

## Amazon SageMaker

개요

기능 ▼

요금

FAQ

역할별 ▼

ML 수명 주기별 ▼

시작하기 ▼

고객

파트너

제품 / 기계 학습 / Amazon SageMaker

# Amazon SageMaker FAQ

## 일반

### Q: Amazon SageMaker란 무엇인가요?

SageMaker는 완전관리형 인프라, 도구 및 워크플로를 사용하여 모든 사용 사례에 대해 데이터를 준비하고 기계 학습(ML) 모델을 구축, 훈련 및 배포하는 완전관리형 서비스입니다.

### Q: SageMaker는 어느 AWS 리전에서 사용할 수 있나요?

지원되는 SageMaker 리전 목록은 [AWS 리전 서비스 페이지](#)를 참조하세요. 자세한 내용은 AWS 일반 참조 가이드에서 [리전 엔드포인트](#)를 참조하세요.

### Q: SageMaker의 서비스 가용성은 어떤가요?

SageMaker는 고가용성을 제공하도록 설계되었습니다. 유지 관리 기간이나 예약된 가동 중지 시간이 없습니다. SageMaker API는 입증된 고가용성 Amazon 데이터 센터에서 실행되며 각 리전의 시설 세 곳에 서비스 스택 복제가 구성되어 있어 서버 장애나 가용 영역 중단 시에 내결함성을 제공합니다.

### Q: SageMaker는 어떤 방식으로 코드 보안을 제공하나요?

SageMaker는 보안 그룹과 저장 중 암호화(선택 사항)로 보호된 ML 스토리지 볼륨에 코드를 저장합니다.

### Q: SageMaker에는 어떤 보안 조치가 있나요?

SageMaker는 전송 중이거나 저장 중인 ML 모델 아티팩트 및 기타 시스템 아티팩트를 암호화합니다. SageMaker API 및 콘솔에 대한 요청은 보안(SSL) 연결을 통해 전달됩니다. 사용자는 [AWS Identity and Access Management 역할](#)을 SageMaker로 전달하여 사용자 대신, 훈련 및 배포를 위해 리소스에 액세스할 수 있는 권한을 제공합니다. 모델 아티팩트 및 데이터에 대해 암호화된 Amazon Simple Storage Service(S3) 버킷을 사용하고 AWS Key Management Service(AWS KMS) 키를 SageMaker 노트북, 훈련 작업 및 엔드포인트로 전달하여 연결된 ML 스토리지 볼륨을 암호화할 수 있습니다. 또한 SageMaker는 Amazon Virtual Private Cloud(VPC) 및 AWS PrivateLink 지원도 제공합니다.

### Q: SageMaker는 모델, 훈련 데이터 또는 알고리즘을 사용하거나 공유하나요?

SageMaker는 고객 모델, 훈련 데이터 또는 알고리즘을 사용하거나 공유하지 않습니다. AWS는 프라이버시와 데이터 보안이 고객에게 얼마나 중요한지 잘 알고 있습니다. 따라서 AWS에서는 콘텐츠를 저장할 장소를 결정하고, 저장 콘텐츠 및 전송 콘텐츠를 보호하며, AWS 서비스와 리소스에 대한 사용자의 액세스를 관리할 수 있는 단순하지만 강력한 도구를 통해 고객에게 콘텐츠에 대한 소유권과 제어권을 제공합니다. 또한 AWS는 고객 콘텐츠에 대한 무단 액세스나 공개를 차단하도록 설계된 기술적 제어 및 물리적 제어를 구현합니다. 고객은 자신의 콘텐츠에 대한 소유권을 유지하고 콘텐츠를 처리, 저장 및 호스팅할 AWS 서비스를 선택합니다. AWS에서는 고객의 동의 없이는 어떠한 목적으로도 고객 콘텐츠에 액세스하지 않습니다.

### Q: SageMaker 사용 요금은 어떻게 부과되나요?

노트북을 호스팅하고 모델을 훈련하고 예측을 수행하며 출력을 로깅하는 데 사용하는 ML 컴퓨팅, 스토리지 및 데이터 처리 리소스에 대한 요금이 부과됩니다. SageMaker를 사용할 때는 호스팅된 노트북, 훈련 및 모델 호스팅에 사용되는 인스턴스 수 및 유형을 선택할 수 있습니다. 서비스를 사용하면서 사용한 만큼만 비용을 지불하며 최소 요금 및 사전 약정은 없습니다. 자세한 내용은 [Amazon SageMaker 요금](#) 및 [Amazon SageMaker 요금 계산기](#)를 참조하세요.

**Q: 유류 리소스를 검색하고 중지하여 불필요한 요금을 방지하는 것과 같이 SageMaker 비용을 최적화하려면 어떻게 해야 하나요?**

SageMaker 리소스 사용량을 최적화하는 데 사용할 수 있는 여러 모범 사례가 있습니다. 구성 최적화를 수행하는 것과 관련된 접근 방식이 있고 프로그래밍 방식 솔루션을 제공하는 접근 방식도 있습니다. [이 블로그 게시물](#)에서 시각적 자습서 및 코드 샘플과 함께 이 개념에 대한 전체 가이드를 확인할 수 있습니다.

**Q: 자체 노트북, 훈련 또는 호스팅 환경이 있는 경우는 어떻게 되나요?**

SageMaker는 완전한 전체 워크플로를 제공하지만, SageMaker에서 기존 도구를 계속 사용할 수 있습니다. 비즈니스 요구 사항에 따라 필요한 대로 SageMaker에서 각 단계의 결과를 손쉽게 송수신할 수 있습니다.

**Q: SageMaker에서 R이 지원되나요?**

예. SageMaker 노트북 인스턴스에서 R을 사용할 수 있으며, 여기에는 사전 설치된 R 커널과 [reticulate](#) 라이브러리가 포함됩니다. Reticulate는 Amazon SageMaker Python SDK를 위한 R 인터페이스를 제공하여 ML 실무자의 R 모델 구축, 훈련, 튜닝 및 배포 작업을 지원합니다.

**Q: 모델에서 불균형을 확인하려면 어떻게 해야 하나요?**

[Amazon SageMaker Clarify](#)를 통해 전체 ML 워크플로에서 통계적인 바이어스를 탐지해 모델 투명성을 개선할 수 있습니다. SageMaker Clarify는 데이터 준비 도중, 훈련 후, 시간 경과에 따른 불균형을 검사하고, ML 모델 및 예측을 설명해주는 도구도 포함합니다. 조사 결과는 설명 보고서를 통해 공유할 수 있습니다.

**Q: Amazon SageMaker Clarify는 어떤 종류의 바이어스를 탐지하나요?**

ML 모델에서 바이어스 평가는 바이어스를 완화하는 첫 단계입니다. 바이어스는 훈련 전후에, 그리고 배포된 모델에 대한 추론을 위해 평가할 수 있습니다. 바이어스의 각 평가는 다양한 공정성 개념을 반영합니다. 단순한 공정성 개념을 평가하는 작업도 다양한 컨텍스트에서 다양한 많은 평가를 적용할 수 있습니다. 사용자는 조사항과 애플리케이션에 유효한 바이어스 개념과 지표를 선택해야 합니다. 현재 SageMaker는 훈련 데이터(SageMaker 데이터 준비의 일부), 훈련된 모델(Amazon SageMaker Experiments의 일부), 배포된 모델의 추론(Amazon SageMaker Model Monitor의 일부)을 위해 다양한 바이어스 지표 계산을 지원합니다. 예를 들어 훈련 전에 훈련 데이터가 대표성을 갖는지(즉, 한 그룹에서 대표성이 부족하지 여부)와 여러 그룹에서 레이블 분산에 차이가 있는지를 검사하는 지표를 제공합니다. 훈련 후 또는 배포 중에는 여러 그룹에서 모델의 성능이 다른지 여부와 다르다면 얼마나 다른지를 평가하는 데 이러한 지표를 유용하게 활용할 수 있습니다. 예를 들어 오류 비율(모델의 예측이 실제 레이블과 차이가 날 가능성), 더 세분화된 정밀도로 분류(긍정적인 예측이 올바른 가능성), 회수(모델이 긍정적인 예제에 레이블을 올바르게 지정할 가능성)를 비교하는 작업부터 시작할 수 있습니다.

**Q: SageMaker Clarify는 모델 설명 가능성을 어떻게 개선하나요?**

SageMaker Clarify는 SageMaker Experiments에 통합되어 모델 훈련 후에 모델의 전반적인 의사 결정 프로세스에서 각 입력의 중요성을 자세히 설명하는 특성 중요도 그래프를 제공합니다. 이와 같은 세부 정보는 특정 모델 입력이 전체 모델 동작에 미치는 영향보다 더 많은 영향을 미치는지 확인하는 데 유용할 수 있습니다. SageMaker Clarify는 API를 통해 개별 예측에 대한 설명도 제공할 수 있습니다.

**Q: Amazon SageMaker Studio란 무엇인가요?**

SageMaker Studio는 모든 ML 개발 단계를 수행할 수 있는 웹 기반의 단일 시각적 인터페이스를 제공합니다. SageMaker Studio는 데이터를 준비하고 모델을 구축, 훈련 및 배포하는 데 필요한 액세스 권한, 제어 및 가시성을 각 단계별로 완벽하게 제공합니다. 신속하게 데이터를 업로드하고, 새로운 노트북을 생성하며, 모델을 훈련 및 튜닝하고, 단계를 앞뒤로 이동하며 실험을 조정하고, 결과를 비교하며, 모델을 프로덕션에 배포하여 한 곳에서 생산성을 크게 높일 수 있습니다. 노트북, 실험 관리, 자동 모델 생성, 디버깅 및 프로파일링, 모델 드리프트 탐지를 포함한 모든 ML 개발 활동은 통합된 SageMaker Studio 시각적 인터페이스 내에서 수행할 수 있습니다.

**Q: RStudio on Amazon SageMaker란 무엇인가요?**

RStudio on SageMaker는 클라우드 최초의 완전관리형 RStudio Workbench입니다. 익숙한 RStudio 통합 개발 환경(IDE)을 신속하게 시작할 수 있고, 작업을 중단하지 않고도 기본 컴퓨팅 리소스를 탄력적으로 늘리고 줄일 수 있으며, R에서 대규모로 ML 및 분석 솔루션을 손쉽게 구축할 수 있습니다. RStudio IDE와 R 및 Python 개발용 SageMaker Studio 노트북 간에 원활하게 전환할 수 있습니다. 코드, 데이터 세트, 리포지토리 및 기타 아티팩트를 비롯한 모든 작업이 두 환경 간에 자동으로 동기화되므로 컨텍스트 전환이 줄어들고 생산성이 개선됩니다.

**Q: SageMaker Studio 요금은 어떻게 적용되나요?**

SageMaker Studio 사용에 대한 추가 요금은 없습니다. SageMaker Studio 내에서 사용하는 서비스의 기본 컴퓨팅 및 스토리지 요금만 지불하면 됩니다.

**Q: SageMaker Studio는 어느 리전에서 지원되나요?**

SageMaker Studio를 지원하는 리전은 [Amazon SageMaker 개발자 안내서](#)에서 확인할 수 있습니다.

## ML 거버넌스

### Q: SageMaker는 어떤 ML 거버넌스 도구를 제공하나요?

SageMaker는 ML 수명 주기 전반에 걸쳐 목적별 ML 거버넌스 도구를 제공합니다. 관리자는 Amazon SageMaker 역할 관리자를 사용하여 최소 권한을 몇 분 만에 정의할 수 있습니다. Amazon SageMaker 모델 카드를 사용하면 필수 모델 정보를 개념 단계부터 배포 단계까지 손쉽게 캡처, 검색 및 공유할 수 있고, Amazon SageMaker 모델 대시보드라는 한 장소에서 프로덕션 모델 동작에 대한 정보를 지속적으로 확인할 수 있습니다. 자세한 내용은 [Amazon SageMaker를 사용한 ML 거버넌스](#)를 참조하세요.

### Q: SageMaker 역할 관리자는 어떤 기능을 제공하나요?

SageMaker 역할 관리자를 사용하면 최소 권한을 몇 분 만에 정의할 수 있습니다. SageMaker 역할 관리자는 사전 구축된 IAM 정책 카탈로그를 통해 ML 활동 및 페르소나에 대한 기본 권한 세트를 제공합니다. 기존 권한을 유지하거나 특정한 요구 사항에 따라 추가로 사용자 지정할 수 있습니다. 몇 가지 셀프 구성 화면을 통해 네트워크 액세스 경계, 암호화 키 등 일반적인 거버넌스 구성을 빠르게 입력할 수 있습니다. 그러면 SageMaker Role Manager가 IAM 정책을 자동으로 생성합니다. AWS IAM 콘솔을 통해 생성된 역할과 관련 정책을 검색할 수 있습니다. 사용 사례에 대한 권한을 추가로 사용자 지정하려면, SageMaker Role Manager에서 생성하는 IAM 역할에 관리형 IAM 정책을 연결합니다. 또한 AWS 서비스 전반에서 역할을 식별하고 구성하는 데 유용한 태그를 추가할 수 있습니다.

### Q: SageMaker 모델 카드는 어떤 기능을 제공하나요?

SageMaker 모델 카드는 모델 정보의 단일 소스를 생성하여 ML 수명 주기 전반에서 모델 설명서를 중앙 집중화하고 표준화하는 데 도움이 됩니다. SageMaker 모델 카드는 훈련 세부 정보를 자동으로 입력하여 문서화 프로세스의 속도를 높여줍니다. 모델의 용도 및 성능 목표와 같은 세부 정보를 추가할 수도 있습니다. 모델 평가 결과를 모델 카드에 첨부하고, 시각화 요소를 제공하여 모델 성능에 대한 주요 인사이트를 얻을 수 있습니다. SageMaker 모델 카드를 PDF 형식으로 내보내 다른 사용자와 손쉽게 공유할 수 있습니다.

