AWS 백서

SageMaker Studio 관리 모범 사례

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

SageMaker Studio 관리 모범 사례: AWS 백서

Copyright © 2024 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 Amazon 계열사, 관련 업체 또는 Amazon의 지원 업체 여부에 상관없이 해당 소유자의 자산입니다.

Table of Contents

요약 및 소개 i
- 요약 1
- 귀사는 Well-Architected입니까? 1
- 소개 1

운영 모델 3
- 권장 계정 구조 3
  - 중앙 집중식 모델 계정 구조 4
  - 분산된 모델 계정 구조 5
  - 페더레이션된 모델 계정 구조 6
- ML 플랫폼 멀티테넌시 6

도메인 관리 8

다중 도메인 및 공유 스페이스 10
- 도메인에 공유 스페이스 설정 10
- IAM) 연동을 위한 도메인 설정 11
- Single Sign-On(SSO) 연동을 위한 도메인 설정 11
- SageMaker AI Studio 사용자 프로필 11
- Jupyter 서버 앱 12
- Jupyter 커널 게이트웨이 앱 12
- Amazon EFS 볼륨 13
  - 백업 및 복구 13
- Amazon EBS 볼륨 14
- 미리 서명된에 대한 액세스 보안 URL 14
- SageMaker AI 도메인 할당량 및 제한 15

자격 증명 관리 17
- 사용자, 그룹 및 역할 17
- 사용자 페더레이션 18
  - IAM 사용자 19
  - AWS IAM 또는 계정 연동 19
  - SAML를 사용한 인증 AWS Lambda 21
- AWS IAM IdC 페더레이션 21
  - 도메인 인증 지침 22

권한 관리 24
- IAM 역할 및 정책 24
- SageMaker AI Studio 노트북 권한 부여 워크플로 25

IAM 페더레이션: Studio Notebook 워크플로 ... 26
배포된 환경: SageMaker AI 훈련 워크플로 ... 27
데이터 권한 ... 28
AWS Lake Formation 데이터 액세스 ... 28
일반 가드레일 ... 29
특정 인스턴스로 노트북 액세스 제한 ... 29
규정 미준수 SageMaker AI Studio 도메인 제한 ... 30
승인되지 않은 SageMaker AI 이미지 시작 제한 ... 31
SageMaker AI VPC 엔드포인트를 통해서만 노트북 시작 ... 31
SageMaker AI Studio 노트북 액세스를 제한된 IP 범위로 제한 ... 32
SageMaker AI Studio 사용자가 다른 사용자 프로필에 액세스하지 못하도록 방지 ... 33
태그 지정 적용 ... 34
SageMaker AI Studio의 루트 액세스 ... 35
네트워크 관리 ... 37
VPC 네트워크 계획 ... 37
VPC 네트워크 옵션 ... 39
제한 사항 ... 41
데이터 보호 ... 42
저장 데이터 보호 ... 42
를 사용한 유휴 암호화 AWS KMS ... 42
전송 중 데이터 보호 ... 43
데이터 보호 가드레일 ... 43
유휴 SageMaker AI 호스팅 볼륨 암호화 ... 43
모델 모니터링 중에 사용되는 S3 버킷 암호화하기 ... 44
SageMaker AI Studio 도메인 스토리지 볼륨 암호화 ... 45
노트북을 공유하는 데 사용되는 S3에 저장된 데이터를 암호화합니다. ... 45
제한 사항 ... 46
로깅 및 모니터링 ... 47
를 사용하여 로깅 CloudWatch ... 47
를 사용한 감사 AWS CloudTrail ... 50
비용 분담 ... 51
자동 태그 지정 ... 51
비용 모니터링 ... 51
비용 관리 ... 52
사용자 지정 ... 53
수명 주기 구성 ... 53

SageMaker AI Studio 노트북용 사용자 지정 이미지 53
JupyterLab 확장 54
Git 리포지토리 54
Conda 환경 55
결론 56
부록 57
멀티테넌시 비교 57
SageMaker AI Studio 도메인 백업 및 복구 58
옵션 1:를 EFS 사용하여 기존에서 백업 EC2 58
옵션 2: S3 및 수명 주기 구성을 EFS 사용하여 기존에서 백업 59
SageMaker SAML어설션을 사용한 Studio 액세스 60
참조 자료 62
기여자 63
문서 수정 64
고지 사항 65
AWS 용어집 66
..... lxvii

SageMaker Studio 관리 모범 사례

게시 날짜: 2023년 4월 25일 (문서 수정)

요약

Amazon SageMaker AI Studio는 모든 기계 학습(ML) 개발 단계를 수행할 수 있는 단일 웹 기반 시각적 인터페이스를 제공하여 데이터 과학 팀 생산성을 향상시킵니다. SageMaker AI Studio는 모델을 빌드, 훈련 및 평가하는 데 필요한 각 단계에 대한 완전한 액세스, 제어 및 가시성을 제공합니다.

이 백서에서는 운영 모델, 도메인 관리, 자격 증명 관리, 권한 관리, 네트워크 관리, 로깅, 모니터링 및 사용자 지정을 포함한 여러 주제의 모범 사례를 설명합니다. 여기에서 설명하는 모범 사례는 다중 테넌트 배포를 포함한 엔터프라이즈 SageMaker AI Studio 배포를 위한 것입니다. 이 문서는 ML 플랫폼 관리자, ML 엔지니어 및 ML 아키텍트를 대상으로 합니다.

귀사는 Well-Architected입니까?

AWS Well-Architected 프레임워크는 클라우드에서 시스템을 구축할 때 내리는 결정의 장단점을 이해하는 데 도움이 됩니다. 이 프레임워크를 사용하여 클라우드에서 안정적이고 안전하며 효율적이고 비용 효율적인 시스템을 설계하고 운영하기 위한 아키텍처 모범 사례를 살펴볼 수 있습니다. AWS Management Console에서 무료로 제공되는 AWS Well-Architected Tool를 사용하면 각 요소에 대한 일련의 질문에 답하여, 이러한 모범 사례와 비교하여 워크로드를 검토할 수 있습니다.

Machine Learning Lens에서는, AWS 클라우드에서 기계 학습 워크로드를 설계, 배포 및 구축하는 방법에 중점을 둡니다. 이 Lens는 Well-Architected Framework에 설명된 모범 사례에 추가됩니다.

소개

SageMaker AI Studio를 ML 플랫폼으로 관리할 때는 워크로드가 증가함에 따라 ML 플랫폼을 확장하는 데 도움이 되는 정보에 입각한 결정을 내리기 위한 모범 사례 지침이 필요합니다. ML 플랫폼을 프로비저닝, 운영 및 확장하려면 다음 사항을 고려하세요.

- 적합한 운영 모델을 선택하고 비즈니스 목표에 맞게 ML 환경을 구성하세요.
- 사용자 ID에 대한 SageMaker AI Studio 도메인 인증을 설정하는 방법을 선택하고 도메인 수준 제한을 고려합니다.
- 세분화된 액세스 제어 및 감사를 위해 사용자의 ID 및 인증을 ML 플랫폼에 페더레이션하는 방법을 결정하세요.

- ML 페르소나의 다양한 역할에 대한 권한 및 가드레일을 설정하는 것을 고려해 보세요.
- ML 워크로드의 민감도, 사용자 수, 인스턴스 유형, 앱 및 시작된 작업을 고려하여 가상 프라이빗 클라우드(VPC) 네트워크 토폴로지를 계획합니다.
- 암호화를 통해 저장 데이터와 전송 중 데이터를 분류하고 보호하세요.
- 규정 준수를 위해 다양한 애플리케이션 프로그래밍 인터페이스(APIs) 및 사용자 활동을 로깅하고 모니터링하는 방법을 고려합니다.
- 자체 이미지 및 수명 주기 구성 스크립트로 SageMaker AI Studio 노트북 환경을 사용자 지정합니다.

운영 모델

운영 모델은 조직이 확장 가능하고 일관되며 효율적인 방식으로 비즈니스 가치를 제공하도록 지원하기 위해 사람, 프로세스 및 기술을 통합하는 프레임워크입니다. ML 운영 모델은 조직 전체의 팀을 위한 표준 제품 개발 프로세스를 제공합니다. 규모, 복잡성, 비즈니스 동인에 따라 운영 모델을 구현하기 위한 세 가지 모델이 있습니다.

- 중앙 집중식 데이터 과학 팀 - 이 모델에서는 모든 데이터 과학 활동이 단일 팀 또는 조직 내에서 중앙 집중화됩니다. 이는 모든 사업부가 데이터 과학 프로젝트를 위해이 팀을 찾는 Center of Excellence(COE) 모델과 유사합니다.
- 분산된 데이터 과학 팀 - 이 모델에서는 데이터 과학 활동이 다양한 비즈니스 기능 또는 부서 또는 다양한 제품군에 따라 분산됩니다.
- 페더레이션된 데이터 과학 팀 - 이 모델에서는 코드 리포지토리, 지속적 통합 및 지속적 전달(CI/CD) 파이프라인 등과 같은 공유 서비스 기능은 중앙 집중식 팀에 의해 관리되며 각 사업부 또는 제품 수준 기능은 분산된 팀에 의해 관리됩니다. 이는 각 사업부에 자체 데이터 과학 팀이 있는 허브 앤 스포크 모델과 비슷하지만 이러한 사업부 팀은 중앙 집중식 팀과 활동을 조정합니다.

프로덕션 사용 사례를 위해 첫 번째 스튜디오 도메인을 시작하기로 결정하기 전에 운영 모델과 환경 구성 AWS 모범 사례를 고려하세요. 자세한 내용은 여러 계정을 사용하여 AWS 환경 구성을 참조하세요.

다음 섹션에서는 각 운영 모델의 계정 구조를 구성하는 방법에 대한 지침을 제공합니다.

권장 계정 구조

이 섹션에서는 조직의 운영 요구 사항에 따라 시작하고 수정할 수 있는 운영 모델 계정 구조를 간략하게 소개합니다. 어떤 운영 모델을 선택하든 다음과 같은 일반적인 모범 사례를 구현하는 것이 좋습니다.

