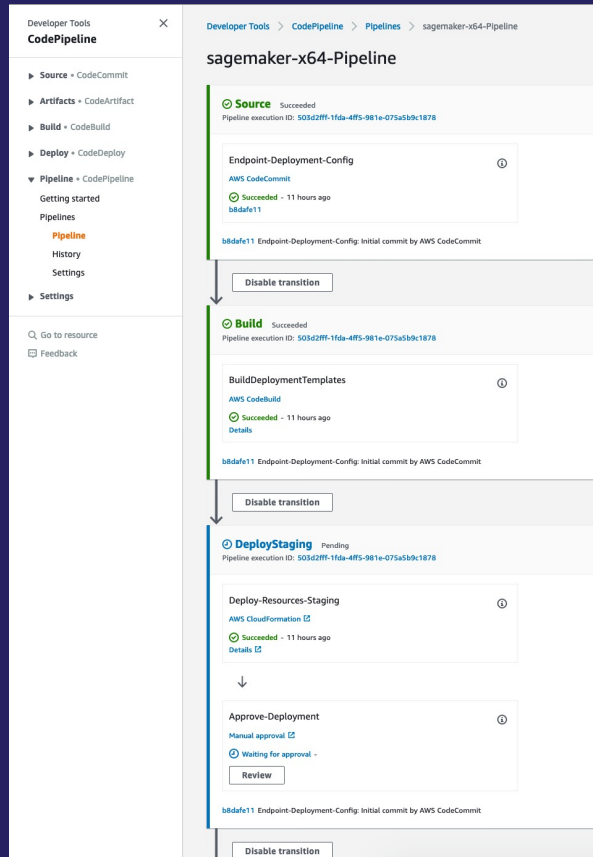


프로덕션 환경을 위해 모델 승인



데이터 사이언티스트에게 제공되면 데이터 사이언티스트가 모델 레지스트리 내에서 모델을 '승인'하여 CI/CD 파이프라인을 통해 모델을 이동합니다.

완전관리형 CI/CD 파이프라인을 사용하여 배포

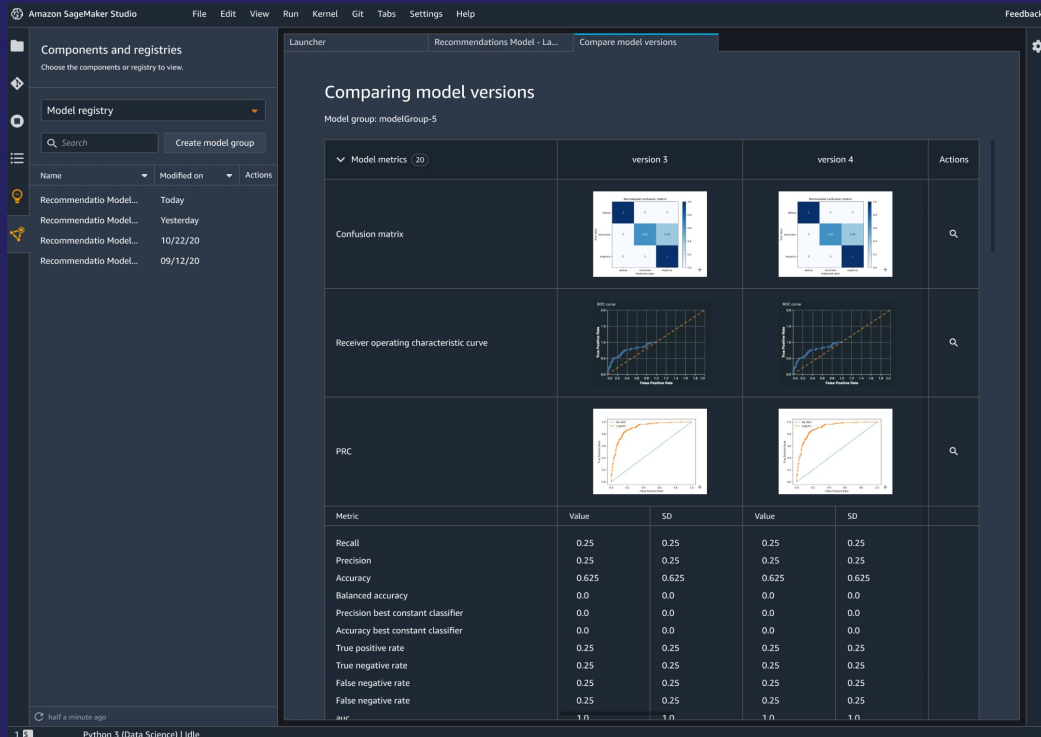


이제 모델이 훈련된 이후와 모델 레지스트리에 추가된 이후 어떤 일이 발생하는지 이야기해 보겠습니다. SageMaker Pipelines를 사용하여 CI/CD 워크플로를 시작하려면 어떻게 해야 할까요?