### Q: SageMaker 모델 대시보드는 어떤 기능을 제공하나요?

SageMaker 모델 대시보드는 배포된 모델과 엔드포인트에 대한 전반적인 개요 정보를 제공하므로, 단일 창에서 리소스를 추적하고 동작 위반을 모델링할 수 있습니다. 대시보드에서는 4가지 차원으로 모델 동작을 모니터링할 수 있습니다. 예를 들어 데이터 및 모델 품질로 모니터링하거나 SageMaker Model Monitor 및 SageMaker Clarify와 통합할 경우 바이어스 및 특성 기여도 드리프트로 모니터링할 수 있습니다. 또한 SageMaker 모델 대시보드는 누락되고 비활성화된 모델 모니터링 작업과 모델 품질, 데이터 품질, 바이어스 드리프트 및 특성 기여도 드리프트에 대한 모델 동작의 편차에 대한 알림을 설정하고 받을 수 있는 통합 경험을 제공합니다. 개별 모델을 추가로 검사하고 시간 경과에 따라 모델 성능에 영향을 미치는 요인을 분석할 수 있습니다. 그런 다음 ML 실무자와 함께 시정 조치를 취할 수 있습니다.

## 기초 모델

### Q: SageMaker를 빠르게 시작하는 방법은 무엇인가요?

SageMaker JumpStart를 사용하면 ML을 빠르고 쉽게 시작할 수 있습니다. SageMaker JumpStart는 몇 단계로 손쉽게 배포할 수 있는, 가장 일반적인 사용 사례를 위한 솔루션 세트를 제공합니다. 솔루션은 모든 사용자 지정이 가능하며 AWS CloudFormation 템플릿과 참조 아키텍처 사용을 보여주므로 ML 여정의 속도를 높일 수 있습니다. 또한 SageMaker JumpStart는 기초 모델을 제공하고 한 단계로 배포와 변환 도구, 객체 탐지 및 이미지 분류 모델과 같은 150개가 넘는 널리 사용되는 오픈 소스 모델에 대한 세분화된 튜닝도 지원합니다.

### Q: SageMaker JumpStart에서 사용할 수 있는 기초 모델에는 어떤 것이 있나요?

SageMaker JumpStart는 독점 모델과 공개 모델을 제공합니다. 사용 가능한 기초 모델 목록은 [Amazon SageMaker JumpStart 시작하기](#)를 참조하세요.

### Q: SageMaker JumpStart에서 기초 모델을 시작하려면 어떻게 해야 하나요?

SageMaker Studio, SageMaker SDK 및 AWS Management Console을 통해 기초 모델에 액세스할 수 있습니다. 독점 기초 모델을 시작하려면 [AWS Marketplace](#)의 판매 약관에 동의해야 합니다.

### Q: SageMaker JumpStart를 사용하는 고객에게 제공되는 기본 모델을 업데이트하는 데 내 데이터가 사용되거나 공유되나요?

아니요. 훈련 및 훈련 데이터는 SageMaker JumpStart가 고객에게 제공하는 기본 모델의 업데이트 또는 훈련에 사용되거나 공유되지 않습니다.

### Q: SageMaker JumpStart를 통해 독점 모델의 모델 가중치 및 스크립트를 볼 수 있나요?

아니요. 독점 모델에서는 모델 가중치 및 스크립트를 볼 수 없습니다.

### Q: SageMaker JumpStart 기초 모델을 사용할 수 있는 리전은 어디인가요?

SageMaker Studio를 사용할 수 있는 모든 [리전](#)에서 모델을 검색할 수 있지만 모델 배포 기능은 필요한 인스턴스 유형의 모델 및 인스턴스 가용성에 따라 다릅니다. [AWS Marketplace](#)의 모델 세부 정보 페이지에서 AWS 리전 가용성 및 필요한 인스턴스를 참조할 수 있습니다.

#### Q: SageMaker JumpStart 기초 모델의 요금은 어떻게 부과되나요?

독점 모델의 경우 모델 제공자가 결정한 소프트웨어 요금과 사용된 인스턴스를 기준으로 한 SageMaker 인프라 요금이 부과됩니다. 공개적으로 사용 가능한 모델의 경우 사용된 인스턴스를 기준으로 한 SageMaker 인프라 요금이 부과됩니다. 자세한 내용은 [Amazon SageMaker 요금](#) 및 [AWS Marketplace](#)를 참조하세요.

#### Q: SageMaker JumpStart는 데이터의 보호 및 보안에 어떤 도움이 되나요?

보안은 AWS의 최우선 과제이며 SageMaker JumpStart는 안전하도록 설계되었습니다. 따라서 SageMaker에서는 콘텐츠를 저장할 장소를 결정하고, 저장 콘텐츠 및 전송 콘텐츠를 보호하며, AWS 서비스와 리소스에 대한 사용자의 액세스를 관리할 수 있는 단순하지만 강력한 도구를 통해 고객에게 콘텐츠에 대한 소유권과 제어권을 제공합니다.

1. AWS는 고객의 훈련 및 추론 정보를 AWS Marketplace의 모델 판매자와 공유하지 않습니다. 마찬가지로 판매자의 모델 아티팩트(예: 모델 가중치)는 구매자와 공유되지 않습니다.
2. SageMaker JumpStart는 고객의 모델, 훈련 데이터 또는 알고리즘을 사용하여 서비스를 개선하지 않으며 고객의 훈련 및 추론 데이터를 서드 파티와 공유하지 않습니다.
3. SageMaker JumpStart에서는 ML 모델 아티팩트가 전송 및 저장 중에 암호화됩니다.
4. [AWS Shared Responsibility Model](#)에 따라 AWS는 AWS의 모든 것이 실행되는 글로벌 인프라를 보호할 책임이 있습니다. 이 인프라에서 호스팅되는 콘텐츠에 대한 제어를 유지할 책임은 사용자에게 있습니다.

AWS Marketplace 또는 SageMaker JumpStart의 모델을 사용할 때 사용자는 모델 출력 품질에 대한 책임을 지고 개별 모델 설명에 설명된 기능 및 한계를 인정합니다.

#### Q: SageMaker JumpStart에서 지원하는 오픈 소스 모델로는 어떤 것이 있나요?

SageMaker JumpStart에는 PyTorch Hub 및 TensorFlow Hub의 사전 훈련된 150개 이상의 오픈 소스 모델이 포함되어 있습니다. 이미지 분류 및 객체 탐지와 같은 비전 태스크의 경우 RESNET, MobileNet, Single-Shot Detector(SSD)와 같은 모델을 사용할 수 있습니다. 문장 분류, 텍스트 분류, 질문 답변과 같은 텍스트 태스크의 경우 BERT, RoBERTa, DistilBERT와 같은 모델을 사용할 수 있습니다.

#### Q: ML 아티팩트를 조직 내의 다른 사용자와 공유하려면 어떻게 해야 하나요?

SageMaker JumpStart를 사용하면 노트북 및 모델을 포함한 ML 아티팩트를 조직 내에서 손쉽게 공유할 수 있습니다. 관리자는 정의된 사용자 집합이 액세스할 수 있는 리포지토리를 설정할 수 있습니다. 리포지토리 액세스 권한이 있는 모든 사용자는 모델 및 노트북과 SageMaker JumpStart 내부의 공개 콘텐츠를 찾아보고 검색하고 사용할 수 있습니다. 사용자는 SageMaker JumpStart에서 아티팩트를 선택하여 모델을 훈련하고 엔드포인트를 배포하며 노트북을 실행할 수 있습니다.

#### Q: 조직 내의 다른 사용자와 ML 아티팩트를 공유할 때 SageMaker JumpStart를 사용해야 하는 이유는 무엇인가요?

SageMaker JumpStart를 사용하면 구축하는 ML 애플리케이션의 출시를 가속화할 수 있습니다. 조직 내에서 한 팀이 구축한 모델 및 노트북을 단 몇 단계로 손쉽게 조직 내의 다른 팀과 공유할 수 있습니다. 내부에서 지식을 공유하고 자산을 재사용하면 조직의 생산성이 크게 개선될 수 있습니다.

## 로우 코드 ML

#### Q: Amazon SageMaker Autopilot이란 무엇인가요?

[SageMaker Autopilot](#)은 업계 최초의 자동화된 기계 학습 기능으로, ML 모델을 완벽하게 제어하고 파악할 수 있는 기능을 제공합니다. SageMaker Autopilot은 원시 데이터를 자동으로 검사하고, 특성 프로세서를 적용하며, 최상의 알고리즘 세트를 선택하고, 여러 가지 모델을 훈련 및 튜닝하며, 해당 성능을 추적한 다음 성능을 기준으로 모델 등급을 지정합니다. 이 모든 작업이 클릭 몇 번으로 수행됩니다. 이 작업의 결과로 최고 성능의 모델이 나오는데, 일반적으로 모델 훈련에 필요한 시간에 비해 아주 짧은 시간으로 모델을 배포할 수 있습니다. 모델이 생성된 방식과 모델에 포함된 내용을 완벽하게 파악할 수 있으며 SageMaker Autopilot을 SageMaker Studio와 통합할 수 있습니다. SageMaker Studio 내에서 SageMaker Autopilot으로 생성된 서로 다른 모델을 50개까지 탐색할 수 있으므로 사용 사례에 가장 적합한 모델을 쉽게 선택할 수 있습니다. SageMaker Autopilot을 사용하면 ML 경험이 없는 사람도 모델을 쉽게 만들 수 있고, 숙련된 개발자라면 팀에서 나중에 반복해서 사용할 수 있는 기준 모델을 개발할 수 있습니다.

#### Q: SageMaker Autopilot에서 지원되는 기본 제공 알고리즘은 무엇인가요?

SageMaker Autopilot은 XGBoost와 선형 회귀라는 2가지 기본 제공 알고리즘을 지원합니다.

#### Q: SageMaker Autopilot 작업을 수동으로 중지할 수 있나요?

예. 언제든지 작업을 중지할 수 있습니다. SageMaker Autopilot 작업이 중지되면 진행 중인 모든 실험이 중지되고 새 실험이 시작되지 않습니다.

#### Q: Amazon SageMaker Canvas란 무엇인가요?

SageMaker Canvas는 마우스 클릭 방식의 직관적인 인터페이스를 제공하는 노코드 서비스로, 데이터에서 매우 정확한 ML 기반 예측을 생성하는 데 사용할 수 있습니다. SageMaker Canvas를 사용하면 드래그 앤 드롭 방식의 사용자 인터페이스를 사용하여 다양한 소스의 데이터에 액세스하고 이러한 데이터를 결합할 수 있습니다. 데이터가 자동으로 정리되고 준비되므로 수동 정리 작업이 최소화됩니다. SageMaker Canvas는 다양한 최첨단 ML 알고리즘을 적용하여 고도로 정확한 예측 모델을 찾으며, 예측을 수행할 수 있는 직관적인 인터페이스를 제공합니다. SageMaker Canvas를 사용하면 다양한 비즈니스 애플리케이션에서 훨씬 더 정확한 예측을 수행하고 모델, 데이터 및 보고서를 공유하여 기업의 데이터 사이언티스트 및 분석가와 쉽게 협업할 수 있습니다. SageMaker Canvas에 대해 자세히 알아보려면 [Amazon SageMaker Canvas FAQ](#)를 참조하세요.

#### Q: SageMaker Canvas 요금은 어떻게 부과되나요?

SageMaker Canvas 요금은 사용량에 따라 부과됩니다. SageMaker Canvas를 사용하면 여러 소스의 데이터를 대화형으로 수집, 탐색 및 준비하고, 데이터를 사용하여 고도로 정확한 ML 모델을 훈련하고 예측을 생성할 수 있습니다. 요금은 2가지 구성 요소로 결정됩니다. 세션 요금은 SageMaker Canvas를 사용하거나 로그인한 시간에 따라 부과되며 모델 훈련 요금은 모델 구축 시 사용한 데이터 세트의 크기에 따라 부과됩니다. 자세한 내용은 [Amazon SageMaker Canvas 요금](#)을 참조하세요.

## ML 워크플로

#### Q: SageMaker를 사용하여 지속적 통합 및 전송(CI/CD) 파이프라인을 구축하려면 어떻게 해야 하나요?

[Amazon SageMaker Pipelines](#)는 데이터 준비에서 모델 배포에 이르는 완전 자동화된 ML 워크플로를 생성하여 프로덕션 환경에서 수천 개의 ML 모델로 확장하는 데 도움이 됩니다. SageMaker Pipelines는 SageMaker Studio에 연결되는 Python SDK도 함께 제공하므로, 시각적 인터페이스를 활용하여 워크플로의 각 단계를 구축할 수 있습니다. 그런 다음, 단일 API를 사용하여 각 단계를 연결하고 완전한 워크플로를 생성할 수 있습니다. SageMaker Pipelines는 단계 사이에서 데이터를 관리하고, 코드 레시피를 패키지로 작성하며, 실행을 조정합니다. 따라서 수개월이 걸리던 코딩 작업이 몇 시간으로 줄어듭니다. 워크플로를 실행할 때마다 수행되는 작업과 처리되는 데이터의 전체 레코드가 유지되므로, 데이터 사이언티스트와 기계 학습 개발자는 문제를 빠르게 디버깅할 수 있습니다.