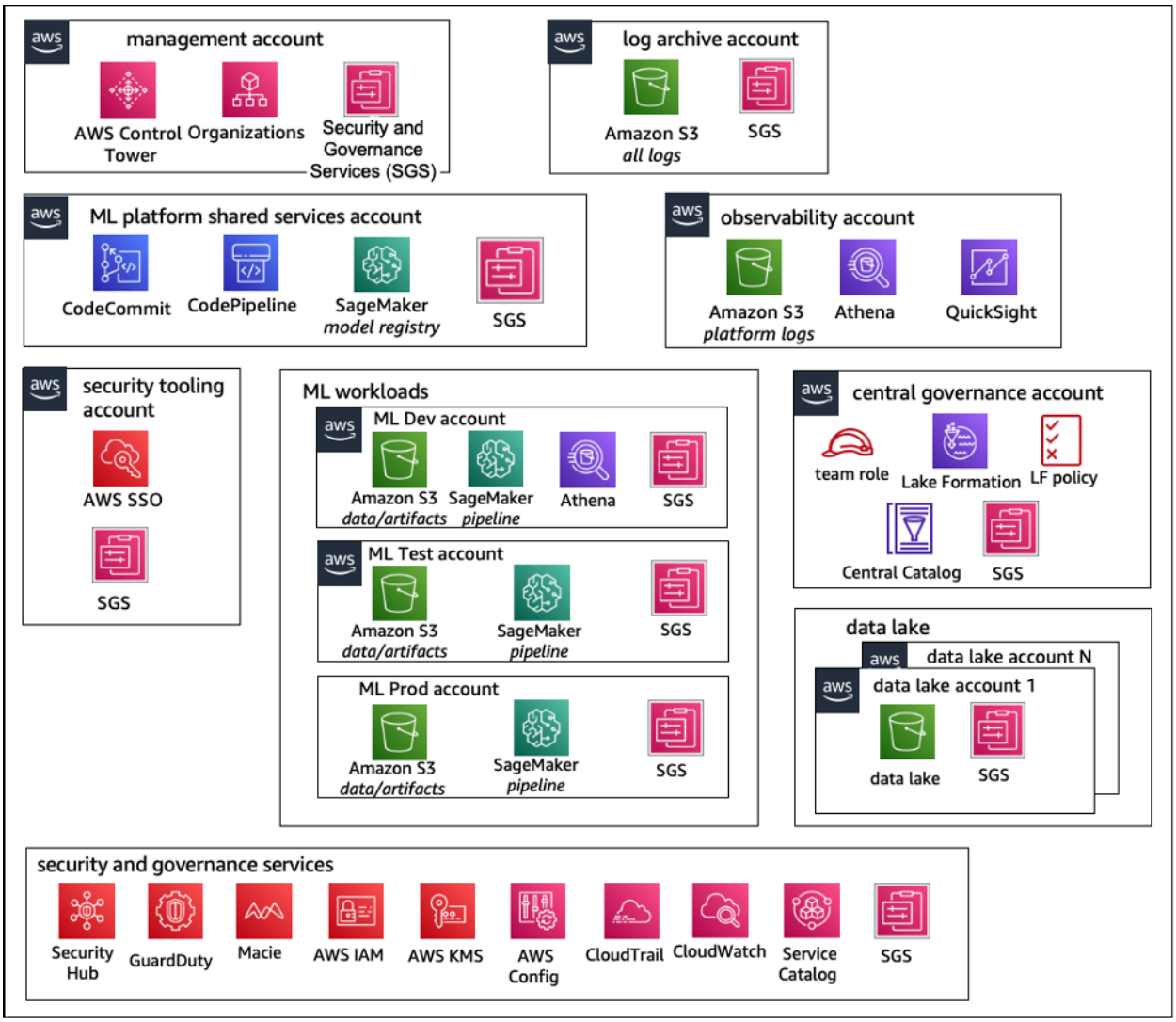
- 계정의 설정, 관리, 거버넌스에 AWS Control Tower를 사용하세요.
- 자격 증명 공급자(IdP)를 통해 자격 증명을 중앙 집중화하고 위임된 관리자 보안 도구 계정을 사용하여 AWS IAM 자격 증명 센터를 중앙 집중화하고 워크로드에 대한 보안 액세스를 활성화합니다.
- 개발, 테스트 및 프로덕션 워크로드에 걸쳐 계정 수준 격리를 통해 ML 워크로드를 실행합니다.
- ML 워크로드 로그를 로그 아카이브 계정으로 스트리밍한 다음, 관찰성 계정에서 로그 분석을 필터링하고 적용합니다.
- 데이터 액세스를 프로비저닝, 제어 및 감사하기 위한 중앙 집중식 거버넌스 계정을 운영하세요.

- 조직 및 워크로드 요구 사항에 따라 보안 및 규정 준수를 보장하기 위해 각 계정에 적절한 예방 및 탐지 가드레일을 사용하여 보안 및 거버넌스 서비스(SGS)를 포함합니다.

중앙 집중식 모델 계정 구조

이 모델에서 ML 플랫폼 팀은 다음을 제공할 책임이 있습니다.

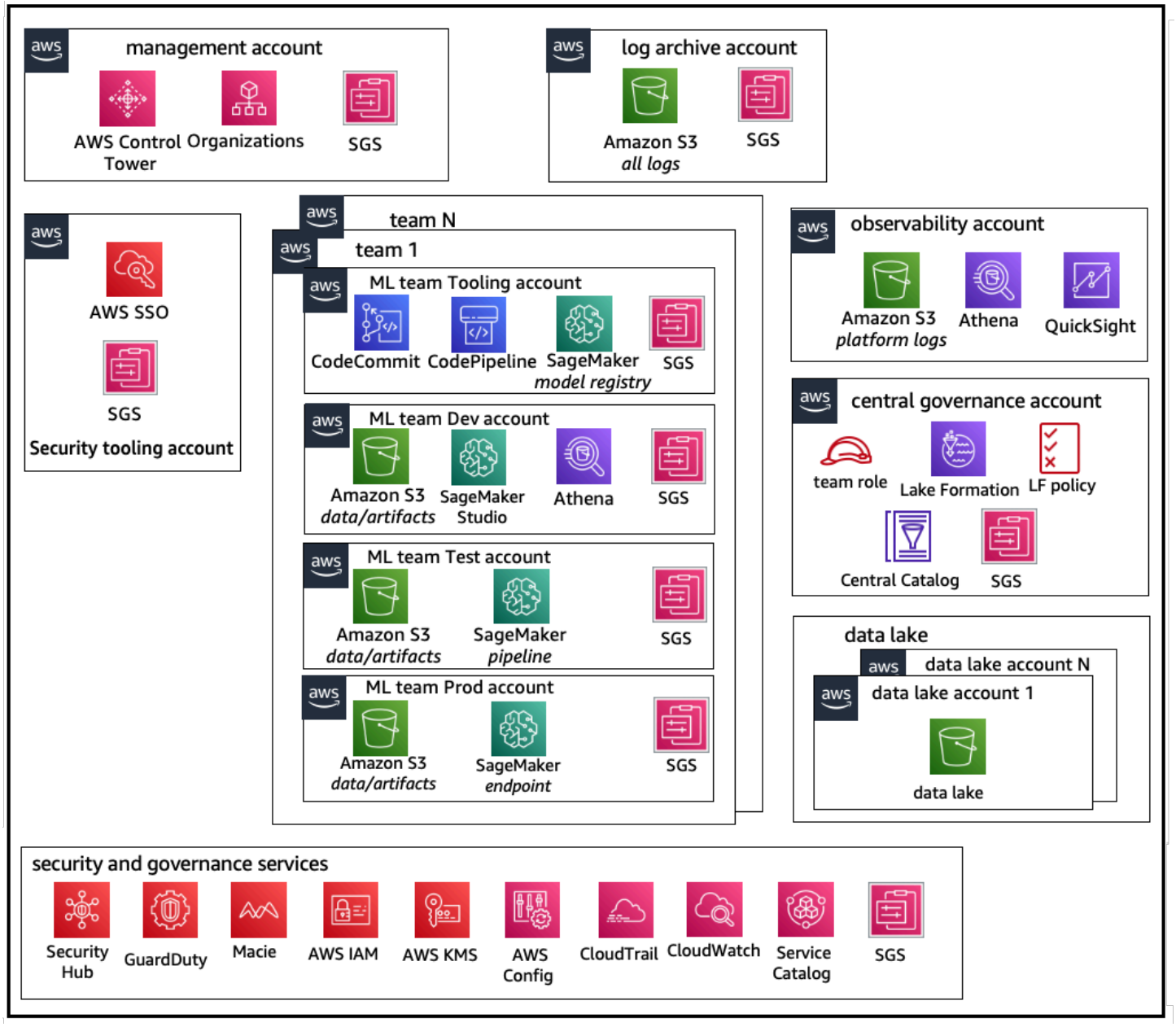
- 데이터 과학 팀의 Machine Learning 작업(MLOps) 요구 사항을 해결하는 공유 서비스 도구 계정입니다.
- 데이터 과학 팀 간에 공유된 ML 워크로드 개발, 테스트 및 프로덕션 계정
- 각 데이터 과학 팀 워크로드가 분리되어 실행되도록 하는 거버넌스 정책.
- 일반적인 모범 사례.