먼저 CI/CD 워크플로가 생성되는 방법에 대해 이야기하겠습니다. 이를 위해 DevOps 엔지니어라는 새로운 페르소나에 대해 알아보겠습니다.

이 사용 사례에서 DevOps 엔지니어는 실제로 AWS CodePipelines와 같이 자신에게 익숙한 AWS 도구를 사용합니다. 또한 CI/CD 파이프라인을 정의하여 다양한 환경 및 통합 테스트를 통해 모델 아티팩트를 이동하는 방법을 체계화할 수 있습니다.

훈련 단계에서 평가 지표 보기 및 비교



이 모든 통합의 가장 큰 장점은 SageMaker Studio가 제공하는 단일 창을 활용할 수 있다는 것입니다. 따라서 모델이 프로덕션 환경에 배포된 후 Studio 내에서 해당 모델에 대한 결과를 바로 볼 수 있습니다.

Amazon SageMaker 주요 이점

가장 완벽한 엔드 투 엔드 기계 학습 서비스



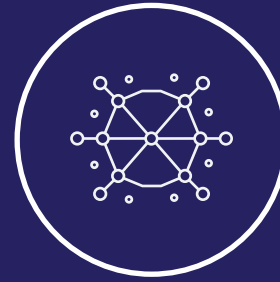
기계 학습 혁신의 공용화

비즈니스 분석가를
포함하여 더 많은
사람들에게 권한 부여



기계 학습 수명 주기 가속화

훈련 시간을 몇
시간에서 몇 분으로
단축



대규모 데이터 준비

정형 및 비정형 데이터
액세스, 레이블 지정 및
처리



기계 학습 프로세스 간소화

MLOps 프로세스 자동화
및 표준화

Amazon SageMaker 주요 이점

가장 완벽한 엔드 투 엔드 기계 학습 서비스

1. 비즈니스 분석가를 포함하여 더 많은 사람들이 코드 없이 포인트 앤 클릭 방식의 시각적 인터페이스를 사용하여 직접 기계 학습 모델을 생성할 수 있습니다.
2. 기계 학습을 위한 방대한 양의 정형 데이터(예: 표 형식 데이터) 및 비정형 데이터(예: 사진, 비디오 및 오디오)에 액세스하고 레이블을 지정하며 처리할 수 있습니다.
3. 최적화된 인프라를 통해 훈련 시간을 몇 시간에서 몇 분으로 단축합니다. 목적별 도구를 사용하여 팀 생산성을 최대 10배까지 높입니다.
4. 조직 전체에서 MLOps 프로세스를 자동화 및 표준화하여 프로덕션 환경에서 수만 개의 모델을 배포하고 관리합니다.