#### Q: 프로덕션으로 이전할 최상의 모델을 선택하기 위해 훈련된 모든 모델을 보려면 어떻게 해야 하나요?

SageMaker Pipelines는 모델 레지스트리라고 하는 훈련된 모델의 중앙 리포지토리를 제공합니다. Python SDK를 통해 프로그래밍 방식으로 또는 SageMaker Studio를 통해 시각적으로 모델을 검색하고 모델 레지스트리에 액세스할 수 있으므로 프로덕션에 배포할 원하는 모델을 간편하게 선택할 수 있습니다.

#### Q: SageMaker Pipelines에 추가할 수 있는 SageMaker 구성 요소는 무엇인가요?

SageMaker Clarify, Amazon SageMaker Data Wrangler, Amazon SageMaker 특성 저장소, Amazon SageMaker Experiments, Amazon SageMaker Debugger 및 Amazon SageMaker Model Monitor 등 SageMaker Studio를 통해 사용 가능한 구성 요소를 SageMaker Pipelines에 추가할 수 있습니다.

#### Q: 전체 ML 워크플로에서 모델 구성 요소를 추적하려면 어떻게 해야 하나요?

SageMaker Pipelines는 모든 모델 구성 요소를 자동으로 추적하고 모든 변경 사항에 대한 감사 추적을 유지하므로 수동 추적은 필요하지 않으며 규제 준수 목표를 달성하는 데 도움이 될 수 있습니다. SageMaker Pipelines에서는 데이터, 코드, 훈련된 모델 등을 추적할 수 있습니다.

#### Q: SageMaker Pipelines 요금은 어떻게 부과되나요?

SageMaker Pipelines 사용에 대한 추가 비용은 없습니다. SageMaker Pipelines 내에서 사용하는 개별 AWS 서비스 또는 기본 컴퓨팅에 대한 비용만 지불하면 됩니다.

#### Q: SageMaker에서 Kubeflow를 사용할 수 있나요?

예. Amazon SageMaker Components for Kubeflow Pipelines는 Kubeflow Pipelines를 사용하여 ML 워크플로를 정의하고 SageMaker를 사용하여 데이터 레이블링, 훈련 및 추론 단계를 수행할 수 있도록 하는 오픈 소스 플러그인입니다. Kubeflow Pipelines는 확장 가능하고 이식 가능하며 완전한 ML 파이프라인을 구축 및 배포할 수 있는 Kubeflow 추가 기능입니다. 그러나 Kubeflow Pipelines를 사용할 때 ML 운영 팀은 Kubernetes 클러스터의 CPU 및 GPU 인스턴스를 관리하고 사용량을 계속 높게 유지하여 운영 비용을 절감해야 합니다. 데이터 과학 팀 전체에서 클러스터의 사용량을 극대화하는 것은 어려운 일이며 ML 운영 팀에 추가적인 운영 오버헤드가 발생합니다. ML 최적화 Kubernetes 클러스터 대신, SageMaker Components for Kubeflow Pipelines를 사용하면 기계 학습 작업을 실행하기 위해 Kubernetes 클러스터를 특별히 구성 및 관리할 필요 없이 데이터 레이블링, 완전관리형 대규모 하이퍼파라미터 튜닝 및 분산 훈련 작업, 안전하고 확장 가능한 원클릭 모델 배포, Amazon Elastic Compute Cloud(Amazon EC2) 스팟 인스턴스를 통한 비용 효율적인 훈련과 같은 강력한 SageMaker 기능을 활용할 수 있습니다.

#### Q: SageMaker Components for Kubeflow Pipelines 요금은 어떻게 부과되나요?



SageMaker Components for Kubeflow Pipelines 사용에 대한 추가 비용은 없습니다.

## 데이터 준비

### Q: SageMaker에서는 ML용 데이터를 어떻게 준비하나요?

SageMaker Data Wrangler는 ML용 데이터를 집계하고 준비하는 시간을 줄여줍니다. SageMaker Studio의 단일 인터페이스를 통해 단 몇 단계로 Amazon S3, Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon SageMaker 특성 저장소 및 Snowflake에서 데이터를 찾아서 가져올 수 있습니다. 40개 이상의 데이터 소스에서 전송되어 Amazon AppFlow를 통해 AWS Glue 데이터 카탈로그에 등록된 데이터를 쿼리하고 가져올 수도 있습니다. SageMaker Data Wrangler는 원시 데이터를 자동으로 로드하고 집계한 후 표시합니다. 데이터를 SageMaker Data Wrangler로 가져오면 자동으로 생성된 열 요약 및 히스토그램이 표시됩니다. SageMaker Data Wrangler의 데이터 품질 및 인사이트 보고서를 사용하면 제공된 요약 통계 및 데이터 품질 경고를 바탕으로 데이터를 더 자세히 살펴보고 잠재적 오류를 식별할 수 있습니다. SageMaker Clarify에서 지원하는 바이어스 분석을 SageMaker Data Wrangler에서 직접 실행하여 데이터 준비 중에 잠재적 바이어스를 감지할 수도 있습니다. 거기에서 SageMaker Data Wrangler의 미리 구축된 변환을 사용하여 데이터를 준비할 수 있습니다. 데이터가 준비되면 Amazon SageMaker Pipelines에서 완전히 자동화된 ML 워크플로를 구축하거나 [Amazon SageMaker 특성 저장소](#)로 해당 데이터를 가져올 수 있습니다.

### Q: SageMaker Data Wrangler에서 모델 특성을 생성하려면 어떻게 해야 하나요?

SageMaker Data Wrangler를 사용하면 코드 작성 없이 데이터를 새 특성으로 자동으로 변환할 수 있습니다. SageMaker Data Wrangler를 사용할 때는 미리 구성된 데이터 변환, 누락 데이터 산입, 원-핫 인코딩, 주성분 분석(PCA)을 사용한 차원 축소와 시계열 관련 변환기를 선택할 수 있습니다. 예를 들어 한 단계로 텍스트 필드 열을 숫자 열로 변환할 수 있습니다. 또한 SageMaker Data Wrangler의 코드 조각 라이브러리를 사용하여 코드 조각을 작성할 수도 있습니다.

### Q: SageMaker Data Wrangler에서 데이터를 시각화하려면 어떻게 해야 하나요?

SageMaker Data Wrangler는 미리 구성된 강력한 시각화 템플릿 세트를 통해 데이터를 이해하고 잠재적 오류와 극단 값을 식별하는 데 도움이 됩니다. 코드 작성 없이 히스토그램, 산점도 및 ML 특정 시각화(예: 대상 유출 탐지)를 모두 사용할 수 있습니다. 고유한 시각화를 생성하고 편집할 수도 있습니다.

### Q: SageMaker Data Wrangler 요금은 어떻게 부과되나요?

SageMaker Data Wrangler에 대해 사용하는 모든 ML 컴퓨팅, 스토리지 및 데이터 처리 리소스에 대한 요금만 청구됩니다. SageMaker Data Wrangler 요금에 대한 모든 세부 정보는 [여기](#)에서 검토할 수 있습니다. [AWS 프리 티어](#)를 사용하는 고객은 SageMaker Data Wrangler를 무료로 시작할 수도 있습니다.

### Q: SageMaker Data Wrangler에서 준비된 데이터로 ML 모델을 훈련하려면 어떻게 해야 하나요?

SageMaker Data Wrangler는 [SageMaker Autopilot](#)에서 데이터를 준비하고 기계 학습 모델을 원활하게 훈련할 수 있는 통합 경험을 제공합니다. SageMaker Autopilot은 데이터를 기반으로 최상의 ML 모델을 자동으로 구축, 훈련 및 튜닝합니다. SageMaker Autopilot을 사용하면 데이터 및 모델의 완벽한 제어와 가시성을 유지 관리할 수 있습니다. SageMaker Data Wrangler에서 준비한 특성을 기존 모델에 사용할 수도 있습니다. SageMaker Data Wrangler 처리 작업을 SageMaker 훈련 파이프라인의 일부로 실행되도록 구성할 수 있습니다. 사용자 인터페이스(UI)에서 작업을 구성하거나 오케스트레이션 코드로 노트북을 내보내면 됩니다.

### Q: 기록 데이터를 기반으로 특성을 준비한 경우 SageMaker Data Wrangler에서 새 데이터는 어떻게 처리되나요?

SageMaker Data Wrangler UI에서 직접 SageMaker 처리 작업을 구성하고 시작할 수 있습니다. 예를 들어 데이터 처리 작업을 예약하고 데이터 소스를 파라미터화하여 새 데이터 배치를 대규모로 손쉽게 변환할 수 있습니다.

### Q: SageMaker Data Wrangler는 CI/CD 프로세스에서 어떻게 작동하나요?

데이터가 준비된 후에는 SageMaker Data Wrangler에서 다양한 옵션을 사용하여 SageMaker Data Wrangler 흐름을 프로덕션으로 승격하고 MLOps 및 CI/CD 기능과 원활하게 통합할 수 있습니다. SageMaker Data Wrangler UI에서 직접 SageMaker 처리 작업을 구성하고 시작할 수 있습니다. 예를 들어 데이터 처리 작업을 예약하고 데이터 소스를 파라미터화하여 새 데이터 배치를 대규모로 손쉽게 변환할 수 있습니다. SageMaker Data Wrangler는 SageMaker 처리 및 SageMaker Spark 컨테이너와 원활하게 통합되므로 간편하게 SageMaker SDK를 사용하여 SageMaker Data Wrangler를 프로덕션 워크플로에 통합할 수도 있습니다.

### Q: SageMaker Data Wrangler Quick Model에는 어떤 모델이 사용되나요?

SageMaker Data Wrangler에서 몇 단계를 수행하여 기본 하이퍼파라미터로 XGBoost 모델을 분할하고 훈련할 수 있습니다. SageMaker Data Wrangler는 문제 유형에 따라 모델 요약, 특성 요약 및 혼동 행렬을 제공하므로 이러한 인사이트를 빠르게 확인하여 데이터 준비 흐름을 반복할 수 있습니다.

### Q: SageMaker Data Wrangler가 지원하는 데이터 크기는 얼마인가요?

SageMaker Data Wrangler는 데이터 가져오기에 상위 K 샘플링, 무작위 샘플링 및 층화 샘플링 등 다양한 샘플링 기법을 지원합니다. 따라서 SageMaker Data Wrangler의 UI를 사용하여 데이터를 빠르게 변환할 수 있습니다. 크거나 넓은 데이터 세트를 사용하는 경우 SageMaker Data Wrangler 인스턴스 크기를 늘려 성능을 개선할 수 있습니다. 흐름을 생성한 후에는 SageMaker Data Wrangler 처리 작업을 사용하여 전체 데이터 세트를 처리할 수 있습니다.

#### Q: SageMaker Data Wrangler는 SageMaker 특성 저장소와 어떻게 연동되나요?

SageMaker 특성 저장소를 SageMaker Data Wrangler에서 준비한 특성의 대상으로 구성할 수 있습니다. UI에서 직접 이 작업을 수행하거나 SageMaker 특성 저장소를 데이터 처리 대상으로 사용하기 위해 특별히 생성된 노트북을 내보낼 수 있습니다.

#### Q: 기계 학습 모델에 대한 특성을 저장하려면 어떻게 해야 하나요?

[SageMaker 특성 저장소](#)는 읽기 및 쓰기 지연 시간이 짧은(밀리초) 데이터 특성을 위한 중앙 리포지토리를 제공합니다. SageMaker 특성 저장소를 통해 특성을 저장, 검색, 검색 및 공유할 수 있으므로 보안 액세스 및 제어 기능을 통해 여러 모델과 팀에서 쉽게 재사용할 수 있습니다. SageMaker 특성 저장소는 배치 또는 스트리밍 파이프라인을 통해 생성된 온라인 및 오프라인 특성을 모두 지원합니다. 그리고 특성 다시 채우기를 지원하며, 모델 훈련 및 추론에 사용된 특성 간의 패리티 유지를 위한 온라인 및 오프라인 스토어 모두를 제공합니다.

#### Q: 온라인 및 오프라인 특성 사이의 일관성을 유지하려면 어떻게 해야 하나요?

SageMaker 특성 저장소는 추가 관리나 코드 없이 온라인 특성과 오프라인 특성의 일관성을 자동으로 유지합니다. SageMaker 특성 저장소는 훈련 및 추론 환경에서 일관성을 유지하는 완전관리형 기능입니다.

#### Q: 지정된 순간에서 특성을 재현하려면 어떻게 해야 하나요?

SageMaker 특성 저장소는 모든 순간에 모든 특성에 대한 타임스탬프를 유지합니다. 따라서 특정 기간의 특성을 검색하여 비즈니스 또는 규정 준수 요구 사항을 충족할 수 있습니다. 지정된 순간에 모델을 재현하여 처음 특성을 생성한 시점부터 현재까지 모델 특성과 값을 쉽게 설명할 수 있습니다.

#### Q: 오프라인 특성이란 무엇인가요?

오프라인 피처는 오랫동안 매우 큰 볼륨에 액세스해야 하는 경우 훈련에 사용됩니다. 이러한 특성은 처리량이 높고 대역폭이 큰 리포지토리에서 지원됩니다.

#### Q: 온라인 특성이란 무엇인가요?

온라인 특성은 실시간 예측을 수행해야 하는 애플리케이션에 사용됩니다. 온라인 특성은 지연 시간이 10밀리초 미만인 고처리량 리포지토리에서 제공되어 빠른 예측을 지원합니다.