중앙 집중식 운영 모델 계정 구조

분산된 모델 계정 구조

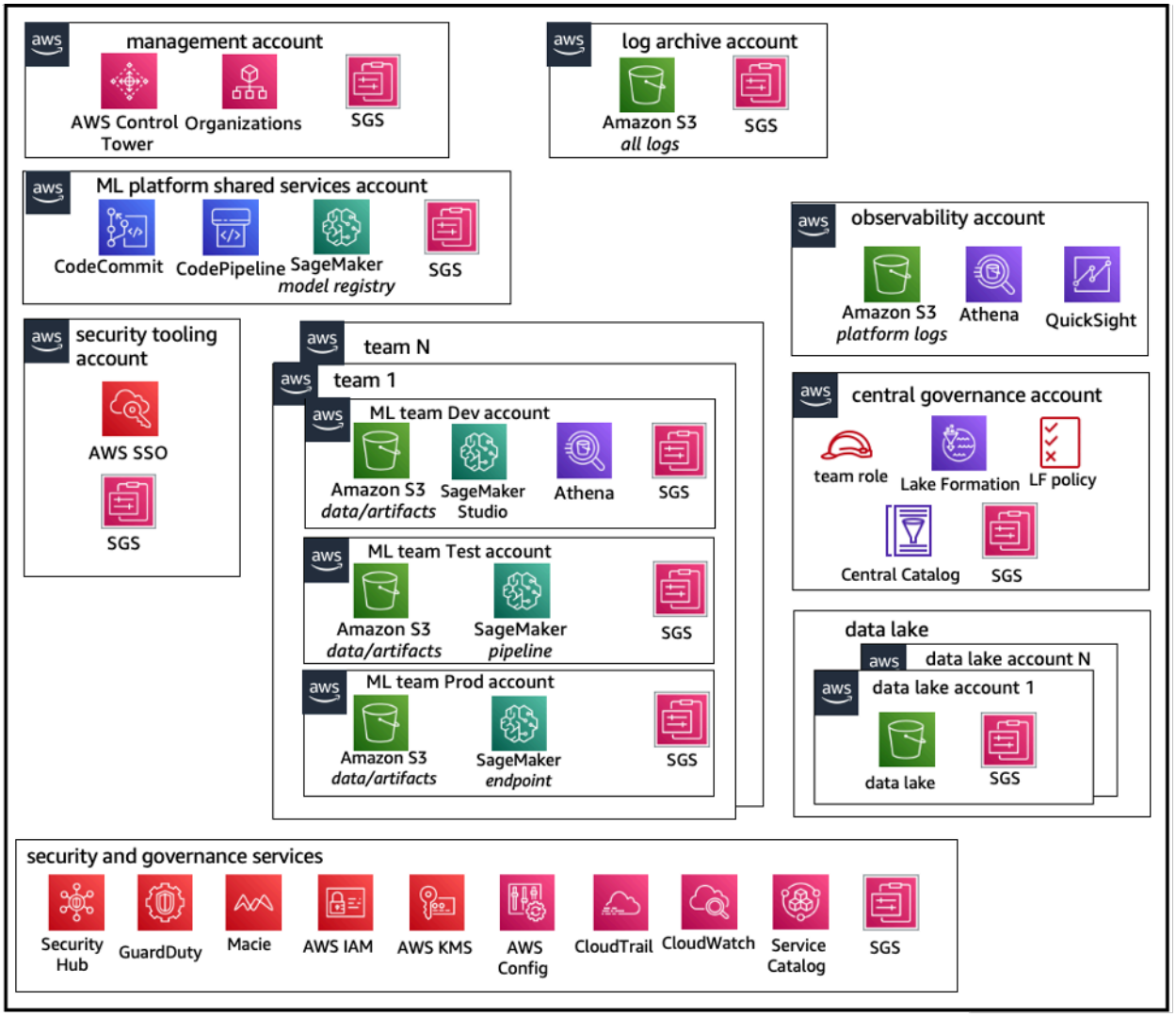
이 모델에서는 각 ML 팀이 ML 계정 및 리소스의 프로비저닝, 관리, 거버넌스를 위해 독립적으로 운영됩니다. 하지만 ML 팀은 중앙 집중식 관찰성 및 데이터 거버넌스 모델 접근 방식을 사용하여 데이터 거버넌스 및 감사 관리를 단순화하는 것이 좋습니다.



분산된 운영 모델 계정 구조

페더레이션된 모델 계정 구조

이 모델은 중앙 집중식 모델과 유사하지만, 주요 차이점은 각 데이터 science/ML team gets their own set of development/test/production 워크로드 계정이 ML 리소스의 강력한 물리적 격리를 가능하게 하고 각 팀이 다른 팀에 영향을 주지 않고 독립적으로 확장할 수 있다는 것입니다.



페더레이션된 운영 모델 계정 구조

ML 플랫폼 멀티테넌시

멀티테넌시는 단일 소프트웨어 인스턴스가 여러 개의 개별 사용자 그룹을 지원할 수 있는 소프트웨어 아키텍처입니다. 테넌트는 소프트웨어 인스턴스에 대한 특정 권한으로 공통 액세스를 공유하는 사용자 그룹입니다. 예를 들어 여러 ML 제품을 구축하는 경우 액세스 요구 사항이 비슷한 각 제품 팀을 테넌트 또는 팀으로 간주할 수 있습니다.

SageMaker AI Studio 인스턴스(예: SageMaker AI 도메인) 내에서 여러 팀을 구현할 수 있지만, 여러 팀을 단일 SageMaker AI Studio 도메인으로 가져올 때 블래스트 반경, 비용 속성 및 계정 수준 제한과 같은 장단점과 이러한 이점을 비교합니다. 다음 섹션에서 이러한 장단점 및 모범 사례에 대해 자세히 알아보십시오.

절대적인 리소스 격리가 필요한 경우 서로 다른 계정의 각 테넌트에 대해 SageMaker AI Studio 도메인을 구현하는 것이 좋습니다. 격리 요구 사항에 따라 여러 사업부(LOBs)를 단일 계정 및 리전 내의 여러 도메인으로 구현할 수 있습니다. 동일한 팀/의 구성원 간에 거의 실시간에 가까운 협업을 위해 공유 공간을 사용합니다LOB. 도메인이 여러 개인 경우에도 자격 증명 액세스 관리(IAM) 정책 및 권한을 사용하여 리소스 격리를 보장합니다.

SageMaker 도메인에서 생성된 AI 리소스에는 Amazon 리소스 이름(ARN)과 사용자 프로필 또는 공간이 자동으로 태깅되어 리소스 격리가 ARN 쉽습니다. 샘플 정책은 도메인 리소스 격리 설명서를 참조하십시오. 여기에서 설명서의 기능 비교와 함께 다중 계정 또는 다중 도메인 전략을 사용하는 시기에 대한 자세한 참조를 볼 수 있으며, 샘플 스크립트를 보고 GitHub 리포지토리의 기존 도메인에 대한 태그를 채울 수 있습니다.

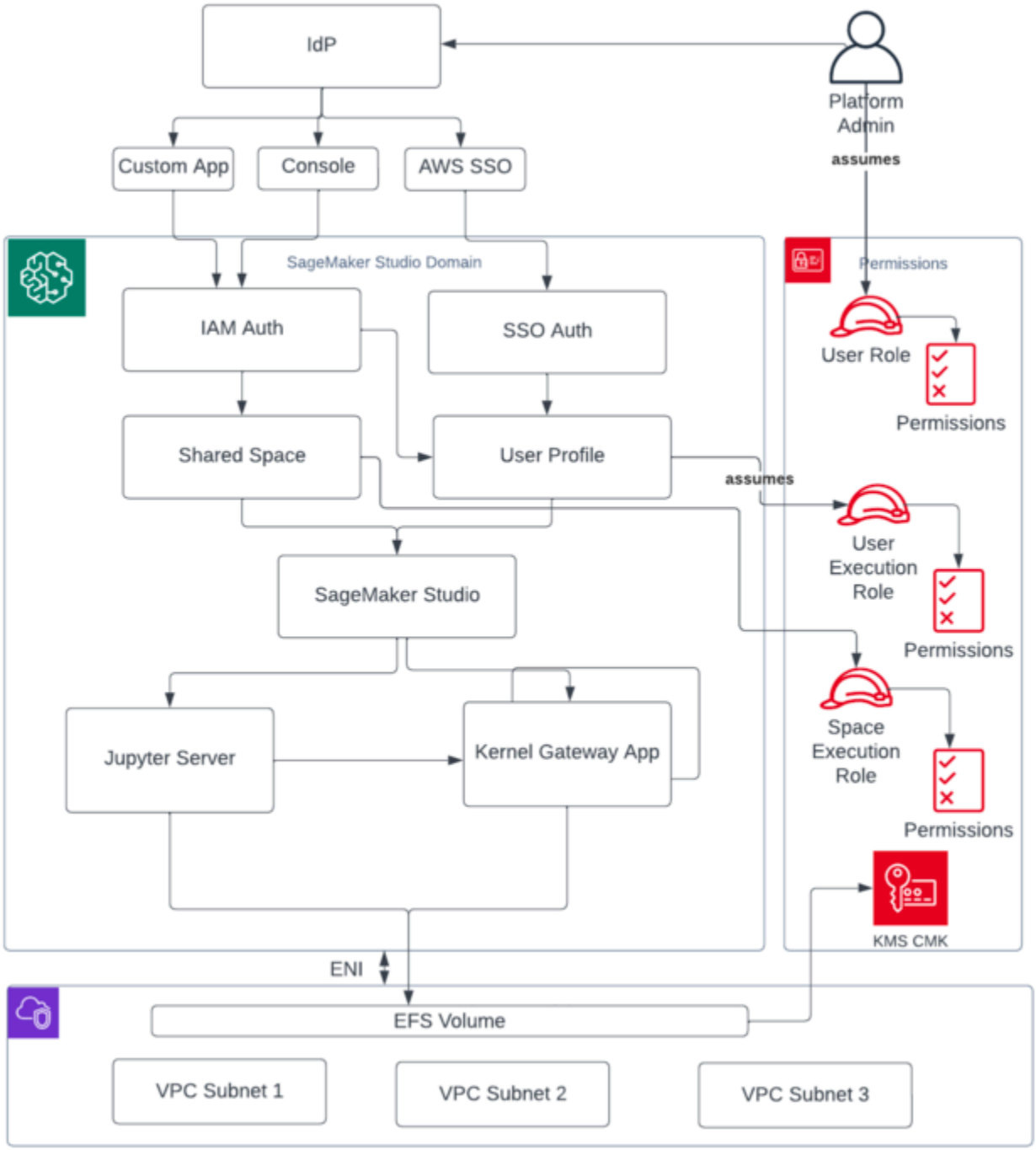
마지막으로를 사용하여 여러 계정에 SageMaker AI Studio 리소스의 셀프 서비스 배포를 구현할 수 있습니다AWS Service Catalog. 자세한 내용은 여러 AWS 계정 및에서 AWS Service Catalog 제품 관리를 AWS 리전 참조하세요.

도메인 관리

Amazon SageMaker AI 도메인은 다음으로 구성됩니다.

- 연결된 Amazon Elastic File System(AmazonEFS) 볼륨
- 권한이 부여된 사용자 목록
- 다양한 보안, 애플리케이션, 정책 및 Amazon Virtual Private Cloud(AmazonVPC) 구성

다음 다이어그램은 SageMaker AIStudio 도메인을 구성하는 다양한 구성 요소를 개략적으로 보여줍니다.