Amazon SageMaker를 DevOps에 사용 가능



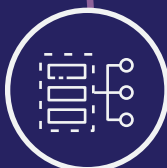
보안

기계 학습 워크로드의 엄격한 보안 요구 사항을 충족할 수 있도록 지원하는 보안 기능



규정 준수

PCI, HIPAA, SOC 1/2/3, FedRAMP
및 ISO 9001/27001/27017/27018 준수



기계 학습 워크플로

자동화된 워크플로를 몇 분 만에 생성하여
수천 개의 모델 지원



확장성

방대한 데이터 집합으로 복잡한 모델 훈련



오케스트레이션

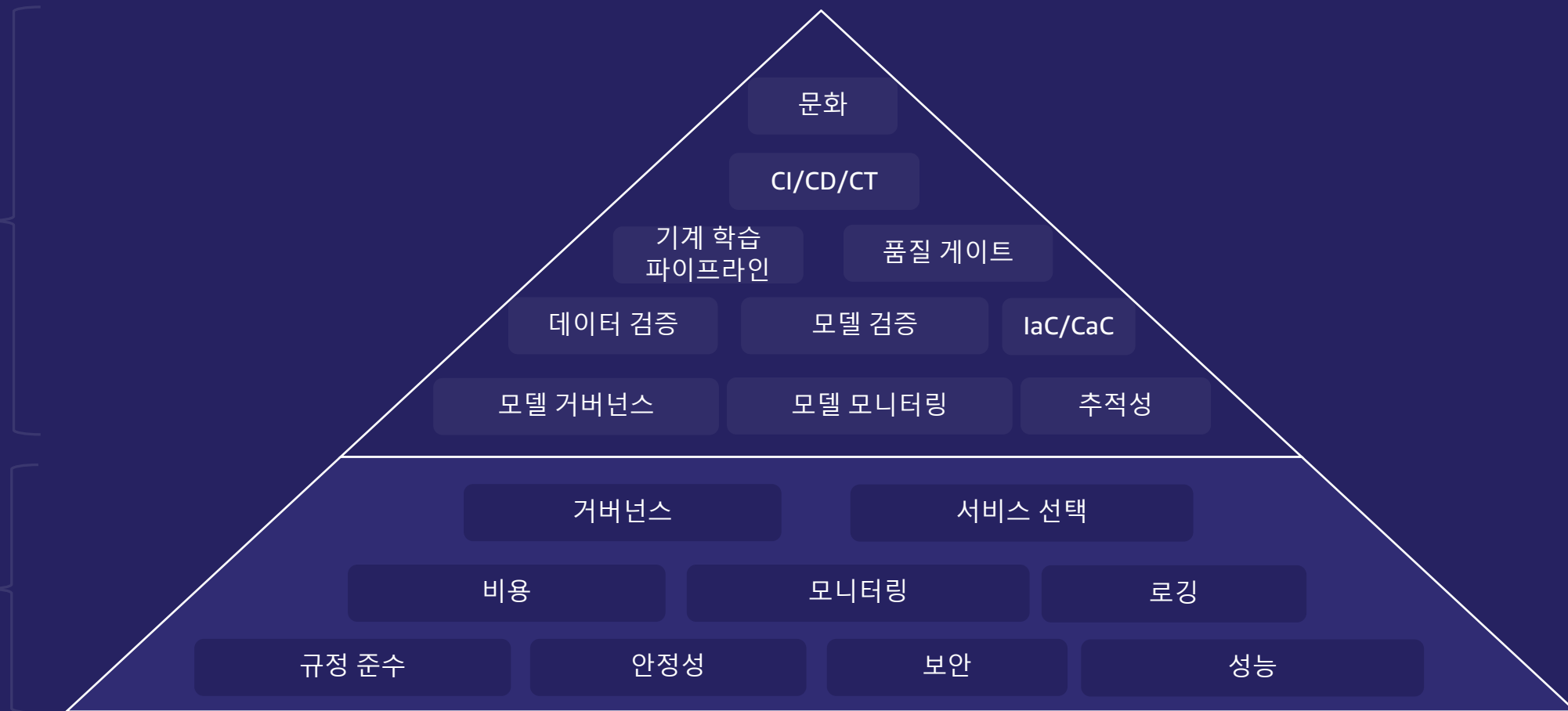
관리형 인프라로 작업을 자동으로 예약 및 실행

기계 학습 운영 가속화 프레임워크

기초 AWS Machine Learning 아키텍처 및 MLOps 사례

MLOps 사례

AWS Machine Learning 아키텍처 사례



MLOps 정의

MLOps 성숙도 모델

	직원	데이터	훈련	배포
초기	<ul style="list-style-type: none"> • 단절된 데이터 과학 및 IT 팀 • 제한된 교차 훈련 	<ul style="list-style-type: none"> • 임시 데이터 수집 및 준비 	<ul style="list-style-type: none"> • 수동 훈련 및 재훈련 • 배포 경로가 명확하지 않음 	<ul style="list-style-type: none"> • 수동 배포
반복 가능	<ul style="list-style-type: none"> • 이해 관계자와의 협업 개선 • 프로젝트 목표 공유 	<ul style="list-style-type: none"> • 자동 데이터 파이프라인 	<ul style="list-style-type: none"> • 실험을 위한 정의된 경로 • 자동 훈련 파이프라인 • 수동 모델 검증 	<ul style="list-style-type: none"> • 자동 배포 파이프라인 • 제한된 모니터링/측정
안정성	<ul style="list-style-type: none"> • 부서 간 프로젝트 팀 • 일부 교차 훈련 	<ul style="list-style-type: none"> • 자동 기계 학습 파이프라인 • 데이터 거버넌스 	<ul style="list-style-type: none"> • 실험 관리 • 자동 기계 학습 파이프라인 • 모델 거버넌스 • 자동 모델 검증 	<ul style="list-style-type: none"> • 자동 기계 학습 파이프라인 • 모니터링 및 로깅(모델, 워크로드, 파이프라인)
확장성	<ul style="list-style-type: none"> • 부서 간 프로젝트 팀 • 교차 훈련 	<ul style="list-style-type: none"> • CI/CD • 코드형 정책 • 코드형 구성 • 자동 검증 	<ul style="list-style-type: none"> • CI/CD • 코드형 정책/구성 • 자동 모델 검증 • 자동 통합 검증 	<ul style="list-style-type: none"> • CI/CD • 코드형 정책/인프라/구성 • 모델 모니터링 • 대시보드 및 투명성