#### Q: SageMaker 특성 저장소 요금은 어떻게 부과되나요?

[AWS 프리 티어](#)의 일부로 SageMaker 특성 저장소를 무료로 시작할 수 있습니다. SageMaker 특성 저장소를 사용할 때는 특성 저장소에 대한 쓰기 작업 요금과 온라인 특성 저장소의 읽기 및 저장 작업에 대한 요금이 부과됩니다. 자세한 요금은 [Amazon SageMaker 요금](#)을 참조하세요.

#### Q: SageMaker는 데이터 레이블링을 위해 어떤 기능을 제공하나요?

SageMaker는 Amazon SageMaker Ground Truth Plus 및 Amazon SageMaker Ground Truth라는 2가지 데이터 레이블링 기능을 제공합니다. 두 옵션 모두 이미지, 텍스트 파일 및 비디오와 같은 원시 데이터를 식별하고 정보 레이블을 추가하여 ML 모델을 위한 고품질 훈련 데이터 세트를 생성할 수 있습니다. 자세히 알아보려면 [Amazon SageMaker 데이터 레이블링](#)을 참조하세요.

#### Q: 지리 공간 데이터란 무엇인가요?

지리 공간 데이터는 지구 표면의 형상이나 물체를 나타냅니다. 지리 공간 데이터의 첫 번째 유형은 벡터 데이터로, 점, 선 또는 다각형과 같은 2차원 기하학 구조를 사용하여 도로와 토지 경계 같은 객체를 나타냅니다. 지리 공간 데이터의 두 번째 유형은 위성, 고소 플랫폼 또는 원격 감지 데이터에서 캡처된 이미지와 같은 래스터 데이터입니다. 이 데이터 유형은 픽셀 매트릭스를 사용하여 형상이 있는 위치를 정의합니다. 래스터 형식은 다양한 데이터를 저장하는 데 사용할 수 있습니다. 지리 공간 데이터의 세 번째 유형은 지오 태깅 위치 데이터입니다. 여기에는 관심 장소(예: 에펠탑), 위치가 태깅된 소셜 미디어 게시물, 위도 및 경도 좌표 또는 다양한 스타일과 형식의 상세 주소가 포함됩니다.

#### Q: SageMaker 지리 공간 기능에는 어떤 것이 있나요?

SageMaker 지리 공간 기능을 사용하면 지리 공간 데이터를 사용하여 예측용 ML 모델을 손쉽게 구축, 훈련 및 배포할 수 있습니다. Amazon S3에서 Planet Labs 위성 데이터와 같은 자체 데이터를 가져오거나, Open Data on AWS, Amazon Location Service 및 기타 SageMaker 지리 공간 데이터 소스에서 데이터를 가져올 수 있습니다.

**Q: SageMaker의 지리 공간 ML을 사용해야 하는 이유는 무엇인가요?**

SageMaker 지리 공간 기능을 사용하면 DIY 솔루션보다 빠르게 지리 공간 데이터 기반 예측을 수행할 수 있습니다. SageMaker 지리 공간 기능을 사용하면 기존 고객 데이터 레이크, 오픈 소스 데이터 세트 및 기타 SageMaker 지리 공간 데이터 소스의 지리 공간 데이터에 보다 손쉽게 액세스할 수 있습니다. SageMaker 지리 공간 기능은 효율적인 데이터 준비, 모델 훈련 및 추론을 지원하는 목적별 알고리즘을 제공하므로, 맞춤형 인프라와 데이터 전처리 기능을 구축해야 하는 필요성이 최소화됩니다. 또한 SageMaker Studio에서 맞춤형 시각화와 데이터를 생성하고 조직에서 공유할 수 있습니다. SageMaker 지리 공간 기능에는 농업, 부동산, 보험 및 금융 서비스 분야에서 일반적인 용도로 사용 가능한 사전 훈련된 모델이 포함되어 있습니다.

## 모델 구축

**Q: Amazon SageMaker Studio 노트북이란 무엇인가요?**

SageMaker Studio 노트북은 협업을 위해 빠르게 시작할 수 있는 관리형 Jupyter Notebook입니다. SageMaker Studio 노트북은 SageMaker 및 기타 AWS 서비스의 목적별 ML 도구와 통합되며 ML용 완전 IDE인 SageMaker Studio에서의 모든 ML 개발을 지원합니다.

**Q: SageMaker Studio 노트북은 인스턴스 기반 노트북 제품 및 서비스와 어떻게 다른가요?**

SageMaker Studio 노트북은 인스턴스 기반 노트북과 차별화된 몇 가지 중요한 기능을 제공합니다. Studio 노트북을 사용하면 인스턴스를 수동으로 프로비저닝하거나 운영 상태가 될 때까지 기다리지 않고도 빠르게 노트북을 시작할 수 있습니다. 노트북을 읽고 실행하기 위해 UI를 시작하는 시작 시간도 인스턴스 기반 노트북보다 빠릅니다.

또한 언제라도 UI 내의 다양한 인스턴스 유형 모음에서 탄력적으로 인스턴스를 고를 수 있습니다. 노트북에서 새 인스턴스와 포트를 시작하기 위해 AWS 관리 콘솔로 이동할 필요가 없습니다.

사용자마다 특정 인스턴스와 독립되어 있는 격리된 홈 디렉터리가 있습니다. 이 디렉터리는 시작될 때 모든 노트북 서버와 커널에 자동으로 탑재되므로, 인스턴스를 전환해서 노트북을 보고 실행하는 경우에도 노트북과 다른 파일에 액세스할 수 있습니다.

SageMaker Studio 노트북은 AWS IAM Identity Center(AWS SSO의 후속 모델)와 통합되며 조직 보안 인증 정보를 통해 노트북에 쉽게 액세스할 수 있습니다. 노트북 공유는 SageMaker Studio 노트북에 통합된 기능입니다. 한 단계로 동료와 노트북을 공유하거나 단일 노트북을 동시에 공동으로 편집할 수도 있습니다.

**Q: SageMaker Studio 노트북은 어떤 방식으로 작동하나요?**

SageMaker Studio 노트북은 한 단계로 빠르게 가동할 수 있는 Jupyter Notebook입니다. 기본 컴퓨팅 리소스가 완전히 탄력적이므로 사용 가능한 리소스를 쉽게 확장하거나 축소할 수 있으며 변경 작업이 백그라운드에서 자동으로 진행되므로 작업에 방해가 되지 않습니다. 또한 SageMaker에서 한 단계로 노트북을 공유할 수 있습니다. 노트북을 다른 사람과 쉽게 공유하여 같은 장소에 저장된 정확하게 동일한 노트북을 제공할 수 있습니다.

SageMaker Studio 노트북을 사용할 때는 IAM Identity Center를 사용하여 사내 보안 인증 정보로 로그인할 수 있습니다. 노트북을 공유하면 해당 노트북으로 캡슐화되는 작업 이미지에서 노트북을 실행하는 데 필요한 종속성이 자동으로 추적되므로 팀 내에서는 물론, 팀 간에도 노트북을 손쉽게 공유할 수 있습니다.

**Q: SageMaker에서 공유 공간이란 무엇인가요?**

ML 실무자는 팀원이 SageMaker Studio 노트북을 함께 읽고 편집할 수 있는 작업 공간을 만들 수 있습니다. 팀원들은 공유 공간을 사용하여 동일한 노트북 파일을 공동 편집하고, 노트북 코드를 동시에 실행하며, 결과를 함께 검토하여 반복 작업을 줄이고 협업을 간소화할 수 있습니다. 공유 공간에서는 BitBucket 및 AWS CodeCommit과 같은 서비스가 기본적으로 지원되기 때문에 여러 버전의 노트북을 손쉽게 관리하고 변경 사항을 지속적으로 비교할 수 있습니다. 실험 및 ML 모델과 같이 노트북 내에서 생성되는 모든 리소스가 자동으로 저장되고, 생성된 특정 작업 공간에 연결되므로 팀 작업을 보다 체계적으로 유지하고 ML 모델 개발을 가속화할 수 있습니다.

**Q: SageMaker Studio 노트북을 다른 AWS 서비스와 함께 사용할 수 있나요?**

SageMaker Studio 노트북에서 분산 훈련, 배치 변환, 호스팅, 실험 관리 같은 모든 SageMaker 기능을 활용할 수 있습니다. SageMaker 노트북에서 Amazon S3의 데이터 세트, Amazon Redshift, AWS Glue, Amazon EMR 또는 AWS Lake Formation 같은 다른 서비스에 액세스할 수 있습니다.

**Q: SageMaker Studio 노트북 요금은 어떻게 적용되니까?**

SageMaker Studio 노트북을 사용할 때는 컴퓨팅과 스토리지 비용을 모두 지불해야 합니다. 컴퓨팅 인스턴스 유형별 요금은 [Amazon SageMaker 요금](#)을 참조하세요. 노트북과 데이터 파일 및 스크립트와 같은 관련 아티팩트는 Amazon Elastic File System(Amazon EFS)에 보존됩니다. 스토리지 요금은 [Amazon EFS 요금](#)을 참조하세요. [AWS 프리 티어](#)를 일부로 SageMaker Studio 노트북을 무료로 시작할 수 있습니다.

**Q: SageMaker Studio에서 만들고 실행하는 각 노트북에 대해 별도의 요금이 부과되나요?**



아니요. 같은 컴퓨팅 인스턴스에 여러 노트북을 생성하고 실행할 수 있습니다. 개별 항목이 아닌 사용하는 컴퓨팅에 대해서만 지불합니다. [측정 가이드](#)에서 자세한 내용을 읽어보실 수 있습니다.

노트북 외에도 SageMaker Studio에서 터미널과 대화형 셸을 시작하고 실행할 수 있습니다. 모두 동일한 컴퓨팅 인스턴스에 위치합니다. 각 애플리케이션은 컨테이너나 이미지 내에서 실행됩니다. SageMaker Studio는 데이터 과학 및 기계 학습용으로 특별히 구축되고 사전 구성된 여러 가지 내장 이미지를 제공합니다. [SageMaker Studio 노트북](#)에 대한 가이드에서 SageMaker Studio 개발자 환경에 대해 자세히 읽어볼 수 있습니다.

#### Q: 어떻게 내 노트북에서 사용하는 리소스를 모니터링하고 종료할 수 있습니까?

SageMaker Studio 비주얼 인터페이스 및 AWS 관리 콘솔을 통해 SageMaker Studio 노트북에서 사용하는 리소스를 모니터링하고 [종료](#)할 수 있습니다. 자세한 내용은 [설명서](#)를 참조하세요.

#### Q: SageMaker Studio 노트북을 실행하고 있습니다. 브라우저를 닫거나, 노트북 탭을 닫거나, 브라우저를 열어두어도 요금이 부과됩니까?

예. 컴퓨팅 요금은 계속해서 부과됩니다. 이는 AWS 관리 콘솔에서 Amazon EC2 인스턴스를 시작한 다음 브라우저를 닫는 것과 유사합니다. Amazon EC2 인스턴스가 계속해서 실행되고 있으므로 인스턴스를 명시적으로 종료하지 않는 한 여전히 요금이 부과됩니다.

#### Q: SageMaker Studio 도메인을 만들고 설정하는 데 요금이 부과되나요?

아니요. 사용자 프로필 추가, 업데이트, 삭제를 포함하여 SageMaker Studio 도메인을 만들거나 구성하는 데는 요금이 부과되지 않습니다.

#### Q: SageMaker Studio 노트북 또는 기타 SageMaker 서비스에 대한 항목별 요금은 어떻게 확인하나요?

관리자는 AWS 빌링 콘솔에서 SageMaker Studio를 포함하여 SageMaker에 대한 항목별 요금 목록을 확인할 수 있습니다. SageMaker용 AWS Management Console의 상단 메뉴에서 **Services(서비스)**를 선택하고, 검색 상자에 **'billing(결제)'**를 입력하고, 드롭다운 메뉴에서 결제를 선택한 다음 왼쪽 패널에서 **Bills(청구서)**를 선택합니다. Details(세부 정보) 섹션에서 **SageMaker**를 선택하여 리전 목록을 확장하고 항목별 요금으로 드릴다운할 수 있습니다.

#### Q: Amazon SageMaker Studio Lab이란 무엇인가요?

SageMaker Studio Lab은 누구나 ML을 배우고 실험할 수 있도록 컴퓨팅, 스토리지(최대 15GB) 및 보안을 모두 무료로 제공하는 무료 기계 학습 개발 환경입니다. 시작하려면 유효한 이메일 ID만 있으면 됩니다. 인프라를 구성하거나 ID 및 액세스를 관리하거나 AWS 계정에 가입할 필요가 없습니다. SageMaker Studio Lab은 GitHub 통합을 통해 모델 구축을 가속화하고 가장 인기 있는 ML 도구, 프레임워크 및 라이브러리로 사전 구성되어 제공되므로 즉시 시작할 수 있습니다. SageMaker Studio Lab은 작업을 자동으로 저장하므로 세션 사이에 다시 시작할 필요가 없습니다. 노트북을 닫았다가 나중에 다시 열고 작업하는 것처럼 쉽습니다.

#### Q: SageMaker Studio Lab을 사용해야 하는 이유는 무엇인가요?

SageMaker Studio Lab은 설정 없이 기계 학습 수업 및 실험에 사용할 수 있는 무료 노트북 개발 환경이 필요한 학생, 연구원 및 데이터 사이언티스트를 위한 것입니다. SageMaker Studio Lab은 프로덕션 환경이 필요하지는 않지만 기계 학습 기술 향상을 위해 SageMaker 기능의 일부를 원하는 사용자에게 이상적입니다. SageMaker 세션은 자동으로 저장되므로 각 사용자 세션에서 중단한 부분부터 다시 시작할 수 있습니다.