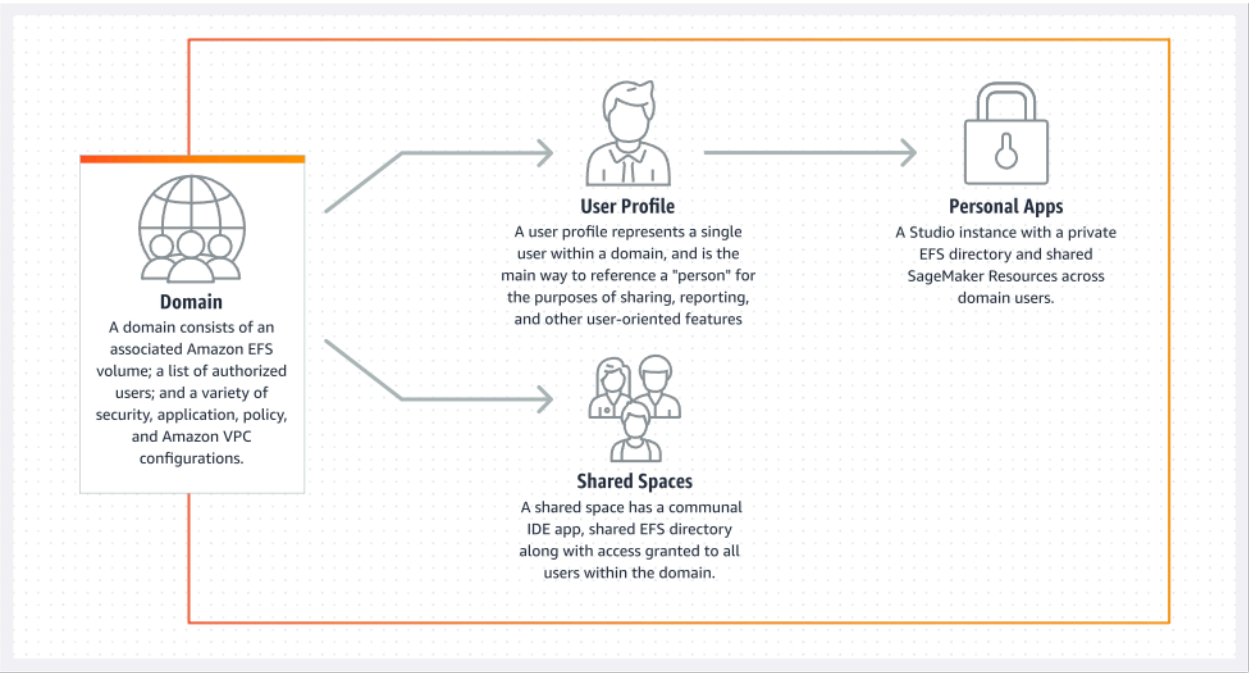


SageMaker AI Studio 도메인을 구성하는 다양한 구성 요소의 상위 수준 보기

다중 도메인 및 공유 스페이스

Amazon SageMaker AI는 이제 각 계정에 AWS 리전 대해 단일에 여러 SageMaker AI 도메인 생성을 지원합니다. 각 도메인에는 인증 모드와 같은 자체 도메인 설정과 및 서브넷VPC과 같은 네트워킹 설정이 있을 수 있습니다. 사용자 프로필은 도메인 간에 공유할 수 없습니다. 인간 사용자가 도메인으로 분리된 여러 팀에 속해 있는 경우 각 도메인에서 해당 사용자에 대한 사용자 프로필을 생성하십시오. 기존 도메인의 태그 채우기에 대해 알아보려면 다중 도메인 개요를 참조하세요.

IAM 인증 모드로 설정된 각 도메인은 사용자 간의 거의 실시간 협업을 위해 공유 공간을 사용할 수 있습니다. 공유 공간을 사용하면 사용자는 공유 Amazon EFS 디렉터리와 사용자 인터페이스용 공유 JupyterServer 앱에 액세스할 수 있으며 거의 실시간으로 공동 편집할 수 있습니다. 공유 스페이스에서 생성된 리소스의 자동 태그 지정을 통해 관리자는 프로젝트 수준에서 비용을 추적할 수 있습니다. 또한 공유 JupyterServer UI는 실험 및 모델 레지스트리 항목과 같은 리소스를 필터링하므로 공유 ML 노력과 관련된 항목만 표시됩니다. 다음 다이어그램은 각 도메인 내의 비공개 앱 및 공유 스페이스에 대한 개요를 제공합니다.



단일 도메인 내의 비공개 앱 및 공유 스페이스 개요

도메인에 공유 스페이스 설정

공유 공간은 일반적으로 단일 도메인의 멤버가 동일한 기본 파일 스토리지 및에 거의 실시간으로 액세스해야 하는 특정 ML 노력 또는 프로젝트에 대해 생성됩니다IDE. 사용자는 거의 실시간으로 노트북에

액세스하고, 읽고, 편집하고, 공유할 수 있으므로 동료와 함께 반복 작업을 아주 빠르게 시작할 수 있습니다.

공유 스페이스를 만들려면 먼저 스페이스를 사용하는 모든 사용자의 권한을 제어할 스페이스 기본 실행 역할을 지정해야 합니다. 이 글을 쓰는 시점에서 도메인 내의 모든 사용자는 해당 도메인의 모든 공유 스페이스에 액세스할 수 있습니다. 기존 도메인에 공유 스페이스를 추가하는 방법에 대한 최신 설명서를 확인하려면 공유 공간 만들기를 참조하세요.

IAM 페더레이션을 위한 도메인 설정

SageMaker AI Studio 도메인에 대한 페더레이션 AWS Identity and Access Management (IAM)을 설정하기 전에 자격 증명 관리 섹션에서 설명한 대로 IdP에서 IAM 페더레이션 사용자 역할(예: 플랫폼 관리자)을 설정해야 합니다.

IAM 옵션으로 SageMaker AI Studio를 설정하는 자세한 지침은 IAM Identity Center를 사용하여 Amazon SageMaker 도메인에 온보딩을 참조하세요.

Single Sign-On(SSO) 연동을 위한 도메인 설정

Single Sign-On(SSO) 연동을 사용하려면 SageMaker AI Studio를 실행해야 하는 리전과 동일한 리전 AWS IAM Identity Center 의 AWS Organizations 관리 계정에서를 활성화해야 합니다. 도메인 설정 단계는 인증 섹션에서 AWS IAM Identity Center (IdC)를 선택하는 것을 제외하고 IAM 페더레이션 단계와 유사합니다.

자세한 지침은 IAM Identity Center를 사용하여 Amazon SageMaker 도메인에 온보딩을 참조하세요.

SageMaker AI Studio 사용자 프로필

사용자 프로필은 도메인 내의 단일 사용자를 나타내며 공유, 보고 및 기타 사용자 중심 기능을 목적으로 “사람”을 참조하는 주된 방법입니다. 이 엔터티는 사용자가 toSageMaker AI Studio를 온보딩할 때 생성됩니다. 관리자가 이메일로 사용자를 초대하거나 IdC에서 가져오는 경우 사용자 프로파일이 자동으로 생성됩니다. 사용자 프로필은 개별 사용자의 기본 설정 소유자이며 사용자의 프라이빗 Amazon Elastic File System(AmazonEFS) 홈 디렉터리에 대한 참조가 있습니다. SageMaker AI Studio 애플리케이션의 각 물리적 사용자에 대한 사용자 프로필을 생성하는 것이 좋습니다. 각 사용자는 Amazon에 전용 디렉터리를 가지고 있으며EFS, 사용자 프로필은 동일한 계정의 도메인 간에 공유할 수 없습니다.

SageMaker AI Studio 도메인을 공유하는 각 사용자 프로필은 노트북을 실행하기 위한 전용 컴퓨팅 리소스(들)(예: SageMaker AI Amazon Elastic Compute Cloud(AmazonEC2) 인스턴스(들))를 가져옵니

다. 사용자 1에 할당된 컴퓨팅 인스턴스는 사용자 2에게 할당된 컴퓨팅 인스턴스와 완전히 분리되어 있습니다. 마찬가지로, 한 AWS 계정의 사용자들에게 할당된 컴퓨팅 리소스는 다른 계정의 사용자들에게 할당된 컴퓨팅 리소스와 완전히 분리됩니다. 각 사용자는 격리된 Docker 컨테이너 내에서 최대 4개의 애플리케이션(앱) 을 실행하거나 동일한 인스턴스 유형의 이미지를 실행할 수 있습니다.

Jupyter 서버 앱

미리 서명된에 액세스URL하거나 AWS IAM IdC를 사용하여 로그인하여 사용자를 위한 Amazon SageMaker AI Studio 노트북을 시작하면 SageMaker AI 서비스 관리형 VPC 인스턴스에서 Jupyter Server 앱이 시작됩니다. 각 사용자는 전용 Jupyter 서버 앱을 프라이빗 앱으로 받습니다. 기본적으로 SageMaker AI Studio 노트북용 Jupyter Server 앱은 전용 `ml.t3.medium` 인스턴스(시스템 인스턴스 유형으로 예약됨)에서 실행됩니다. 이 인스턴스의 컴퓨팅 요금은 고객에게 청구되지 않습니다.

Jupyter 커널 게이트웨이 앱

커널 게이트웨이 앱은 API 또는 SageMaker AI Studio 인터페이스를 통해 생성할 수 있으며 선택한 인스턴스 유형에서 실행됩니다. 이 앱은 널리 사용되는 데이터 과학으로 사전 구성된 기본 제공 SageMaker AI Studio 이미지 중 하나와 , TensorFlow Apache MXNet및와 같은 딥 러닝 패키지를 사용하여 실행할 수 있습니다PyTorch.

사용자는 동일한 SageMaker Studio 내에서 여러 Jupyter 노트북 커널, 터미널 세션 및 대화형 콘솔을 시작하고 실행할 수 있습니다image/Kernel Gateway app. Users can also run up to four Kernel Gateway apps or images on the same physical instance—each isolated by its container/image.

추가 앱을 만들려면 다른 인스턴스 유형을 사용해야 합니다. 사용자 프로필에서는 인스턴스 유형에 관계없이 하나의 인스턴스만 실행할 수 있습니다. 예를 들어, 사용자는 동일한 인스턴스에서 SageMaker AI Studio 내장 데이터 과학 이미지를 사용하여 간단한 노트북과 내장 TensorFlow 이미지를 사용하여 다른 노트북을 모두 실행할 수 있습니다. 인스턴스가 실행되는 시간만큼 사용자에게 요금이 청구됩니다. 사용자가 SageMaker AI Studio를 적극적으로 실행하지 않을 때 발생하는 비용을 방지하려면 인스턴스를 종료해야 합니다. 자세한 내용은 Studio 앱 종료 및 업데이트를 참조하세요.

SageMaker AI Studio 인터페이스에서 커널 게이트웨이 앱을 종료했다가 다시 열 때마다 해당 앱은 새 인스턴스에서 시작됩니다. 즉, 동일한 앱을 다시 시작해도 패키지 설치가 지속되지 않습니다. 마찬가지로 사용자가 노트북에서 인스턴스 유형을 변경하면 설치된 패키지와 세션 변수가 손실됩니다. 그러나 자체 이미지 및 수명 주기 스크립트 가져오기와 같은 기능을 사용하여 사용자 자체 패키지를 SageMaker AI Studio로 가져와 인스턴스 스위치 및 새 인스턴스 시작을 통해 유지할 수 있습니다.