The screenshot shows the Kubeflow Central Dashboard interface. The top navigation bar includes the Kubeflow logo and a 'Select namespace' dropdown. The main content area displays the 'Experiments' section for 'webinar-experiments', with a specific experiment 'cifar10-hpo-train-deploy' selected. The 'Graph' tab is active, showing a vertical sequence of five pipeline steps, each with a green checkmark indicating completion. The steps are: 'sagemaker-hyper...', 'update-best-mod...', 'sagemaker-traini...', 'sagemaker-creat...', and 'sagemaker-deplo...'. A sidebar on the left contains navigation options like Pipelines, Experiments, Artifacts, Executions, Archive, Documentation, Github Repo, and AI Hub Samples. At the bottom, there is a 'Build commit: 743746b' and a 'Report an Issue' link.

Amazon SageMaker capability

Hyperparameter tuning job

Custom function to update epochs

Training job

Create model

Deploy inference endpoint

Kubeflow pipeline

Runtime execution graph. Only steps that are currently running or have already completed are shown.

AWS Machine Learning 스택

가장 방대하고 완전한 기계 학습 기능 세트

AI 서비스



기계 학습 서비스



기계 학습 프레임워크 및 인프라

PyTorch, Apache MXNet, TensorFlow

Amazon EC2

CPU

GPU

AWS Inferentia

AWS Trainium

Habana Gaudi

FPGA

Elastic Inference



AWS Machine Learning 스택

가장 방대하고 완전한 기계 학습 기능 세트

- 거시적 수준에서 기계 학습은 3개 스택 계층이 있는 것으로 생각할 수 있습니다.
- 맨 아래 계층에는 전문 기계 학습 실무자를 위한 프레임워크, 인터페이스 및 인프라가 있습니다.
- 프레임워크 계층에는 가장 많이 사용되는 3가지 프레임워크인 TensorFlow, PyTorch 및 MXNet이 있습니다.
- 오늘날 클라우드 기반 TensorFlow의 92%와 클라우드 기반 PyTorch의 91%가 AWS에서 실행됩니다.
- 다양한 프레임워크에서 새로운 알고리즘이 계속 개발되고 있습니다. 회사는 이러한 알고리즘을 차용하거나 실행하기를 원하지만 다른 프레임워크로 이식하기를 원하지는 않습니다. 결과적으로 AWS에는 모든 주요 프레임워크에 대한 전담 팀이 있으므로 작업에 맞는 적절한 도구를 사용할 수 있습니다.

AWS Machine Learning 스택

가장 방대하고 완전한 기계 학습 기능 세트

- AWS는 고유한 성능 및 예산 요구 사항을 충족하는 프로세서 및 액셀러레이터를 선택할 수 있는 가장 광범위하고 세분화된 기계 학습 인프라 서비스 포트폴리오를 제공합니다. [Amazon EC2 P4d 인스턴스](#)는 클라우드 최초의 400Gbps 인스턴스 네트워킹과 결합된 최신 NVIDIA A100 Tensor Core GPU를 사용하여 클라우드에서 고성능 기계 학습 교육을 제공합니다. P4d 인스턴스는 하이퍼스케일 클러스터(EC2 UltraCluster라고 함)에서 배포되어 복잡한 기계 학습 훈련 작업에서도 슈퍼 컴퓨터급 성능을 발휘합니다.
- 추론의 경우 AWS Inferentia 칩으로 구동되는 [Amazon EC2 Inf1 인스턴스](#)는 클라우드에서 고성능과 최저 비용의 추론을 제공합니다.