#### Q: SageMaker Studio Lab은 다른 AWS 서비스에서 어떻게 작동하나요?

SageMaker Studio Lab은 AWS에 구축된 서비스로, Amazon SageMaker Studio와 동일한 핵심 서비스(예: Amazon S3 및 Amazon EC2)를 다수 사용합니다. 다른 서비스와 달리 AWS 계정이 필요하지 않습니다. 대신 이메일 주소를 사용하여 SageMaker Studio Lab 전용 계정을 생성하게 됩니다. 이 전용 계정으로 ML 노트북을 실행할 수 있는 제한된 환경(15GB 스토리지 및 12시간 세션)에 액세스할 수 있습니다.

#### Q: SageMaker Canvas란 무엇인가요?

SageMaker Canvas는 비즈니스 분석가가 코드를 작성하거나 기계 학습 전문 지식 없이도 기계 학습 모델을 구축하고 정확한 예측을 생성할 수 있는, 끌어서 놓기 방식의 시각적 서비스입니다. SageMaker Canvas를 사용하면 다양한 소스의 데이터에 쉽게 액세스 및 결합하고, 데이터를 자동으로 정리하고 다양한 데이터 조정을 적용하고, 기계 학습 모델을 구축하여 한 단계로 정확한 예측을 생성할 수 있습니다. 또한 간편하게 결과를 게시하고, 모델을 설명 및 해석하고, 검토를 위해 조직 내의 다른 사람들과 모델을 공유할 수 있습니다.

#### Q: SageMaker Canvas는 어떤 데이터 소스를 지원하나요?

SageMaker Canvas는 Amazon S3 및 Amazon Redshift를 포함하여 계정에서 액세스할 수 있는 AWS 데이터 소스를 원활하게 검색하는 데 도움이 됩니다. SageMaker Canvas의 시각적 드래그 앤 드롭 인터페이스를 사용하여 데이터를 찾아보고 가져올 수 있습니다. 또한 로컬 디스크에서 파일을 끌어서 놓을 수 있으며 사전 구축된 커넥터를 사용하여 Snowflake와 같은 서드 파티 소스에서 데이터를 가져올 수 있습니다.

#### Q: SageMaker Canvas에서 정확한 예측을 생성하기 위해 기계 학습 모델을 구축하려면 어떻게 해야 하나요?

소스를 연결하고 데이터 세트를 선택하고 데이터를 준비한 후 예측할 대상 열을 선택하여 모델 생성 작업을 시작할 수 있습니다. SageMaker Canvas는 자동으로 문제 유형을 식별하고, 새로운 관련 기능을 생성하고, 선형 회귀, 로지스틱 회귀, 딥 러닝, 시계열 예측, 그라데이션 부스팅 등의 기계 학습 기술을 사용하여 포괄적인 예측 모델 세트를 테스트하며, 데이터 세트를 기반으로 정확한 예측을 수행하는 모델을 구축합니다.

**Q: SageMaker Canvas에서 모델을 구축할 때 소요되는 시간은 어느 정도인가요? 모델 생성 중 진행 상황을 모니터링하려면 어떻게 해야 하나요?**

모델을 구축하는 데 걸리는 시간은 데이터 집합의 크기에 따라 다릅니다. 작은 데이터 세트는 30분 미만이 소요될 수 있고 큰 데이터 세트는 몇 시간이 걸릴 수 있습니다. 모델 생성 작업이 진행됨에 따라 작업 완료율과 작업 완료까지 남은 시간을 포함하여 자세한 시각적 업데이트가 SageMaker Canvas에 표시됩니다.

## 모델 훈련

**Q: Amazon SageMaker Experiments란 무엇인가요?**

SageMaker Experiments를 사용하면 기계 학습 모델에 대한 반복을 구성하고 추적할 수 있습니다. SageMaker Experiments는 입력 파라미터, 구성 및 결과를 자동으로 캡처하고 이를 '실험'으로 저장하여 반복을 관리하도록 돕습니다. SageMaker Studio의 시각적 인터페이스에서 작업하여 활성 실험을 탐색하고 특성별로 이전 실험을 검색하며 결과와 함께 이전 실험을 검토하고 실험 결과를 시각적으로 비교할 수 있습니다.

**Q: Amazon SageMaker Debugger란 무엇인가요?**

SageMaker Debugger는 훈련 도중 혼동 지표 및 학습 경사와 같은 실시간 지표를 자동으로 캡처하여 모델 정확도를 개선합니다. SageMaker Debugger의 지표를 SageMaker Studio에서 시각화하면 지표를 더 쉽게 이해할 수 있습니다. SageMaker Debugger에서 일반적인 훈련 문제가 감지될 때 경고 및 해결 방법을 생성할 수도 있습니다. 또한 SageMaker Debugger는 CPU, GPU, 네트워크 및 메모리와 같은 시스템 리소스를 실시간으로 자동으로 모니터링하고 프로파일링하며, 이러한 리소스의 재할당에 대한 권장 사항을 제공합니다. 이 기능은 훈련 중에 리소스를 효율적으로 사용하고 비용과 리소스를 절감하는 데 도움이 됩니다.

**Q: SageMaker에서는 분산 훈련을 지원하나요?**

예. SageMaker는 AWS GPU 인스턴스에 걸쳐 딥 러닝 모델과 대규모 훈련 세트를 자동으로 분산할 수 있습니다. 이때 소요 시간은 이러한 분산 전략을 수동으로 빌드하고 최적화하는 데 걸리는 시간에 비해 몇 분의 일에 불과합니다. SageMaker에서 적용하는 2가지 분산 훈련 기술로는 데이터 병렬 처리와 모델 병렬 처리가 있습니다. 데이터 병렬 처리는 여러 GPU 인스턴스에서 데이터를 균등하게 나누어 훈련 속도를 개선하기 위해 적용됩니다. 이 경우 각 인스턴스를 동시에 훈련할 수 있습니다. 모델 병렬 처리는 모델이 너무 커서 단일 GPU에 저장할 수 없는 경우에 유용하며, 여러 GPU에 분산하기 전에 모델을 더 작은 부분으로 파티셔닝해야 합니다. PyTorch 및 TensorFlow 훈련 스크립트에서 추가 코드 몇 줄만 작성하면, SageMaker는 데이터 병렬 처리 또는 모델 병렬 처리를 자동으로 적용하여 모델을 더 빠르게 개발하고 배포할 수 있습니다. SageMaker는 각 GPU 계산의 균형을 맞추기 위해 그래프 분할 알고리즘을 사용하여 모델을 분할하는 최상의 접근 방법을 결정하는 동시에 GPU 인스턴스 간의 통신을 최소화합니다. 또한 SageMaker는 선형에 가까운 확장 효율성을 달성하기 위해 AWS 컴퓨팅과 네트워크를 충분히 활용하는 알고리즘을 통해 분산 훈련 작업을 최적화하므로 수동 오픈 소스 구현보다 더욱 빠르게 훈련을 완료할 수 있습니다.

**Q: Amazon SageMaker Training Compiler란 무엇인가요?**

SageMaker Training Compiler는 GPU를 보다 효율적으로 사용하기 위해 그래프 및 커널 수준 최적화를 통해 딥 러닝(DL) 모델 훈련을 최대 50% 가속화하는 DL 컴파일러입니다. SageMaker Training Compiler는 SageMaker의 TensorFlow 및 PyTorch 버전과 통합되므로 최소한의 코드 변경으로 이러한 인기 있는 프레임워크에서 훈련 속도를 높일 수 있습니다.

**Q: SageMaker Training Compiler는 어떻게 작동하나요?**

SageMaker Training Compiler는 고급 언어 표현의 DL 모델을 기본 프레임워크를 사용하는 작업보다 빠르게 훈련하는 하드웨어 최적화 지침으로 변환하여 훈련 작업을 가속화합니다. 보다 구체적으로 SageMaker Training Compiler는 그래프 수준 최적화(연산자 융합, 메모리 계획 및 대수 단순화), 데이터 흐름 수준 최적화(레이아웃 변환, 공통 하위 표현식 제거) 및 백엔드 최적화(메모리 대기 시간 숨기기, 루프 지향적 최적화)를 통해 하드웨어 리소스를 보다 효율적으로 사용하고 결과적으로 더 빠르게 훈련하는 최적화된 모델 훈련 작업을 생성합니다.

**Q: SageMaker Training Compiler는 어떻게 사용하나요?**

SageMaker Training Compiler는 SageMaker Python SDK 및 SageMaker Hugging Face Deep Learning Containers에 내장되어 있습니다. 속도 향상 이점을 활용하기 위해 워크플로를 변경할 필요가 없습니다. SageMaker 노트북 인스턴스, SageMaker Studio, AWS SDK for Python(Boto3) 및 AWS Command Line Interface(AWS CLI)와 같은 SageMaker 인터페이스를 사용하여 기존과 동일한 방식으로 훈련 작업을 실행할 수 있습니다. 프레임워크 추정기 객체를 생성할 때 TrainingCompilerConfig 클래스를 파라미터로 추가하여 SageMaker Training Compiler를 사용할 수 있습니다. 실제로 이는 단일 GPU 인스턴스에 대한 기존 훈련 작업 스크립트에 몇 줄의 코드가 추가되었음을 의미합니다. 최신 세부 설명서, 샘플 노트북 및 예제는 [설명서](#)에서 사용할 수 있습니다.

**Q: SageMaker Training Compiler의 요금은 얼마인가요?**

SageMaker Training Compiler는 SageMaker 훈련 기능이며 SageMaker 고객에게만 추가 비용 없이 제공됩니다. SageMaker Training Compiler를 사용하면 훈련 시간이 단축되므로 실제로 비용을 절감할 수 있습니다.

**Q: 관리형 스팟 훈련이란 무엇인가요?**

SageMaker를 통해 관리형 스팟 훈련을 수행할 경우 Amazon EC2 스팟 인스턴스를 사용하여 기계 학습 모델을 훈련할 수 있으며 모델 훈련 비용을 최대 90%나 줄일 수 있습니다.

**Q: 관리형 스팟 훈련을 사용하려면 어떻게 해야 하나요?**

훈련 작업을 제출할 때 관리형 스팟 훈련 옵션을 사용하도록 설정하고 스팟 용량 지연 시간도 지정합니다. 그러면 SageMaker가 Amazon EC2 스팟 인스턴스를 사용하여 작업을 실행하고 스팟 용량을 관리합니다. 작업이 실행 중이거나 용량 대기 중일 때 모두 훈련 작업 상태를 완전히 파악할 수 있습니다.

**Q: 관리형 스팟 훈련은 언제 이용해야 합니까?**

관리형 스팟 훈련은 훈련 실행에 유연성이 있고 훈련 작업 비용을 최소화하려는 경우에 이상적입니다. 관리형 스팟 훈련을 사용할 경우 기계 학습 모델에 대한 훈련 비용을 최대 90%까지 절감할 수 있습니다.

**Q: 관리형 스팟 훈련은 어떻게 작동하나요?**

관리형 스팟 훈련은 훈련에 Amazon EC2 스팟 인스턴스를 사용하며 스팟 인스턴스는 AWS에 용량이 필요한 경우 대체될 수 있습니다. 따라서 관리형 스팟 훈련 작업은 사용 가능한 용량이 생길 때마다 조금씩 실행될 수 있습니다. SageMaker는 최신 모델 체크포인트를 사용하여 훈련 작업을 재개할 수 있으므로 중단이 발생한 경우 훈련 작업을 처음부터 다시 시작할 필요가 없습니다. SageMaker의 기본 제공 프레임워크 및 기본 제공 컴퓨터 비전 알고리즘을 통해 정기적 체크포인트를 사용할 수 있으며 사용자 지정 모델에서 체크포인트를 사용할 수도 있습니다.

**Q: 관리형 스팟 훈련에 대해 정기적으로 체크포인트를 점검해야 합니까?**

장기적인 훈련 작업을 위한 일반적인 모범 사례로서 주기적으로 체크포인트 점검을 수행할 것을 권장합니다. 이렇게 하면 용량이 대체된 경우 관리형 스팟 훈련이 다시 시작되는 것을 방지할 수 있습니다. 체크포인트를 사용하도록 설정하면 SageMaker가 마지막 체크포인트에서 관리형 스팟 훈련 작업을 재개합니다.

**Q: 관리형 스팟 훈련 작업을 통한 비용 절감액을 계산하려면 어떻게 해야 하나요?**

관리형 스팟 훈련 작업이 완료되면 AWS 관리 콘솔에서 절감액을 확인할 수 있고 훈련 작업 기간과 청구 대상 기간 사이의 백분율 차이로도 비용 절감액을 계산할 수 있습니다.

관리형 스팟 훈련 작업의 중단 횟수와 관계없이 데이터를 다운로드한 기간에 대해서 한 번만 비용이 청구됩니다.

**Q: 관리형 스팟 훈련에 사용할 수 있는 인스턴스는 무엇인가요?**

관리형 스팟 훈련은 SageMaker에서 지원되는 모든 인스턴스에서 사용할 수 있습니다.

**Q: 관리형 스팟 훈련은 어느 리전에서 지원되나요?**

관리형 스팟 훈련은 현재 SageMaker가 [제공](#)되는 모든 리전에서 지원됩니다.

**Q: 훈련에 사용할 수 있는 데이터 세트의 크기에 제한이 있나요?**

SageMaker에서 모델 훈련에 사용할 수 있는 데이터 세트의 크기 제한은 없습니다.