Amazon Elastic File System 볼륨

도메인이 생성되면 도메인 내의 모든 사용자가 사용할 단일 Amazon Elastic File System(Amazon EFS) 볼륨이 생성됩니다. 각 사용자 프로필은 사용자의 노트북, GitHub 리포지토리 및 데이터 파일을 저장하기 위해 Amazon EFS 볼륨 내에서 프라이빗 홈 디렉터리를 수신합니다. 도메인 내 각 공간은 Amazon EFS 볼륨 내에서 여러 사용자 프로필로 액세스할 수 있는 프라이빗 디렉터리를 수신합니다. 폴더에 대한 액세스는 파일 시스템 권한을 통해 사용자에 의해 분리됩니다. SageMaker AI Studio는 각 사용자 프로필 또는 공간에 대한 글로벌 고유 사용자 ID를 생성하고, 이를 휴대용 운영 체제 인터페이스(POSIX)로 적용하여 데이터에 액세스user/group ID for the user's home directory on EFS, which prevents other users/spaces하지 못하도록 합니다.

백업 및 복구

기존 EFS 볼륨은 새 SageMaker AI 도메인에 연결할 수 없습니다. 프로덕션 설정에서 Amazon EFS 볼륨이 백업되어 있는지 확인합니다(다른 EFS 볼륨 또는 Amazon Simple Storage Service(Amazon S3)). 실수로 EFS 볼륨이 삭제된 경우 관리자는 SageMaker AI Studio 도메인을 해체하고 다시 생성해야 합니다. 프로세스는 다음과 같습니다.

, , 및 `DescribeSpace` API 호출을 통해 사용자 프로필`ListUserProfiles`, 공백 `DescribeUserProfile` `List Spaces`및 연결된 EFS 사용자IDs(UIDs) 목록을 백업합니다.

1. 새 SageMaker AI Studio 도메인을 생성합니다.
2. 사용자 프로필 및 스페이스를 생성합니다.
3. 각 사용자 프로필에 대해 EFS/Amazon S3의 백업에서 파일을 복사합니다.
4. 필요에 따라 이전 SageMaker AI Studio 도메인에서 모든 앱 및 사용자 프로필을 삭제합니다.

자세한 지침은 부록 섹션 SageMaker AI Studio 도메인 백업 및 복구를 참조하세요.

> **Note**
>
> 이는 사용자가 앱을 시작할 때마다 `LifecycleConfigurations`를 통해 S3와 데이터를 주고 받아서도 달성할 수 있습니다.

Amazon EBS 볼륨

Amazon Elastic Block Store(Amazon EBS) 스토리지 볼륨도 각 SageMaker AI Studio 노트북 인스턴스에 연결됩니다. 이것은 인스턴스에서 실행되는 컨테이너 또는 이미지의 루트 볼륨으로 사용됩니다. Amazon EFS 스토리지는 영구적이지만 컨테이너에 연결된 Amazon EBS 볼륨은 일시적입니다. 고객이 앱을 삭제하면 Amazon EBS 볼륨에 로컬로 저장된 데이터는 유지되지 않습니다.

미리 서명된에 대한 액세스 보안 URL

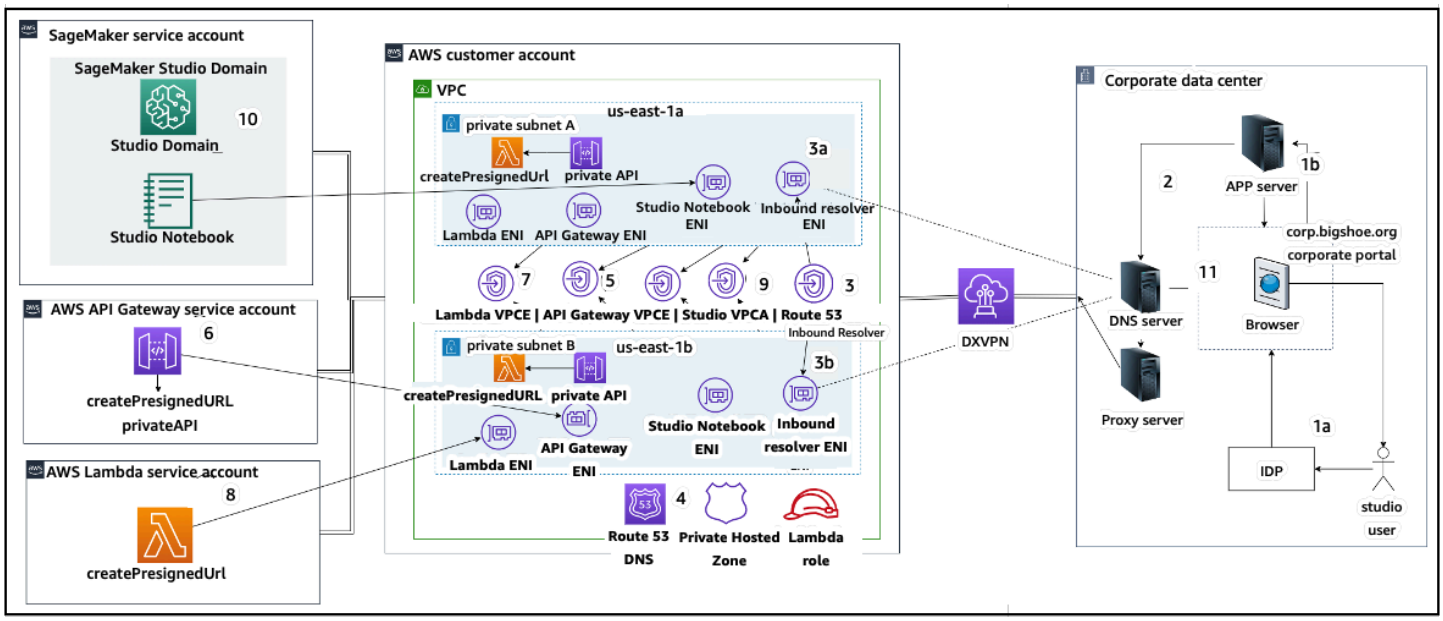
SageMaker AI Studio 사용자가 노트북 링크를 열면 SageMaker AI Studio는 페더레이션 사용자의 IAM 정책을 검증하여 액세스를 승인하고 해당 사용자에 URL 대해 미리 서명된를 생성하고 해결합니다. SageMaker AI 콘솔은 인터넷 도메인에서 실행되므로 미리 생성된이 URL는 브라우저 세션에 표시됩니다. 이로 인해 적절한 액세스 제어가 시행되지 않을 경우 데이터 도용 및 고객 데이터에 대한 액세스 권한을 획득할 수 있는 원치 않는 위협 요소가 발생할 수 있습니다.

Studio는 미리 서명된 URL 데이터 도난에 대한 액세스 제어를 적용하는 몇 가지 방법을 지원합니다.

- IAM 정책 조건을 사용한 클라이언트 IP 검증`aws:sourceIp`
- IAM 조건을 사용한 클라이언트 VPC 검증`aws:sourceVpc`
- IAM 정책 조건을 사용한 클라이언트 VPC 엔드포인트 검증`aws:sourceVpce`

SageMaker AI 콘솔에서 SageMaker AI Studio 노트북에 액세스할 때 사용 가능한 유일한 옵션은 IAM 정책 조건과 함께 클라이언트 IP 검증을 사용하는 것입니다`aws:sourceIp`. 하지만 Zscaler와 같은 브라우저 트래픽 라우팅 제품을 사용하여 작업 인력 인터넷 액세스의 규모와 규정 준수를 보장할 수 있습니다. 이러한 트래픽 라우팅 제품은 자체 소스 IP를 생성하며, 이 IP 범위는 기업 고객이 제어하지 않습니다. 따라서 이러한 기업 고객은 `aws:sourceIp` 조건을 사용할 수 없습니다.

IAM 정책 조건을 사용하여 클라이언트 VPC 엔드포인트 검증을 사용하려면 SageMaker AI Studio가 배포VPC된 동일한 고객URL에서 미리 서명된를 `aws:sourceVpce`생성하고 고객의 SageMaker AI Studio VPC 엔드포인트를 통해 미리 서명된를 해결URL해야 합니다VPC. 기업 네트워크 사용자의 액세스 시간 URL 동안 미리 서명된의 이러한 해결은 다음 아키텍처와 같이 DNS 전달 규칙(Zscaler 및 기업 모두DNS)을 사용하여 수행한 다음 Amazon Route 53인바운드 해석기를 사용하여 고객 VPC 엔드포인트로 수행할 수 있습니다.



기업 네트워크를 통해 VPC 엔드포인트URL로 사전 서명된 Studio에 액세스

이전 아키텍처를 설정하는 지침은 step-by-step 1URLs부: 기본 인프라에 미리 서명된 Amazon SageMaker AI Studio 보안을 참조하세요.

SageMaker AI 도메인 할당량 및 제한

- SageMaker AI Studio 도메인 SSO 페더레이션은 AWS Identity Center가 프로비저닝된 AWS 조직의 멤버 계정에서 리전에서만 지원됩니다.
- 공유 공간은 현재 AWS Identity Center로 설정된 도메인에서 지원되지 않습니다.
- VPC 도메인을 생성한 후에는 및 서브넷 구성을 변경할 수 없습니다. 하지만 다른 VPC 서브넷 구성으로 새 도메인을 생성할 수 있습니다.
- 도메인을 생성한 후에는 도메인 액세스를 IAM 및 SSO 모드 간에 전환할 수 없습니다. 다른 인증 모드로 새 도메인을 생성할 수 있습니다.
- 모든 사용자에 대해 인스턴스 유형당 실행되는 커널 게이트웨이 앱은 4개로 제한됩니다.
- 각 사용자는 각 인스턴스 유형에서 하나의 인스턴스만 시작할 수 있습니다.
- 도메인 내에서 소비되는 리소스에는 제한이 있습니다(예: 인스턴스 유형별로 시작하는 인스턴스 수, 생성할 수 있는 사용자 프로필 수). 서비스 한도의 전체 목록은 서비스 할당량 페이지를 참조하십시오.
- 고객은 도메인 또는 사용자 프로필 수와 같은 기본 리소스 제한을 상향 조정하는 데 필요한 비즈니스 근거가 있는 기업 지원 사례를 제출할 수 있으며, 여기에는 계정 수준의 가드레일이 적용됩니다.

- 계정당 동시 앱 수는 앱 2,500개로 엄격히 제한됩니다. 도메인 및 사용자 프로필 한도는 이 엄격한 한도에 따라 달라집니다. 예를 들어 계정에는 1,000개의 사용자 프로필이 있는 단일 도메인 또는 각각 50개의 사용자 프로필이 있는 20개의 도메인이 있을 수 있습니다.

자격 증명 관리

이 섹션에서는 기업 디렉터리의 작업 인력 사용자가 SageMaker AI Studio에 페더레이션 AWS 계정 하고 액세스하는 방법을 설명합니다. 먼저 사용자, 그룹, 역할이 매핑되는 방식과 사용자 페더레이션이 작동하는 방식을 간략하게 설명하겠습니다.