AI 및 기계 학습을 위한 가장 광범위하고 세분화된 컴퓨팅 인프라

성능 및 예산 요구 사항에 맞는 CPU, GPU 및 액셀러레이터 선택

기존의 기계 학습

딥 러닝(DL)

훈련 및 추론

추론

훈련



Cascade Lake CPU, Skylake CPU
Habana Gaudi 액셀러레이터



EPYC CPU



Graviton CPU
Inferentia 칩



A100, V100, T4 GPU



AI 및 기계 학습을 위한 가장 광범위하고 세분화된 컴퓨팅 인프라 성능 및 예산 요구 사항에 맞는 CPU, GPU 및 액셀러레이터 선택

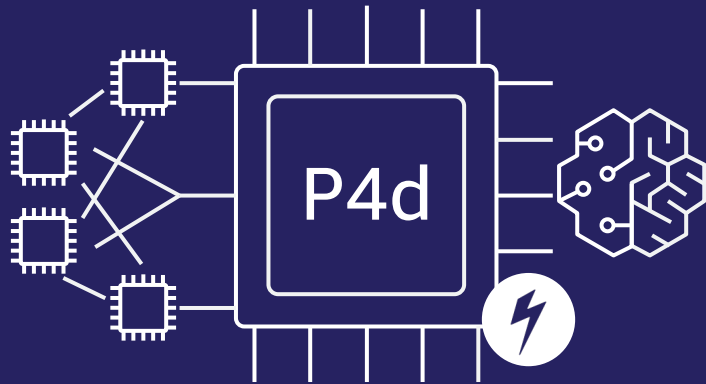
- AWS는 AI 및 기계 학습 훈련과 추론을 위한 가장 광범위하고 세분화된 컴퓨팅 인프라를 제공하는 선두 업체로, 클라우드에서 기계 학습을 위한 GPU 기반 인스턴스를 최초로 출시했습니다. 최신 CPU, GPU 및 사용자 지정 액셀러레이터와 업계 최고의 네트워킹 및 스토리지를 기반으로 하는 다양한 Amazon EC2 컴퓨팅 인스턴스 중에서 예산 요구 사항과 모델의 성능 요구 사항에 적합한 인스턴스를 선택할 수 있습니다. 기존 기계 학습을 실행하는 경우 CPU 기반 인스턴스를 선택하여 중소 규모 기계 학습 모델의 훈련 및 추론을 실행할 수 있습니다. 딥 러닝 모델의 훈련 및 추론을 실행할 고성능 인스턴스를 찾고 있다면 GPU 및 액셀러레이터 기반 인스턴스를 활용할 수 있습니다.
- EC2 P4d 인스턴스는 클라우드에서 딥 러닝 모델을 훈련하기 위한 최고 성능의 컴퓨팅 인스턴스입니다. 이전 세대 GPU 기반 P3 인스턴스보다 2.5배 향상된 성능을 제공하고 60% 저렴한 비용으로 이용할 수 있습니다. 이러한 인스턴스는 대규모 DL 모델에 대한 다중 노드 분산 훈련을 위해 설계되었습니다. P3 인스턴스는 중대형 딥 러닝 모델 훈련 및 단일 노드 기계 학습 훈련을 위한 비용 효율적인 고성능 인스턴스입니다. AWS는 고객이 요구 사항에 가장 적합한 옵션을 선택할 수 있도록 다양한 고성능, 저비용 인프라 옵션을 제공하기 위해 투자 및 혁신하고 있습니다.

AI 및 기계 학습을 위한 가장 광범위하고 세분화된 컴퓨팅 인프라 성능 및 예산 요구 사항에 맞는 CPU, GPU 및 액셀러레이터 선택

- 예를 들어, Habana Labs(인텔 자회사)의 Gaudi 액셀러레이터로 구동되는 EC2 DL1 인스턴스는 딥 러닝 모델 훈련을 위해 동급 GPU 기반 인스턴스보다 40% 더 나은 가격 대비 성능을 제공합니다. 또한 기계 학습 추론 성능에 최적화된 실리콘을 기초부터 구축하고 있습니다. AWS는 Graviton CPU 및 Nitro 시스템 혁신에서 얻은 교훈을 사용하여 클라우드에서 가장 저렴한 추론 비용을 제공하는 맞춤형 칩인 AWS Inferentia 칩을 구축했습니다.
- 고성능 GPU 기반 추론을 위해 G4d 인스턴스를 사용할 수 있습니다. AWS에서 기계 학습 워크로드를 실행하면 기계 학습 및 DL 모델을 훈련하고 배포하기 위해 사용이 간편한 고성능, 저비용 인프라 서비스에 온디맨드 방식으로 액세스할 수 있습니다.