**Q: SageMaker에서는 모델을 생성할 때 어떤 알고리즘을 사용하나요?**

SageMaker에는 선형 회귀, 로지스틱 회귀, k-평균 클러스터링, 주성분 분석, 인수 분해 머신, 신경 주제 모델링, LDA(Latent Dirichlet Allocation), 그래디언트 부스 티드 트리, sequence2sequence, 시계열 예측, word2vec 및 이미지 분류에 대한 알고리즘이 내장되어 있습니다. SageMaker는 최적화된 Apache MXNet, Tensorflow, Chainer, PyTorch, Gluon, Keras, Horovod, Scikit-learn 및 Deep Graph Library 컨테이너도 제공합니다. 그뿐 아니라 SageMaker는 문서화된 사양을 준수하는 도커 이미지를 통해 제공된 사용자 지정 훈련 알고리즘을 지원합니다.

**Q: 자동 모델 튜닝이란 무엇인가요?**

대부분 기계 학습 알고리즘에서는 기본 알고리즘이 작동하는 방식을 제어하는 다양한 파라미터를 공개합니다. 이러한 파라미터는 일반적으로 하이퍼파라미터라고 부르며 그 값은 훈련된 모델의 품질에 영향을 미칩니다. 자동 모델 튜닝은 알고리즘에서 최적의 모델을 산출할 수 있는 하이퍼파라미터 세트를 찾는 프로세스입니다.

**Q: 자동 모델 튜닝으로 어떤 모델을 튜닝할 수 있나요?**

과학적으로 실현 가능한 한, 내장된 SageMaker 알고리즘, 딥 신경망 또는 도커 이미지 형식으로 SageMaker에 가져오는 임의 알고리즘을 비롯하여 원하는 알고리즘에 대해 SageMaker에서 자동 모델 튜닝을 실행할 수 있습니다.

**Q: SageMaker 외부에서 자동 모델 튜닝을 사용할 수 있나요?**

현재는 지원되지 않습니다. 최상의 모델 튜닝 성능과 경험은 SageMaker 내에서 제공됩니다.

**Q: 자동 모델 튜닝의 기본 튜닝 알고리즘은 무엇인가요?**

현재 튜닝 하이퍼파라미터에 대한 알고리즘은 베이지안 최적화를 사용자 지정하여 구현한 것입니다. 튜닝 프로세스 전반에 걸쳐 고객이 지정한 목표 지표를 최적화하는 것을 목표로 합니다. 완료된 훈련 작업의 객체 지표를 확인하고 지식을 사용하여 다른 훈련 작업을 위한 하이퍼파라미터 조합을 추론합니다.

**Q: 자동 모델 튜닝에 권장되는 특정 하이퍼파라미터가 있나요?**

아니요. 특정 하이퍼파라미터가 모델 성능에 미치는 영향은 다양한 요소에 따라 다르며 하나의 하이퍼파라미터가 다른 하이퍼파라미터보다 중요하므로 튜닝을 해야 한다고 단정적으로 말하기는 어렵습니다. SageMaker에 내장된 알고리즘의 경우 AWS에서 하이퍼파라미터의 튜닝 가능 여부를 알려드립니다.

**Q: 하이퍼파라미터 튜닝 작업은 얼마나 걸리나요?**

하이퍼파라미터 튜닝 작업에 걸리는 시간은 데이터 크기, 기본 알고리즘, 하이퍼파라미터의 값 등 여러 요소에 따라 달라집니다. 또한 고객은 동시 훈련 작업 수와 총 훈련 작업 수를 선택할 수 있습니다. 이 모든 선택이 하이퍼파라미터 튜닝 작업에 걸리는 시간에 영향을 미칩니다.

**Q: 모델의 속도와 정확도를 최적화하는 것처럼 여러 목표를 동시에 최적화할 수 있습니까?**

지금은 지원되지 않습니다. 현재는 단일 목표 지표를 지정하여 알고리즘 코드를 최적화하거나 변경하여 2개 이상의 유용한 지표 간 가중 평균인 새로운 지표를 생성해야 하며 튜닝 프로세스를 해당 목표 지표에 맞게 최적화해야 합니다.

**Q: 자동 모델 튜닝은 비용이 어떻게 됩니까?**

하이퍼파라미터 튜닝 작업 자체는 무료입니다. [모델 훈련 요금](#)을 기반으로 하이퍼파라미터 튜닝 작업에서 시작하는 훈련 작업에 대한 비용이 부과됩니다.

**Q: SageMaker Autopilot이나 자동 모델 튜닝을 사용하는 상황은 어떻게 판단해야 하나요?**

SageMaker Autopilot은 분류 및 회귀 사용 사례에 중점을 두면서 기능 사전 처리, 알고리즘 선택, 하이퍼파라미터 튜닝 같은 일반적인 기계 학습 워크플로의 모든 것을 자동화합니다. 반면 자동 모델 튜닝은 모델의 기반이 기본 제공 알고리즘이든, 딥 러닝 프레임워크든, 사용자 지정 컨테이너든 관계없이 모든 모델을 튜닝하도록 설계되어 있습니다. 유연성을 높이려면 특정 알고리즘을 직접 선택하고, 튜닝할 하이퍼파라미터와 해당하는 검색 범위를 결정해야 합니다.

**Q: 강화 학습이란 무엇입니까?**

강화 학습은 에이전트가 자신의 작업과 경험의 피드백을 사용하여 시행착오를 통해 대화형 환경에서 학습할 수 있는 기계 학습 기법입니다.

**Q: SageMaker에서 강화 학습 모델을 훈련할 수 있나요?**

예, SageMaker에서 지도 학습 및 비지도 학습에 추가하여 강화 학습 모델을 훈련할 수 있습니다.

**Q: 강화 학습은 지도 학습과 어떻게 다른가요?**

지도 학습과 강화 학습은 입력 및 출력 간의 매핑을 사용하지만, 태스크를 수행하기 위한 올바른 작업 세트가 에이전트에게 피드백으로 제공되는 지도 학습과는 달리, 강화 학습은 일련의 작업을 통해 장기적인 목표를 실현할 수 있도록 보상 신호를 최적화하는 지연된 피드백을 사용합니다.

**Q: 강화 학습은 언제 사용해야 합니까?**

지도 학습 기법의 목적은 훈련 데이터의 패턴을 기반으로 올바른 답변을 찾는 것인 반면, 비지도 학습 기법의 목적은 데이터 포인트 간의 유사점과 차이점을 찾는 것입니다. 반대로, 강화 학습(RL) 기법의 목적은 결과를 달성하는 방법이 확실하지 않더라도 원하는 결과를 달성하는 방법을 학습하는 것입니다. 결과적으로 RL은 로봇 공학, 자율 주행 차량, HVAC, 산업 제어 등과 같이 에이전트가 자율적 결정을 수행할 수 있는 지능형 애플리케이션을 지원하는 데 더 적합합니다.

**Q: RL 모델을 훈련하기 위해 어떤 유형의 환경을 사용할 수 있습니까?**

Amazon SageMaker RL은 RL 모델을 훈련할 수 있는 다양한 환경을 지원합니다. AWS RoboMaker와 같은 AWS 서비스, Open AI Gym 인터페이스를 사용하여 개발된 오픈 소스 환경 또는 사용자 지정 환경, MATLAB 및 SimuLink와 같은 상업용 시뮬레이션 환경을 사용할 수 있습니다.

**Q: RL 모델을 훈련하기 위해 고유의 RL 에이전트 알고리즘을 작성해야 하나요?**

아니요. SageMaker RL에는 DQN, PPO, A3C 등과 같은 RL 에이전트 알고리즘의 구현을 제공하는 Coach 및 Ray RLlib와 같은 RL 도구 키트가 포함되어 있습니다.

**Q: 고유의 RL 라이브러리 및 알고리즘 구현을 가져와서 SageMaker RL에서 실행할 수 있나요?**

예. 고유의 RL 라이브러리 및 알고리즘 구현을 도커 컨테이너로 가져온 다음 SageMaker RL에서 실행할 수 있습니다.

**Q: SageMaker RL을 사용하여 배포된 롤아웃을 수행할 수 있나요?**

예. 다른 유형의 클러스터를 선택할 수도 있습니다. 이 경우 한 GPU 인스턴스에서 훈련을 실행하고 여러 CPU 인스턴스에서 시뮬레이션을 실행할 수 있습니다.

## 모델 배포

**Q: SageMaker에서 제공하는 배포 옵션은 무엇인가요?**

모델을 구축 및 훈련한 후 SageMaker에서는 3가지 옵션으로 모델을 배포하고 예측을 시작할 수 있습니다. 실시간 추론은 밀리초 단위의 지연 시간 요구 사항, 최대 6MB의 페이로드 크기, 최대 60초의 처리 시간을 요구하는 워크로드에 적합합니다. 배치 변환은 사전에 제공되는 대규모 데이터 배치에 대한 오프라인 예측에 이상적입니다. Asynchronous Inference는 1초 미만의 대기 시간 요구 사항, 최대 1GB의 페이로드 크기, 최대 15분의 처리 시간이 필요하지 않은 워크로드를 위해 설계되었습니다.

**Q: Amazon SageMaker 비동기 추론이란 무엇인가요?**

SageMaker 비동기 추론은 수신 요청을 대기열에 넣고 비동기식으로 처리합니다. 이 옵션은 페이로드 크기가 크고 처리 시간이 길어서 도착 즉시 처리되어야 하는 요청에 이상적입니다. 선택적으로 자동 크기 조정 설정을 구성하여 요청을 바쁘게 처리하지 않을 때 인스턴스 수를 0개로 축소하여 비용을 절감할 수 있습니다.

**Q: 요청을 능동적으로 처리하지 않을 때 인스턴스 수를 0으로 스케일 다운하도록 자동 크기 조정 설정을 구성하려면 어떻게 해야 하나요?**

SageMaker 비동기 추론 엔드포인트 인스턴스 수를 0개로 스케일 다운하면 요청을 능동적으로 처리하지 않을 때 비용을 절감할 수 있습니다. 'ApproximateBacklogPerInstance' 사용자 지정 지표에서 크기를 조정할 크기 조정 정책을 정의하고 'MinCapacity' 값을 0으로 설정해야 합니다. 단계별 지침은 개발자 안내서에서 [비동기 엔드포인트 자동 크기 조정](#) 섹션을 참조하세요.

**Q: Amazon SageMaker 서버리스 추론이란 무엇인가요?**

[SageMaker 서버리스 추론](#)은 기계 학습 모델을 쉽게 배포하고 확장할 수 있도록 특별히 제작된 서버리스 모델 제공 옵션입니다. SageMaker 서버리스 추론 엔드포인트는 컴퓨팅 리소스를 자동으로 시작하고 트래픽에 따라 리소스를 확장 및 축소하므로 인스턴스 유형을 선택하거나 프로비저닝된 용량을 실행하거나 확장을 관리할 필요가 없습니다. 서버리스 추론 엔드포인트에 대한 메모리 요구 사항을 지정할 수도 있습니다. 유휴 기간이 아닌 추론 코드를 실행하는 기간과 처리된 데이터 양에 대해서만 비용을 지불합니다.

**Q: SageMaker 서버리스 추론을 사용해야 하는 이유는 무엇인가요?**

SageMaker 서버리스 추론은 용량을 미리 프로비저닝하고 조정 정책을 관리할 필요가 없기 때문에 개발자 경험이 단순화됩니다. SageMaker 서버리스 추론은 사용 패턴에 따라 몇 초 안에 수십 개에서 수천 개의 추론으로 즉시 확장할 수 있으므로 트래픽이 간헐적이거나 예측할 수 없는 기계 학습 애플리케이션에 이상적입니다. 예를 들어 급여 처리 회사의 챗봇 서비스는 월말에 문의가 증가하지만 나머지 월에는 트래픽이 간헐적입니다. 이러한 상황에서 전체 달에 대한 인스턴스를 프로비저닝하는 것은 유휴 기간의 비용을 지불하게 되므로 비용 효율적이지 않습니다. SageMaker Serverless Inference는 트래픽을 미리 예측하거나 확장 정책을 관리할 필요 없이 즉시 사용 가능한 자동 확장을 제공하여 이러한 유형의 사용 사례를 해결하는 데 도움이 됩니다. 또한 추론 코드를 실행하는 컴퓨팅 시간(밀리초 단위로 청구)과 데이터 처리에 대해서만 비용을 지불하면 트래픽이 간헐적으로 발생하는 워크로드에 대해 비용 효율적인 옵션이 됩니다.

**Q: SageMaker Serverless Inference의 프로비저닝된 동시성이란 무엇인가요?**

프로비저닝된 동시성을 사용하면 지정된 수의 동시 요청에 대해 엔드포인트를 워밍업 상태로 유지하여 예측 가능한 성능과 높은 확장성으로 서버리스 엔드포인트에 모델을 배포할 수 있습니다.

**Q: 프로비저닝된 동시성을 사용해야 하는 이유는 무엇인가요?**

온디맨드 서버리스 엔드포인트의 경우 엔드포인트가 한동안 트래픽을 수신하지 못하다가 갑자기 새 요청을 받게 되면 컴퓨팅 리소스를 가동하여 요청을 처리하는 데 시간이 걸릴 수 있습니다. 이를 콜드 스타트라고 합니다. 동시 요청이 현재 동시 요청 사용량을 초과하는 경우에도 콜드 스타트가 발생할 수 있습니다. 콜드 스타트 시간은 모델 크기, 모델을 다운로드하는 데 걸리는 시간, 컨테이너의 시작 시간에 따라 달라집니다.