사용자, 그룹 및 역할

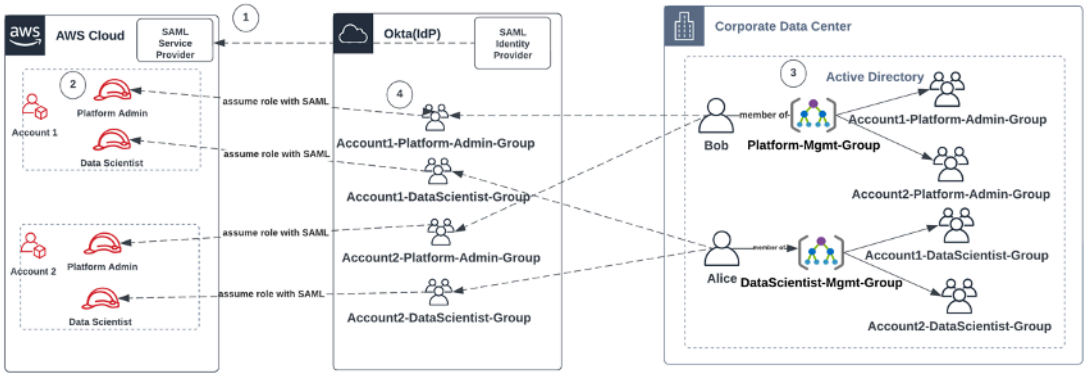
에서 리소스 권한은 사용자 AWS, 그룹 및 역할을 사용하여 관리됩니다. 고객은 IAM를 통해 또는 Okta와 같은 외부 IdP를 통해 활성화된 Active Directory(AD)와 같은 회사 디렉터리에서 사용자 및 그룹을 관리할 수 있습니다. 이를 통해 클라우드 및 온프레미스에서 실행되는 다양한 애플리케이션에 사용자를 인증할 수 있습니다.

AWS Security Pillar Identity Management 섹션에서 설명한 대로 중앙 IdP에서 사용자 ID를 관리하는 것이 모범 사례입니다. 이는 백엔드 HR 프로세스와 쉽게 통합되고 작업 인력 사용자에 대한 액세스를 관리하는 데 도움이 되기 때문입니다.

IdPs 예: Okta는 최종 사용자가 보안 어설션 마크업 언어()와 SSO 함께를 사용하여 하나 이상의에 인증AWS 계정 하고 특정 역할에 액세스할 수 있도록 허용합니다SAML. IdP 관리자는에서 IdP AWS 계정 로 역할을 다운로드하고 사용자에게 할당할 수 있습니다. 에 로그인할 때 AWS최종 사용자에게는 하나 이상의에 할당된 AWS 역할 목록을 표시하는 AWS 화면이 표시됩니다 AWS 계정. 최종 사용자는 로그인할 때 맡을 역할을 선택할 수 있으며, 이 역할에 따라 인증된 세션 기간 동안의 권한이 정의됩니다.

IdP에는 액세스를 제공하려는 각 특정 계정 및 역할 조합에 대한 그룹이 존재해야 합니다. 이러한 그룹들은 AWS 역할별 그룹이라고 부르겠습니다. 이러한 역할별 그룹의 구성원인 모든 사용자에게는 한 가지 권한이 부여됩니다. 즉, 한 개의 특정 AWS 계정내에서 하나의 특정 역할에 액세스할 수 있는 것입니다. 하지만 이 단일 권한 부여 프로세스로는 각 사용자를 특정 AWS 역할 그룹에 할당하여 사용자 액세스를 관리할 수 있을 정도로 확장되지 않습니다. 관리를 간소화하려면 조직의 다른 사용자 집합에 대해 다양한 AWS 권한 집합이 필요한 여러 그룹을 생성하는 것이 좋습니다.

중앙 IdP 설정을 설명하기 위해 사용자 및 그룹이 IdP 디렉터리와 동기화되는 AD 설정을 사용하는 엔터프라이즈를 예로 들어 보겠습니다. 에서 AWS이러한 AD 그룹은 IAM 역할에 매핑됩니다. 워크플로의 주요 단계는 다음과 같습니다.



AD 사용자, AD 그룹 및 IAM 역할 온보딩을 위한 워크플로

1. 에서 각에 AWS 계정 대해 IdP와의 SAML 통합을 AWS설정합니다.
2. 에서 각에 역할을 AWS설정하고 IdP AWS 계정 와 동기화합니다.
3. 기업 AD 시스템에서:
   a. 각 계정 역할에 대한 AD 그룹을 생성하고 IdP에 동기화합니다(예: `Account1-Platform-Admin-Group` (AWS 역할 그룹)).
   b. 각 페르소나 수준에서 관리 그룹을 생성하고(예: `Platform-Mgmt-Group`) AWS 역할 그룹을 멤버로 할당합니다.
   c. AWS 계정 역할에 대한 액세스를 허용하도록 사용자를 해당 관리 그룹에 할당합니다.
4. IdP에서 AWS 역할 그룹(예: `Account1-Platform-Admin-Group`)을 AWS 계정 역할(예: Account1의 플랫폼 관리자)에 매핑합니다.
5. 데이터 사이언티스트 Alice가 Idp에 로그인하면 '계정 1 데이터 사이언티스트'와 '계정 2 데이터 사이언티스트'라는 두 가지 옵션 중에서 선택할 수 있는 AWS 페더레이션 앱 UI가 표시됩니다.
6. Alice는 '계정 1 데이터 과학자' 옵션을 선택하면 AWS 계정 1(SageMaker AI 콘솔)의 승인된 애플리케이션에 연결됩니다.

SAML 계정 연동 설정에 대한 자세한 지침은 Okta의 AWS 계정 연동을 위한 SAML 2.0 구성 방법을 참조하세요.

사용자 페더레이션

SageMaker AI Studio에 대한 인증은 IAM 또는 IAM IdC를 사용하여 수행할 수 있습니다. 사용자가를 통해 관리되는 경우 IAM 모드를 선택할 IAM수 있습니다. 엔터프라이즈가 외부 IdP를 사용하는 경우 IAM 또는 IAM IdC를 통해 페더레이션할 수 있습니다. 기존 SageMaker AI Studio 도메인에 대해 인증

모드를 업데이트할 수 없으므로 프로덕션 SageMaker AI Studio 도메인을 생성하기 전에 결정을 내리는 것이 중요합니다.

SageMaker AI Studio가 IAM 모드로 설정된 경우, SageMaker AI Studio 사용자는 브라우저를 통해 액세스하면 SageMaker AI Studio 앱에 자동으로 로그인URL하는 미리 서명된를 통해 앱에 액세스합니다.