Amazon EC2 P4d 인스턴스 소개

P4d 인스턴스



클라우드의 가장 강력한 GPU
인스턴스 중 하나

최대 60% 더 낮은 교육 비용, 평균 2.5배 더 향상된
딥 러닝 성능, 25% 더 많은 GPU 메모리를 갖춘
기계 학습 모델

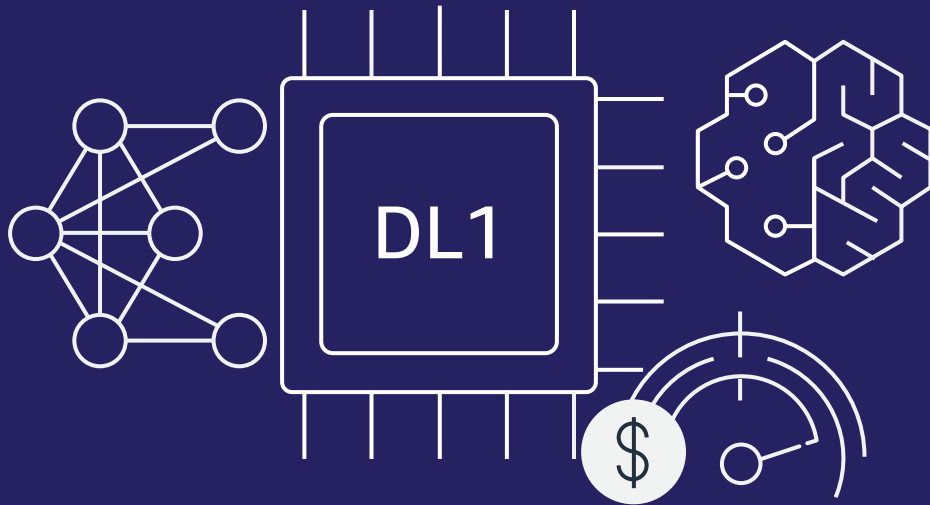
8개의 NVIDIA A100 GPU와 400Gbps의 네트워크
대역폭으로 구동되며 2.5페타플롭의 성능 제공

긴밀하게 결합된 수천 개의 GPU로 구성된
UltraCluster에 배포되어 기계 학습 훈련 및 HPC에
이상적

Amazon EC2 DL1 인스턴스 소개

딥 러닝 모델 훈련을 위한 더 나은 가격 대비 성능

DL1 인스턴스



Habana Labs(인텔 자회사)의 Gaudi 액셀러레이터
최대 8개 탑재

딥 러닝 모델 훈련을 위해 특별히 제작됨

최신 GPU 인스턴스보다 최대 40% 향상된 가격 대비
성능

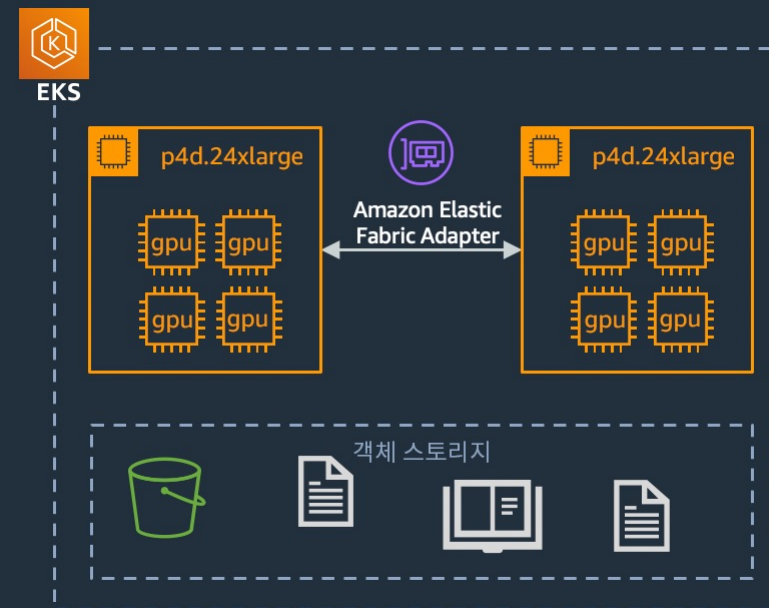
TensorFlow 및 PyTorch와 원활하게 통합된 맞춤형
소프트웨어