지연 시간 프로파일의 가변성을 줄이려면 서버리스 엔드포인트에 프로비저닝된 동시성을 선택적으로 사용할 수 있습니다. 프로비저닝된 동시성을 사용하면 서버리스 엔드포인트가 항상 준비되어 있기 때문에 콜드 스타트 없이 트래픽 급증을 즉시 처리할 수 있습니다.

**Q: 프로비저닝된 동시성 요금은 어떻게 부과되나요?**

온디맨드 서버리스 추론과 마찬가지로 프로비저닝된 동시성을 사용하는 경우 추론 요청을 처리하는 데 사용된 컴퓨팅 파워에 대한 요금이 부과됩니다. 요금은 밀리초 단위로 처리된 데이터 양에 따라 부과됩니다. 또한 프로비저닝된 동시성 사용 요금은 구성된 메모리, 프로비저닝된 기간 및 사용된 동시성 양에 따라 부과됩니다. 자세한 내용은 [Amazon SageMaker 요금](#)을 참조하세요.



**Q: Amazon SageMaker 새도우 테스트란 무엇인가요?**

SageMaker는 프로덕션 릴리스 전에 현재 배포된 모델을 기준으로 새로운 ML 모델의 성능을 테스트하여 새 모델을 평가하는 새도우 테스트를 실행하는 데 도움이 됩니다. SageMaker는 새 모델을 현재 프로덕션 모델과 함께 새도우 모드에 배포하고 프로덕션 트래픽에서 사용자가 지정한 부분을 새 모델에 미러링합니다. 필요한 경우 오프라인 비교를 위해 모델 추론을 로깅합니다. 지연 시간 및 오류율과 같은 주요 성능 지표를 프로덕션 모델과 새도우 모델 간에 비교하여 새 모델을 프로덕션으로 승격할지 여부를 결정하는 데 도움이 되는 라이브 대시보드도 제공합니다.

**Q: 새도우 테스트에 SageMaker를 사용해야 하는 이유는 무엇인가요?**

SageMaker를 사용하면 새도우 변형을 설정하고 모니터링하는 프로세스가 간소화되므로 라이브 프로덕션 트래픽에서 새 ML 모델의 성능을 평가할 수 있습니다. SageMaker를 사용하면 새도우 테스트용으로 인프라를 오케스트레이션할 필요가 없습니다. 새도우 변형으로 미러링되는 트래픽의 비율 및 테스트 기간과 같은 테스트 파라미터를 제어할 수 있기 때문입니다. 따라서 소규모로 시작한 후 모델 성능에 대해 확신이 생기면 추론 요청을 새 모델로 늘려갈 수 있습니다. SageMaker는 주요 지표의 성능 차이를 표시하는 라이브 대시보드를 생성합니다. 따라서 모델 성능을 손쉽게 비교하여 새 모델과 프로덕션 모델 간의 차이를 평가할 수 있습니다.

**Q: Amazon SageMaker Inference Recommender란 무엇인가요?**

[SageMaker Inference Recommender](#)는 성능 벤치마킹을 자동화하고 SageMaker 기계 학습 인스턴스에서 모델 성능을 조정하여 프로덕션에서 ML 모델을 가져오는 데 필요한 시간을 줄여줍니다. 이제 SageMaker Inference Recommender를 사용하여 엔드포인트에 최상의 성능을 제공하고 비용을 최소화하는 모델을 배포할 수 있습니다. 몇 분 만에 SageMaker Inference Recommender를 시작하면서 인스턴스 유형을 선택하고 몇 시간 안에 최적의 엔드포인트 구성에 대한 권장 사항을 얻을 수 있으므로 몇 주에 걸친 수동 테스트 및 조정 시간이 필요 없습니다. SageMaker Inference Recommender를 사용하면 로드 테스트 중에 사용된 SageMaker 기계 학습 인스턴스에 대해서만 비용을 지불하고 추가 요금은 없습니다.

**Q: SageMaker Inference Recommender를 사용해야 하는 이유는 무엇인가요?**

성능을 개선하고 비용을 줄이기 위해 올바른 엔드포인트 구성에 대한 권장 사항이 필요한 경우 SageMaker Inference Recommender를 사용해야 합니다. 이전에는 모델을 배포하려는 데이터 사이언티스트가 올바른 엔드포인트 구성을 선택하기 위해 수동 벤치마크를 실행해야 했습니다. 또한 먼저 모델 및 샘플 페이로드의 리소스 요구 사항을 기반으로 사용 가능한 70개 이상의 인스턴스 유형 중에서 올바른 기계 학습 인스턴스 유형을 선택한 다음 다양한 하드웨어를 설명하도록 모델을 최적화해야 했습니다. 그런 다음 대기 시간 및 처리량 요구 사항이 충족되고 비용이 저렴한지 검증하기 위해 광범위한 로드 테스트를 수행해야 했습니다. SageMaker Inference Recommender는 다음을 용이하게 함으로써 이러한 복잡성을 제거합니다. 1) 인스턴스 권장 사항으로 몇 분 안에 시작합니다. 2) 인스턴스 유형 전반에 걸쳐 로드 테스트를 수행하여 몇 시간 내에 엔드포인트 구성에 대한 권장 사항을 얻습니다. 3) 컨테이너 및 모델 서버 파라미터를 자동으로 조정하고 주어진 인스턴스 유형에 대한 모델 최적화를 수행합니다.

**Q: SageMaker Inference Recommender는 다른 AWS 서비스와 어떻게 작동하나요?**

데이터 사이언티스트는 SageMaker Studio, Python용 AWS SDK(Boto3) 또는 AWS CLI에서 SageMaker Inference Recommender에 액세스할 수 있습니다. 등록된 모델 버전에 대한 SageMaker 모델 레지스트리의 SageMaker Studio 내에서 배포 권장 사항을 얻을 수 있습니다. 데이터 사이언티스트는 SageMaker Studio, AWS SDK 또는 AWS CLI를 통해 권장 사항을 검색하고 필터링할 수 있습니다.

**Q: SageMaker Inference Recommender는 다중 모델 엔드포인트 또는 다중 컨테이너 엔드포인트를 지원할 수 있나요?**

아니요. 현재 엔드포인트당 단일 모델만 지원합니다.

**Q: SageMaker Inference Recommender는 어떤 유형의 엔드포인트를 지원하나요?**

현재는 실시간 엔드포인트만 지원합니다.

**Q: 한 리전에서 SageMaker Inference Recommender를 사용하고 다른 리전에서 벤치마킹할 수 있나요?**

AWS 중국 리전을 제외하고 Amazon SageMaker에서 지원하는 모든 리전이 지원됩니다.

**Q: SageMaker Inference Recommender는 Amazon EC2 Inf1 인스턴스를 지원하나요?**

예. 모든 유형의 컨테이너를 지원합니다. AWS Inferentia 칩을 기반으로 하는 Amazon EC2 Inf1에는 Neuron 컴파일러 또는 Amazon SageMaker Neo를 사용하는 컴파일된 모델 아티팩트가 필요합니다. Inferentia 대상 및 연결된 컨테이너 이미지 URI에 대한 컴파일된 모델이 있으면 SageMaker Inference Recommender를 사용하여 다양한 Inferentia 인스턴스 유형을 벤치마킹할 수 있습니다.

**Q: Amazon SageMaker Model Monitor란 무엇인가요?**

SageMaker Model Monitor를 사용하면 개념 드리프트를 탐지하고 해결할 수 있습니다. SageMaker Model Monitor는 배포된 모델에서 개념 드리프트를 자동으로 감지하고 문제 원인을 파악하는 데 도움이 되는 자세한 알림을 제공합니다. SageMaker에서 훈련된 모든 모델은 SageMaker Studio에서 수집하고 볼 수 있는 주요 지표를 자동으로 생성합니다. SageMaker Studio 내부에서 수집할 데이터, 확인하는 방법, 알림 수신 시기를 구성할 수 있습니다.

**Q: SageMaker를 실행하는 인프라에 액세스할 수 있나요?**

아니요. SageMaker가 사용자 대신 컴퓨팅 인프라를 운영하므로, 상태 확인을 수행하고 보안 패치를 적용하며 그 외 주기적인 유지 관리를 수행합니다. 사용자는 자체 호스팅 환경에 사용자 지정 추론 코드로 학습한 모델 아티팩트를 배포할 수 있습니다.

**Q: SageMaker 모델이 프로덕션에 배포된 후에는 크기와 성능을 어떻게 조정하나요?**

SageMaker 호스팅은 애플리케이션 Auto Scaling을 사용해 애플리케이션에 필요한 성능에 맞춰 자동으로 조정됩니다. 또한 엔드포인트 구성을 수정하여 가동 중단 없이 인스턴스 수와 유형을 수동으로 변경할 수 있습니다.

**Q: SageMaker 프로덕션 환경을 모니터링하려면 어떻게 해야 하나요?**

SageMaker는 성능 지표를 Amazon CloudWatch 지표로 내보내므로, 지표를 추적하고, 경보를 설정하고, 프로덕션 트래픽의 변경 사항에 자동으로 대응할 수 있습니다. 또한 SageMaker는 사용자가 프로덕션 환경을 모니터링하고 문제를 해결할 수 있도록 Amazon CloudWatch Logs에 로그를 기록합니다.

**Q: SageMaker에서 호스팅할 수 있는 모델 유형은 무엇인가요?**

SageMaker는 추론 도커 이미지의 문서화된 사양을 준수하는 모든 모델을 호스팅할 수 있습니다. SageMaker 모델 아티팩트 및 추론 코드에서 생성된 모델이 이에 포함됩니다.

**Q: SageMaker에서는 동시에 몇 개의 실시간 API 요청을 처리할 수 있나요?**

SageMaker는 많은 수의 초당 트랜잭션으로 확장할 수 있도록 설계되었습니다. 정확한 수는 배포된 모델과 모델이 배포된 인스턴스의 수 및 유형에 따라 달라집니다.

**Q: 배치 변환이란 무엇입니까?**

배치 변환은 대규모 또는 소규모 배치 데이터에 대한 예측을 실행할 수 있게 해 줍니다. 이제 데이터 집합을 여러 데이터 청크로 분할하고 실시간 엔드포인트를 관리할 필요가 없습니다. 간단한 API를 통해 대량 데이터 레코드에 대한 예측을 요청하고 데이터를 쉽고 빠르게 변환할 수 있습니다.

**Q: Amazon SageMaker Edge Manager란 무엇입니까?**

SageMaker Edge Manager를 사용하면 스마트 카메라, 로봇, 개인용 컴퓨터 및 모바일 디바이스 같은 엣지 디바이스 플릿에서 기계 학습 모델을 손쉽게 최적화, 보안, 모니터링 및 유지 관리할 수 있습니다. SageMaker Edge Manager는 기계 학습 개발 시 다양한 엣지 디바이스에서 대규모로 ML 모델을 가동하는 데 도움이 됩니다.

**Q: SageMaker Edge Manager를 시작하려면 어떻게 해야 합니까?**

SageMaker Edge Manager를 시작하려면 클라우드에서 훈련된 ML 모델을 컴파일하고 패키지로 작성한 후 디바이스를 등록하고 SageMaker Edge Manager SDK에서 디바이스를 준비해야 합니다. 배포를 위해 모델을 준비하도록 SageMaker Edge Manager는 SageMaker Neo를 사용하여 대상 엣지 하드웨어에 맞게 모델을 컴파일합니다. 모델이 컴파일되면 SageMaker Edge Manager는 빠르게 배포할 수 있도록 AWS 생성 키로 모델에 서명한 후 런타임 및 필수 자격 증명을 통해 모델을 패키지로 작성합니다. 디바이스 측에서는, SageMaker Edge Manager에 디바이스를 등록하고, SageMaker Edge Manager SDK를 다운로드한 후 지침에 따라 SageMaker Edge Manager 에이전트를 디바이스에 설치합니다. 자습서 노트북에서는 SageMaker Edge Manager에서 모델을 준비하고 엣지 디바이스에 모델을 연결하는 방법에 대한 단계별 예제를 제공합니다.

**Q: SageMaker Edge Manager에서는 어떤 디바이스를 지원합니까?**

SageMaker Edge Manager는 Linux 및 Windows 운영 체제에서 일반적인 CPU(ARM, x86) 및 GPU(ARM, Nvidia) 기반 디바이스를 지원합니다. 앞으로 SageMaker Edge Manager는 SageMaker Neo에서도 지원하는 더 많은 임베디드 프로세서와 모바일 플랫폼을 지원하도록 확장할 계획입니다.

**Q: SageMaker Edge Manager를 사용하려면 SageMaker를 사용하여 모델을 훈련해야 하나요?**

아니요. 오픈 소스 또는 모델 공급 업체의 사전 훈련된 모델을 사용하거나 그 외 위치에서 모델을 훈련할 수 있습니다.

**Q: SageMaker Edge Manager를 사용하려면 SageMaker Neo를 사용하여 모델을 컴파일해야 하나요?**

예. SageMaker Neo는 엣지 디바이스에서 패키지로 작성하고 배포할 수 있도록 모델을 실행 가능 항목으로 변환하고 컴파일합니다. 모델 패키지를 배포하면 SageMaker Edge Manager 에이전트가 모델 패키지를 압축 해제하고 디바이스에서 모델을 실행합니다.

**Q: 엣지 디바이스에 모델을 배포하려면 어떻게 해야 하나요?**

SageMaker Edge Manager는 지정된 Amazon S3 버킷에 모델 패키지를 저장합니다. AWS IoT Greengrass에서 제공하는 무선 업데이트(OTA) 배포 기능을 사용하거나 원하는 다른 배포 메커니즘을 통해 S3 버킷에서 디바이스로 모델 패키지를 배포할 수 있습니다.