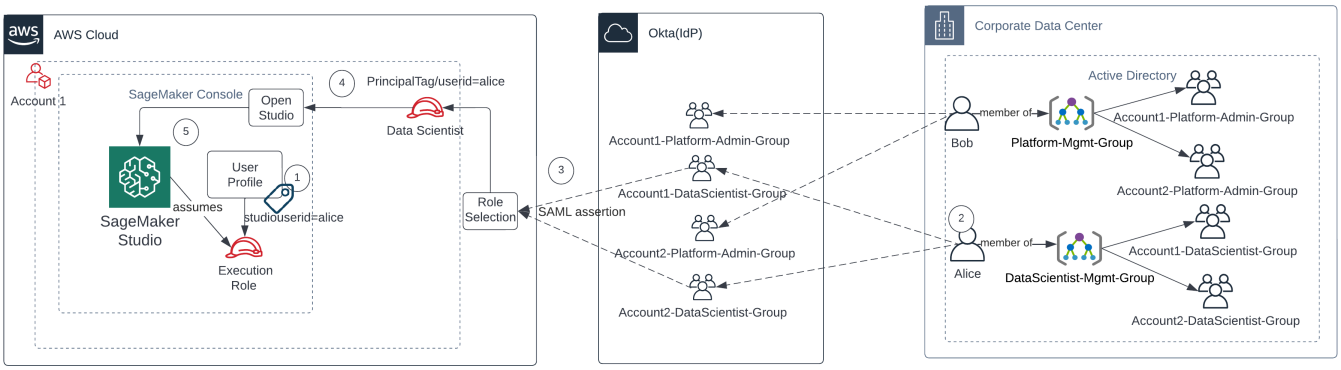
IAM 사용자

IAM 사용자의 경우 관리자는 각 사용자에 대한 SageMaker AI Studio 사용자 프로필을 생성하고, 사용자가 Studio 내에서 수행해야 하는 필요한 작업을 허용하는 IAM 역할과 사용자 프로필을 연결합니다. AWS 사용자가 SageMaker AI Studio 사용자 프로필에만 액세스하지 못하도록 제한하려면 관리자는 SageMaker AI Studio 사용자 프로필에 태그를 지정하고 태그 값이 AWS 사용자 이름과 동일한 경우에만 액세스하도록 허용하는 IAM 정책을 사용자에게 연결해야 합니다. 정책 명령문은 다음과 같습니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AmazonSageMakerPresignedUrlPolicy",
            "Effect": "Allow",
            "Action": [
                "sagemaker:CreatePresignedDomainUrl"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "sagemaker:ResourceTag/studiouserid": "${aws:username}"
                }
            }
        }
    ]
}
```

AWS IAM 또는 계정 연동

AWS 계정 페더레이션 메서드를 사용하면 고객이 Okta와 같은 SAML IdP에서 SageMaker AI 콘솔로 페더레이션할 수 있습니다. 사용자가 사용자 프로필에만 액세스하지 못하도록 제한하려면 관리자가 SageMaker AI Studio 사용자 프로필에 태그를 지정하고 IdP`PrincipalTags`에를 추가한 다음 전이적 태그로 설정해야 합니다. 다음 다이어그램은 페더레이션 사용자(데이터 과학자 Alice)가 자신의 SageMaker AI Studio 사용자 프로필에 액세스할 수 있는 권한을 부여하는 방법을 보여줍니다.



IAM 페더레이션 모드에서 SageMaker AI Studio 액세스

1. Alice SageMaker AI Studio 사용자 프로필에는 사용자 ID로 태그가 지정되고 실행 역할과 연결됩니다.
2. Alice가 IdP(Okta)에 인증합니다.
3. IdP는 Alice를 인증하고 두 역할(계정 1 및 2의 데이터 과학자)로 SAML어설션을 게시합니다. Alice는 멤버입니다. Alice는 계정 1의 데이터 과학자 역할을 선택합니다.
4. Alice는 데이터 사이언티스트 역할을 맡아 계정 1 SageMaker AI 콘솔에 로그인합니다. Alice는 Studio 앱 인스턴스 목록에서 Studio 앱 인스턴스를 엽니다.
5. 수임된 역할 세션의 Alice 보안 주체 태그는 선택한 SageMaker AI Studio 앱 인스턴스 사용자 프로필 태그에 대해 검증됩니다. 프로파일 태그가 유효하면 실행 역할을 수임하여 SageMaker AI Studio 앱 인스턴스가 시작됩니다.

사용자 온보딩의 일환으로 SageMaker AI 실행 역할 및 정책 생성을 자동화하려는 경우 이를 수행하는 한 가지 방법은 다음과 같습니다.

1. 예를 들어 각 계정 및 Studio 도메인 수준에서 AD 그룹(예: `SageMaker AI-Account1-Group`)을 설정합니다.
2. 사용자를 SageMaker AI Studio에 온보딩해야 하는 경우 사용자의 그룹 멤버십에 AI SageMaker AI-Account1-Group을 추가합니다.

`SageMaker AI-Account1-Group` 멤버십 이벤트를 수신하는 자동화 프로세스를 설정하고를 사용하여 AWS APIs AD 그룹 멤버십을 기반으로 역할, 정책, 태그 및 SageMaker AI Studio 사용자 프로필을 생성합니다. 해당 역할을 사용자 프로파일에 추가합니다. 샘플 정책은 섹션을 참조하세요 SageMaker AI Studio 사용자가 다른 사용자 프로필에 액세스하지 못하도록 방지.

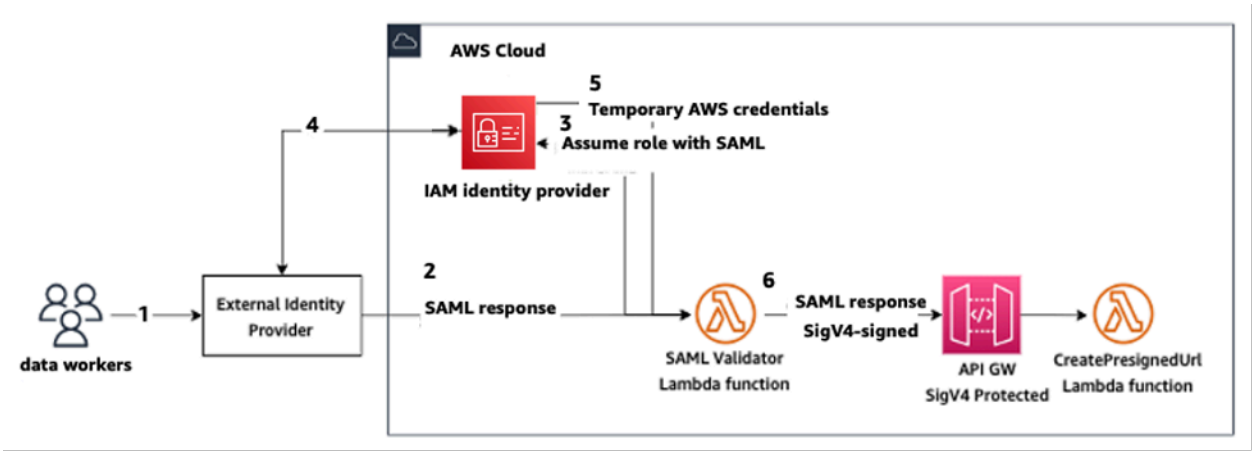
SAML를 사용한 인증 AWS Lambda

IAM 모드에서는 SAML어설션을 사용하여 사용자를 SageMaker AI Studio에 인증할 수도 있습니다. 이 아키텍처에서 고객은 사용자가 Studio에 액세스할 수 있는 SAML 애플리케이션을 생성할 수 있는 기존 IdP를 가지고 있습니다(AWS ID 연동 애플리케이션 대신). 고객의 IdP가에 추가됩니다IAM. AWS Lambda 함수는 IAM 및를 사용하여 어SAML설션을 검증STS한 다음 API 게이트웨이 또는 Lambda 함수를 직접 호출하여 미리 서명된 도메인를 생성하는 데 도움이 됩니다URL.

이 솔루션의 장점은 Lambda 함수가 SageMaker AI Studio에 액세스하기 위해 로직을 사용자 지정할 수 있다는 것입니다. 예제:

- 사용자 프로필이 없는 경우에는 사용자 프로필을 자동으로 생성합니다.
- SAML 속성을 구문 분석하여 역할 또는 정책 문서를 SageMaker AI Studio 실행 역할에 연결하거나 제거합니다.
- 수명 주기 구성(LCC)을 추가하고 태그를 추가하여 사용자 프로필을 사용자 지정합니다.

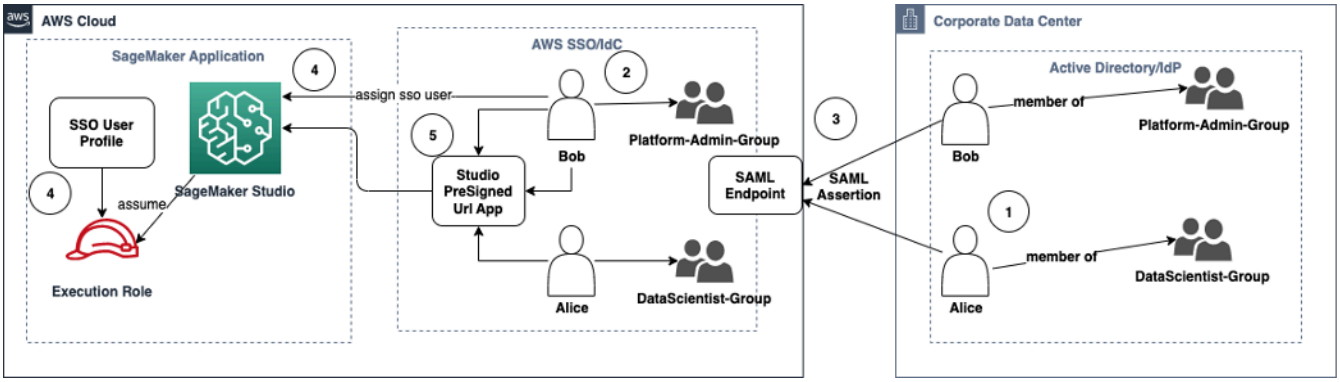
요약하면이 솔루션은 SageMaker AI Studio를 인증 및 권한 부여를 위한 사용자 지정 로직이 있는 SAML2.0 애플리케이션으로 노출합니다. 구현 세부 정보는 어SageMaker SAML설션을 사용한 부록 섹션 Studio 액세스를 참조하세요.



사용자 지정 SAML 애플리케이션을 사용하여 SageMaker AI Studio에 액세스

AWS IAM IdC 페더레이션

IdC 페더레이션 메서드를 사용하면 고객이 SAML IdP(예: Okta)에서 SageMaker AI Studio 애플리케이션에 직접 페더레이션할 수 있습니다. 다음 다이어그램은 페더레이션 사용자에게 자신의 SageMaker AI Studio 인스턴스에 액세스할 수 있는 권한이 부여되는 방법을 보여줍니다.



IAMIdC 모드에서 SageMaker AI Studio 액세스

1. 기업 AD에서 사용자는 플랫폼 관리자 그룹 및 데이터 과학자 그룹과 같은 AD 그룹의 구성원입니다.
2. Identity Provider(IdP)의 AD 사용자 및 AD 그룹은 Identity Center에 AWS IAM 동기화되며 각각 할당을 위한 Single Sign-On 사용자 및 그룹으로 사용할 수 있습니다.
3. IdP는 어SAML설션을 AWS IdC SAML 엔드포인트에 게시합니다.
4. SageMaker AI Studio에서 IdC 사용자는 SageMaker Studio 애플리케이션에 할당됩니다. 이 할당은 IdC 그룹을 사용하여 수행할 수 있으며 SageMaker AI Studio는 각 IdC 사용자 수준에서 적용됩니다. 이 할당이 생성되면 SageMaker AI Studio는 IdC 사용자 프로필을 생성하고 도메인 실행 역할을 연결합니다.
5. 사용자는 IdC.AI Studio에서 클라우드 애플리케이션으로 URL 호스팅되는 미리 서명된 보안를 사용하여 SageMaker AI Studio 애플리케이션에 액세스합니다. SageMaker IdC 사용자 프로필에 연결된 실행 역할을 수임합니다.

도메인 인증 지침

도메인의 인증 모드를 선택할 때 고려해야 할 몇 가지 사항은 다음과 같습니다.

1. 사용자가에 액세스하지 않고 SageMaker AI Studio UI를 직접 AWS Management Console 보지 않도록 하려면 IdC에서 AWS IAM Single Sign-On 모드를 사용합니다.