DLC, DL AMI 또는 Amazon SageMaker를 사용하여
쉽게 시작

컨테이너화된 기계 학습 애플리케이션용 Amazon
ECS 및 Amazon EKS를 통해 DL1 인스턴스 시작

AWS 기반 대규모 NLP 스타트업

- CA 스타트업은 세계 최대 NLP 모델(2T+ 파라미터)을 구축하는 것을 목표로 합니다.
- 200개의 P4d(800개의 A100 GPU)에서 500억 개의 파라미터 모델을 훈련하기 위해 POC 후에 AWS에서 실행하기로 결정했습니다.
- 훈련 워크로드는 현재 500개 이상의 P4d(4000 GPU)에서 운영 중입니다.
- 네트워킹 속도 향상을 위해 EKS 오케스트레이션 및 EFA를 사용하여 EC2에서 직접 실행됩니다.
- 단일 AZ에 배포되어 여러 AZ에 걸쳐 아키텍처를 확장하려고 합니다.
- **시스템의 주요 인프라 요소:**
 - 오케스트레이션을 위한 EKS
 - 대규모 노드 간 빠른 통신을 지원하는 EFA
 - 500개 이상의 P4d 노드를 보유하고 있으며 계속 늘어나고 있음
 - S3 스토리지 계층
 - AWS 기반 DL AMI를 기반으로 구축된 사용자 지정 DL AMI
 - PyTorch + DDP + NCCL



자체 관리형 기계 학습

진화하는 개발 요구 사항을 충족하기 위한 광범위한 선택 사항

도구

<p>데이터 처리 및 레이블 지정</p> <p>Kinesis Glue EMR SageMaker Ground Truth</p>				<p>DL 프레임워크</p> <p>DL AMI 컨테이너 OSS TensorFlow PYTORCH mxnet</p>		<p>개발, 훈련 및 튜닝(OSS)</p> <p>노트북 Horovod Katib Kubeflow 오퍼레이터 AutoGluon DASK</p>				<p>배포</p> <p>KFServing* AWS Greengrass TF Serving*, NVIDIA Triton*</p>		<p>서버리스</p> <p>AWS Lambda Nuclio*</p>		<p>MLOps</p> <p>CodeBuild CodeCommit CDK ECR Jenkins*</p>
---	--	--	--	---	--	--	--	--	--	--	--	---------------------------------------	--	---

오케스트레이션

AWS 서비스

EKS ECS Fargate Batch ParallelCluster

OSS

Kubernetes

AWS/파트너 서비스

Step Functions CloudFormation Terraform

OSS

Kubeflow 파이프라인 Airflow

워크플로

인프라

컴퓨팅

(평가판)

액셀러레이터

스토리지

S3 EFS Lustre

데이터 전송

Direct Connect Snowcone Snowball Snowball Edge

네트워킹

EFA



시작하기: 다음 단계



시작하기: 다음 단계



협업

기계 학습 전문가로서의
AWS



워크숍 검색

사용 사례 파악



훈련

AWS Machine Learning Embark
프로그램

AWS 딥 디바이스

시작하기: 다음 단계

AWS는 가장 광범위하고 세분화된 기계 학습 서비스 세트를 통해 기계 학습 여정을 안내할 수 있습니다. 기계 학습을 사용하여 해결하고자 하는 비즈니스 과제를 파악하고 올바른 사용 사례를 식별하는 데 도움을 드릴 수 있습니다.

AWS에는 시작하는 데 도움이 되는 교육 및 가속화 프로그램이 있습니다. AWS Machine Learning Embark 프로그램을 통한 기술 및 비즈니스 이해 관계자용 대면 교육과 개발자를 위한 전체 무료 온라인 교육 및 인증 프로그램을 제공합니다. 또한 AWS DeepRacer, AWS DeepComposer 및 AWS DeepLens와 같은 딥 디바이스 포트폴리오는 개발자에게 기계 학습을 학습할 수 있는 재미있고 매력적인 실습 방법을 제공하도록 설계되었습니다.





감사합니다!

문의하기

자세히 알아보기