**Q: SageMaker Edge Manager SDK는 SageMaker Neo 런타임(dlr)과 어떻게 다른가요?**

Neo dlr은 SageMaker Neo 서비스에서 컴파일된 모델만 실행하는 오픈 소스 런타임입니다. 오픈 소스 dlr과 달리, SageMaker Edge Manager SDK는 추가 보안, 모델 관리 및 모델 지원 특성을 포함하는 엔터프라이즈급 디바이스 기반 에이전트를 포함합니다. SageMaker Edge Manager SDK는 대규모 프로덕션 배포에 적합합니다.

**Q: SageMaker Edge Manager는 AWS IoT Greengrass와 어떤 관계인가요?**

SageMaker Edge Manager와 AWS IoT Greengrass는 IoT 솔루션에서 함께 사용할 수 있습니다. 기계 학습 모델을 SageMaker Edge Manager에서 패키지로 작성한 후에 AWS IoT Greengrass의 OTA 업데이트 기능을 사용하여 디바이스에 모델 패키지를 배포할 수 있습니다. AWS IoT Greengrass를 사용하면 원격으로 IoT 디바이스를 모니터링하는 동시에, SageMaker Edge Manager가 디바이스에서 기계 학습 모델을 모니터링하고 유지 관리할 수 있습니다.

**Q: SageMaker Edge Manager는 AWS Panorama와 어떤 관계인가요? 언제 SageMaker Edge Manager를 사용하고, 언제 AWS Panorama를 사용해야 하나요?**

AWS는 엣지 디바이스에서 모델을 실행하기 위한 가장 포괄적이고 심층적인 기능을 제공합니다. 그리고 컴퓨터 비전, 음성 인식 및 예측 유지 관리를 포함하여 다양한 사용 사례를 지원하는 서비스도 갖추었습니다.

카메라 및 어플라이언스와 같은 엣지 디바이스에서 컴퓨터 비전을 실행하려는 회사는 AWS Panorama를 사용할 수 있습니다. AWS Panorama는 엣지 디바이스에서 컴퓨터 비전 애플리케이션을 쉽게 배포할 수 있도록 지원합니다. 클라우드 콘솔에 로그인하고 Amazon S3 또는 SageMaker에서 사용하려는 모델을 지정한 후 Python 스크립트로 비즈니스 로직을 작성하는 방법으로 간편하게 AWS Panorama를 시작할 수 있습니다. AWS Panorama는 대상 디바이스에 대해 모델을 컴파일하고 애플리케이션 패키지를 생성하므로, 몇 번의 클릭으로 디바이스에 해당 패키지를 배포할 수 있습니다. 사용자 지정 애플리케이션을 구축하려는 독립 소프트웨어 개발 판매 회사(ISV)도 AWS Panorama SDK를 사용할 수 있습니다. 디바이스 제조업체는 Device SDK를 사용하여 AWS Panorama에 대해 디바이스를 인증할 수 있습니다.

고유한 모델을 구축하고 모델 특성에 대한 세분화된 제어 기능을 원하는 고객이라면 SageMaker Edge Manager를 사용할 수 있습니다. SageMaker Edge Manager는 자연어 처리, 사기 탐지 및 예측 유지 관리와 같은 모든 유형의 사용 사례에 사용하는 스마트 카메라, 스마트 스피커, 로봇과 같은 여러 엣지 디바이스 플릿에서 기계 학습 모델을 준비, 실행, 모니터링 및 업데이트하는 관리형 서비스입니다. SageMaker Edge Manager는 서로 다른 모델 특성 및 드리프트에 대한 모니터 모델 엔지니어링을 포함하여 모델에 대한 제어를 원하는 기계 학습 엣지 개발자에게 적합합니다. 모든 ML 엣지 개발자는 SageMaker 콘솔 및 SageMaker API를 통해 SageMaker Edge Manager를 사용할 수 있습니다. SageMaker Edge Manager는 클라우드의 모델을 엣지 디바이스로 구축, 훈련 및 배포하는 SageMaker의 기능을 활용합니다.

**Q: SageMaker Edge Manager는 어느 리전에서 사용할 수 있나요?**

SageMaker Edge Manager는 미국 동부(버지니아 북부), 미국 동부(오하이오), 미국 서부(오레곤), 유럽(아일랜드), 유럽(프랑크푸르트), 아시아 태평양(도쿄)의 6개 리전에서 사용할 수 있습니다. 자세한 내용은 [AWS 리전 서비스 목록](#)을 참조하세요.

**Q: Amazon SageMaker Neo란 무엇인가요?**

SageMaker Neo를 사용하면 기계 학습 모델을 한 번 훈련하여 클라우드와 엣지의 모든 위치에서 실행할 수 있습니다. SageMaker Neo는 여러 하드웨어 플랫폼에 배포하는 데 사용할 수 있는 인기 DL 프레임워크로 구축된 모델을 자동으로 최적화합니다. 최적화된 모델은 최대 25배 더 빨리 실행되며 일반적인 기계 학습 모델과 비교해 10분의 1 미만의 리소스를 사용합니다.

**Q: SageMaker Neo를 시작하려면 어떻게 해야 하나요?**

SageMaker Neo를 시작하려면 SageMaker 콘솔에 로그인하고, 훈련된 모델을 선택한 다음, 예제를 따라 모델을 컴파일하고, 결과 모델을 대상 하드웨어 플랫폼에 배포합니다.

**Q: SageMaker Neo의 주요 구성 요소는 무엇인가요?**

SageMaker Neo에는 컴파일러와 런타임이라는 2가지 주요 구성 요소가 포함됩니다. 먼저, SageMaker Neo 컴파일러는 다른 프레임워크에서 내보낸 모델을 읽습니다. 그런 다음 프레임워크별 함수와 작업을 프레임워크와 무관한 중간 표시로 변환합니다. 다음에는 일련의 최적화를 수행합니다. 그런 다음, 컴파일러는 최적화된 작업에 대한 바이너리 코드를 생성하고 이 코드를 공유 객체 라이브러리에 기록합니다. 또한 컴파일러는 모델 정의와 파라미터를 개별 파일로 저장합니다. 실행 중에 SageMaker Neo 런타임은 컴파일러에서 생성된 아티팩트(모델 정의, 파라미터, 모델을 실행하기 위한 공유 객체 라이브러리)를 로드합니다.

**Q: SageMaker Neo를 사용하여 모델을 변환하려면 SageMaker를 사용하여 모델을 훈련해야 하나요?**

아니요. 다른 곳에서 모델을 훈련하고 SageMaker Neo를 사용하여 SageMaker ML 인스턴스 또는 AWS IoT Greengrass 지원 디바이스에 맞게 모델을 최적화할 수 있습니다.

#### Q: SageMaker Neo는 어떤 모델을 지원하나요?

현재, SageMaker Neo는 컴퓨터 비전 애플리케이션을 구동하는 가장 인기 있는 DL 모델과 오늘날 SageMaker에서 사용되는 가장 인기 있는 결정 트리 모델을 지원합니다. SageMaker Neo는 MXNet 및 TensorFlow에서 훈련된 AlexNet, ResNet, VGG, Inception, MobileNet, SqueezeNet, DenseNet 모델과 XGBoost에서 훈련된 분류 및 Random Cut Forest 모델을 최적화합니다.

#### Q: SageMaker Neo는 어떤 하드웨어 플랫폼을 지원하나요?

SageMaker Neo 설명서에서 [지원되는 클라우드 인스턴스](#), [엣지 디바이스](#) 및 프레임워크 버전의 목록을 확인할 수 있습니다.

#### Q: SageMaker Neo는 어느 리전에서 사용할 수 있나요?

지원되는 리전 목록은 [AWS 리전 서비스 목록](#)을 참조하세요.

## Amazon SageMaker 절감형 플랜

#### Q: Amazon SageMaker 절감형 플랜이란 무엇인가요?

SageMaker 절감형 플랜은 1년 또는 3년의 일정 사용량 약정(시간당 USD 요금으로 추정)을 조건으로 SageMaker에서 유연한 사용량 기반 요금 모델을 제공합니다. SageMaker 절감형 플랜은 최대 64%까지 비용을 절감할 수 있는 가장 유연한 요금 모델입니다. 이 요금은 인스턴스 패밀리, 크기 또는 리전과 관계없이 SageMaker Studio 노트북, SageMaker 온디맨드 노트북, SageMaker 처리, SageMaker Data Wrangler, SageMaker 훈련, SageMaker 실시간 추론 및 SageMaker 배치 변환을 포함하여 적격 SageMaker ML 인스턴스 사용량에 자동으로 적용됩니다. 예를 들어 추론 워크로드를 위해 미국 동부(오하이오)에서 실행되는 CPU 인스턴스 ml.c5.xlarge에서 미국 서부(오레곤)의 ml.inf1 인스턴스로 언제든지 변경할 수 있으며, 절감형 플랜 요금이 자동으로 계속 적용됩니다.

#### Q: SageMaker 절감형 플랜을 사용해야 하는 이유는 무엇인가요?

SageMaker 인스턴스 사용량이 일정한 크기이고(시간당 USD 요금으로 추정됨), 여러 SageMaker 구성 요소를 사용하거나 시간이 지남에 따라 기술 구성(예: 인스턴스 패밀리, 리전)이 변경될 것으로 예상되는 경우 SageMaker 절감형 플랜을 사용하면 절감 효과를 극대화하면서 애플리케이션 요구 사항 또는 새로운 혁신에 따라 기본적인 기술 구성을 유연하게 변경할 수 있습니다. 절감형 플랜 요금은 수동으로 수정하지 않고도 모든 적격 기계 학습 인스턴스 사용량에 자동으로 적용됩니다.

#### Q: SageMaker 절감형 플랜을 시작하려면 어떻게 해야 하나요?

AWS Management Console의 AWS Cost Explorer에서 또는 API/CLI를 사용하여 절감형 플랜 이용을 시작할 수 있습니다. AWS Cost Explorer에서 제공되는 권장 사항에 따라 손쉽게 Savings Plans 약정을 체결하여 가장 큰 절감 효과를 얻을 수 있습니다. 권장되는 시간 약정은 이전 온디맨드 사용량 기록과 고객이 선택한 요금제 유형, 기간, 결제 옵션을 기준으로 합니다. Savings Plans에 가입하면 컴퓨팅 사용량에 자동으로 할인된 Savings Plans 요금이 청구되고 약정 사용량을 초과하는 사용량에 대해서는 일반적인 온디맨드 요금이 청구됩니다.

#### Q: SageMaker용 절감형 플랜은 Amazon EC2용 컴퓨팅 절감형 플랜과 어떻게 다른가요?

SageMaker용 절감형 플랜과 Amazon EC2용 절감형 플랜의 차이는 포함되는 서비스에 있습니다. SageMaker 절감형 플랜은 SageMaker 기계 학습 인스턴스 사용량에만 적용됩니다.

#### Q: AWS Organizations/통합 결제에서 Savings Plans는 어떻게 작동합니까?

Savings Plans는 AWS Organization/통합 결제 패밀리 내의 모든 계정에서 구매할 수 있습니다. 기본적으로 Savings Plans에서 제공하는 혜택은 AWS Organization/통합 결제 패밀리 내 모든 계정의 사용량에 적용 가능합니다. 그러나 Savings Plans 혜택을 해당 용량을 구매한 계정으로만 제한할 수도 있습니다.

## Amazon SageMaker 요금에 대해 자세히 알아보기

[요금 페이지로 이동하기](#)

시작할 준비가 되셨나요?

[가입](#)

추가 질문이 있으십니까?

[문의하기](#)[콘솔에 로그인](#)

## AWS에 대해 자세히 알아보기

[AWS란 무엇입니까?](#)[클라우드 컴퓨팅이란 무엇입니까?](#)[AWS의 포용, 다양성, 평등](#)[DevOps란 무엇입니까?](#)[컨테이너란 무엇입니까?](#)[데이터 레이크란 무엇입니까?](#)[AWS 클라우드 보안](#)[새로운 소식](#)[블로그](#)[보도 자료](#)

## AWS 리소스

[시작하기](#)[교육 및 자격증](#)[AWS Solutions Library](#)[아키텍처 센터](#)[제품 및 기술 관련 FAQ](#)[애널리스트 보고서](#)[AWS 파트너](#)

## AWS에서의 개발자

[개발자 센터](#)[SDK 및 도구](#)[AWS에서의 .NET](#)[AWS에서의 Python](#)[AWS에서의 Java](#)[AWS 상의 PHP](#)[AWS 상의 JavaScript](#)

## 도움말

[AWS에 문의하기](#)[전문가의 도움 받기](#)[지원 티켓 제출](#)[AWS re:Post](#)[지식 센터](#)[AWS Support 개요](#)[법무](#)[AWS 채용 정보](#)[AWS 계정 생성](#)

Amazon은 기회균등을 보장하는 기업입니다( 소수/여성/장애/재향 군인/성 정체성/성적 지향/나이).

## 언어

[العربي](#)[Bahasa Indonesia](#)[Deutsch](#)[English](#)[Español](#)[Français](#)[Italiano](#)[Português](#)[Tiếng Việt](#)[Türkçe](#)[Русский](#)[ไทย](#)[日本語](#)[한국어](#)[中文\(简体\)](#)[中文\(繁體\)](#)[개인 정보 처리 방침](#)[|](#)[사이트 이용 약관](#)[|](#)[쿠키 기본 설정](#)[|](#)

© 2023, Amazon Web Services, Inc. 또는 자회사. All rights reserved.