2. 사용자가에 액세스하지 않고 IAM 모드에서 SageMaker AI Studio UI를 직접 AWS Management Console 보지 않도록 하려면 백엔드의 Lambda 함수를 사용하여 사용자 프로필에 URL 대해 미리 서명된를 생성하고 SageMaker AI Studio UI로 리디렉션하면 됩니다.
3. IdC 모드에서는 각 사용자가 단일 사용자 프로필에 매핑됩니다.
4. IdC 모드에서는 모든 사용자 프로필에 기본 실행 역할이 자동으로 할당됩니다. 사용자에게 다른 실행 역할을 할당하려면 UpdateUserProfile를 사용하여 사용자 프로필을 업데이트해야 합니다API.

5. 인터넷을 통과하지 않고 IAM 모드(생성된 미리 서명된 사용URL)의 SageMaker AI Studio UI 액세스를 VPC 엔드포인트로 제한하려면 사용자 지정 DNS 해석기를 사용할 수 있습니다. 1URLs부: 기본 인프라 블로그 게시물에 미리 서명된 Secure Amazon SageMaker AI Studio를 참조하세요.

권한 관리

이 섹션에서는 SageMaker AI Studio 도메인을 프로비저닝하고 운영하기 위해 일반적으로 사용되는 IAM 역할, 정책 및 가드레일을 설정하는 모범 사례를 설명합니다.

IAM 역할 및 정책

가장 좋은 방법은 먼저 ML 수명 주기에 관련된 보안 주체라고 하는 관련 사용자 및 애플리케이션과 이를 부여하는 데 필요한 AWS 권한을 식별하는 것입니다. SageMaker AI는 관리형 서비스이므로 사용자를 대신하여 API 호출할 수 있는 AWS 서비스인 서비스 보안 주체도 고려해야 합니다. 다음 다이어그램은 조직의 다양한 페르소나에 해당하는 생성하려는 다양한 IAM 역할을 보여줍니다.

SageMaker AI IAM 역할

이러한 역할은 필요한 특정 역할의 몇 가지 예와 함께 자세히 설명되어 IAMpermissions 있습니다.

- ML Admin 사용자 역할 - 스튜디오 도메인 및 사용자 프로필(`sagemaker:CreateDomain`, `sagemaker:CreateUserProfile`)을 생성하고, 사용자를 위한 AWS Key Management Service AWS KMS키를 생성하고, 데이터 과학자를 위한 S3 버킷을 생성하고, 컨테이너를 수용하기 위한 Amazon ECR리포지토리를 생성하여 데이터 과학자를 위한 환경을 프로비저닝하는 보안 주체입니다. 또한 사용자에 대한 기본 구성 및 수명 주기 스크립트를 설정하고, 사용자 지정 이미지를 빌드하여 SageMaker AI Studio 도메인에 연결하며, 사용자 지정 프로젝트, Amazon EMR 템플릿과 같은 서비스 카탈로그 제품을 제공할 수 있습니다.

  예를 들어이 보안 주체는 훈련 작업을 실행하지 않으므로 SageMaker AI 훈련 또는 처리 작업을 시작하는 데 권한이 필요하지 않습니다. CloudFormation 또는 Terraform과 같은 인프라를 코드 템플릿으로 사용하여 도메인 및 사용자를 프로비저닝하는 경우 프로비저닝 서비스는이 역할을 수임하여 관리자를 대신하여 리소스를 생성합니다. 이 역할은를 사용하여 SageMaker AI에 대한 읽기 전용 액세스 권한을 가질 수 있습니다 AWS Management Console.

  또한이 사용자 역할에는 프라이빗 내에서 도메인을 시작하려면 특정 EC2 권한VPC, EFS 볼륨을 암호화할 수 있는 KMS 권한, Studio()에 대한 서비스 연결 역할을 생성할 수 있는 권한이 필요합니

다`iam:CreateServiceLinkedRole`. 이러한 세분화된 권한에 대해서는 이 문서의 뒷부분에서 설명하겠습니다.

- 데이터 과학자 사용자 역할 -이 보안 주체는 SageMaker AI Studio에 로그인하고, 데이터를 탐색하고, 처리 및 훈련 작업과 파이프라인을 생성하는 사용자입니다. 사용자에게 필요한 기본 권한은 SageMaker AI Studio를 시작할 수 있는 권한이며, 나머지 정책은 SageMaker AI 실행 서비스 역할에서 관리할 수 있습니다.
- SageMaker AI 실행 서비스 역할 - SageMaker AI는 관리형 서비스이므로 사용자를 대신하여 작업을 시작합니다. 많은 고객이 훈련 작업, 처리 작업 또는 모델 호스팅 작업을 실행할 때 단일 실행 역할을 사용하기로 선택하기 때문에 이 역할은 허용되는 권한 측면에서 가장 광범위한 경우가 많습니다. 이는 쉽게 시작할 수 있는 방법이지만, 고객이 여정에서 성숙하기 때문에 특히 배포된 환경에서 해당 작업을 실행할 때 노트북 실행 역할을 다른 API 작업에 대한 별도의 역할로 분할하는 경우가 많습니다.

  생성 시 역할을 SageMaker AI Studio 도메인과 연결합니다. 그러나 고객이 도메인의 다른 사용자 프로필과 연결된 다양한 역할을 유연하게 보유해야 할 수 있으므로(예: 직무에 따라) 각 사용자 프로필과 별도의 IAM 역할을 연결할 수도 있습니다. 단일 실제 사용자를 단일 사용자 프로필에 매핑하는 것이 좋습니다. 생성 시 사용자 프로필에 역할을 연결하지 않으면 기본 동작은 도메인 실행 역할을 사용자 프로필과 연결하는 SageMaker AIStudio 것입니다.

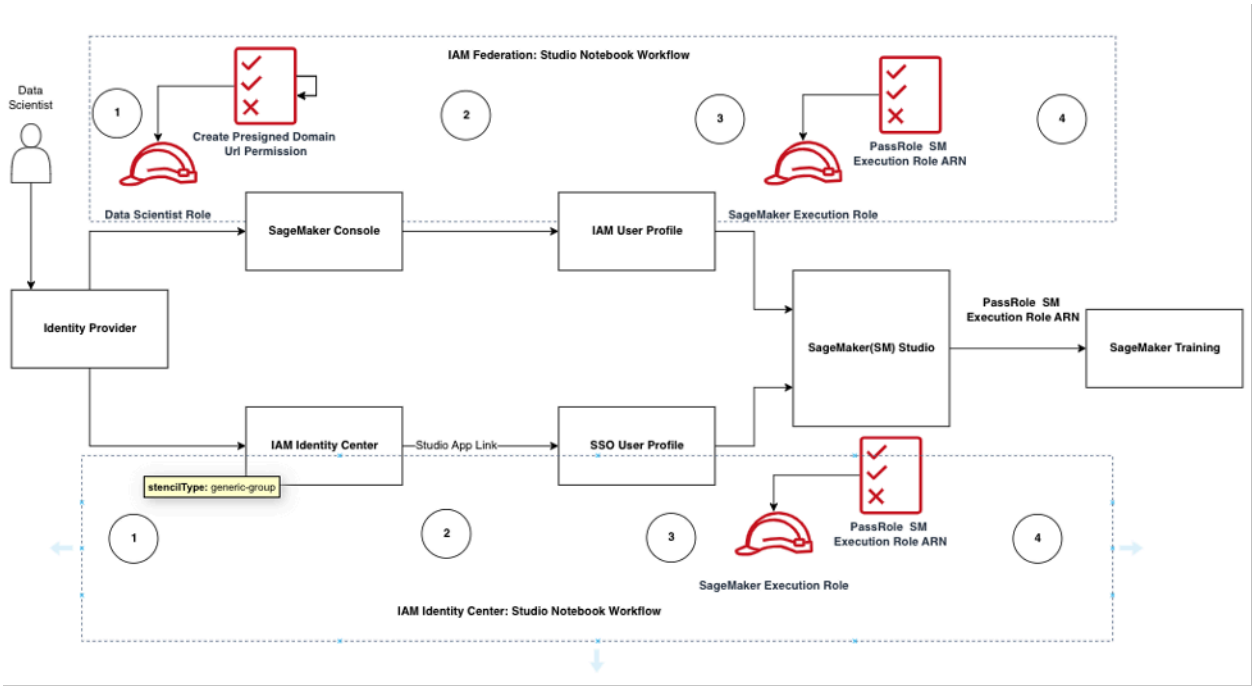
  여러 데이터 과학자와 ML 엔지니어가 프로젝트에서 함께 작업하고 리소스에 액세스하기 위한 공유 권한 모델이 필요한 경우 팀원 간에 IAM 권한을 공유하기 위한 팀 수준 SageMaker AI 서비스 실행 역할을 생성하는 것이 좋습니다. 각 사용자 수준에서 권한을 잠가야 하는 인스턴스에서는 개별 사용자 수준 SageMaker AI 서비스 실행 역할을 생성할 수 있지만 서비스 제한에 유의해야 합니다.

SageMaker AI Studio 노트북 권한 부여 워크플로

이 섹션에서는 SageMaker AI Studio Notebook 인증이 SageMaker AI Studio Notebook에서 직접 모델을 구축하고 훈련하기 위해 데이터 사이언티스트가 수행해야 하는 다양한 활동에 어떻게 적용되는지 설명합니다. SageMaker AI 도메인은 두 가지 권한 부여 모드를 지원합니다.

- IAM 페더레이션
- IAM 자격 증명 센터

다음으로 이 백서에서는 각 모드에 대한 데이터 과학자 인증 워크플로를 안내합니다.



Studio 사용자를 위한 인증 및 권한 부여 워크플로

IAM 페더레이션: SageMaker Studio Notebook 워크플로

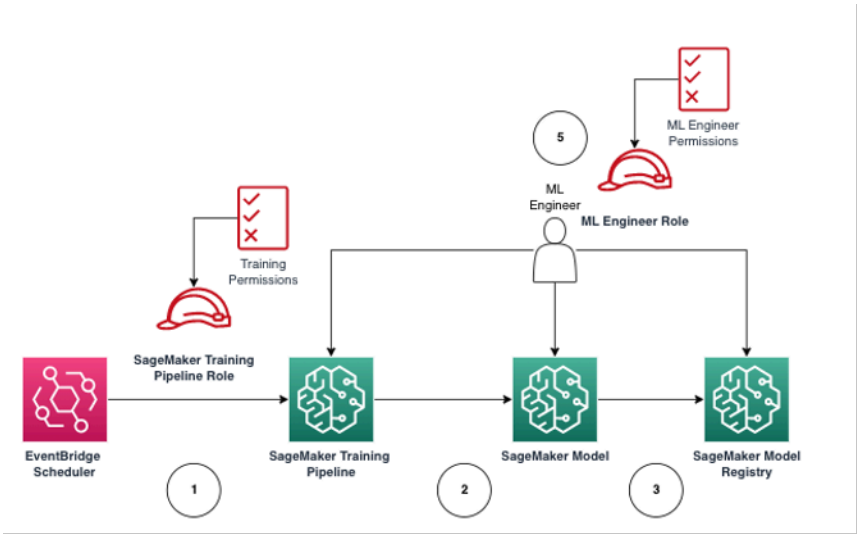
1. 데이터 사이언티스트는 기업 자격 증명 공급자에 인증하고 SageMaker AI 콘솔에서 데이터 사이언티스트 사용자 역할(사용자 페더레이션 역할)을 수임합니다. 이 페더레이션 역할에는 Amazon 리소스 이름(ARN) 역할을 SageMaker Studio에 전달할 수 있는 SageMaker AI 실행 역할에 대한 `iam:PassRole` API 권한이 있습니다.
2. 데이터 사이언티스트는 SageMaker AI 실행 역할과 연결된 Studio IAM 사용자 프로필에서 Open Studio 링크를 선택합니다.
3. 사용자 프로필의 SageMaker 실행 역할 권한을 가정하여 SageMaker Studio IDE 서비스가 시작됩니다. 이 역할에는 SageMaker AI 실행 역할이 SageMaker AI 훈련 서비스에 역할을 전달할 ARN 수 있는 `iam:PassRole` API 권한이 있습니다.
4. Data Scientist가 원격 컴퓨팅 노드(들)에서 훈련 작업을 시작하면 SageMaker AI 실행 역할ARN이 SageMaker AI 훈련 서비스로 전달됩니다. 이렇게 하면이를 사용하여 새 역할 세션이 생성ARN되고 훈련 작업이 실행됩니다. 훈련 작업에 대한 권한을 추가로 축소해야 하는 경우 훈련 관련 역할을 생성하고 훈련를 호출할 ARN 때 해당 역할을 전달할 수 있습니다API.

IAM Identity Center: SageMaker AI Studio 노트북 워크플로

1. 데이터 사이언티스트는 기업 자격 증명 공급자에 인증하고 AWS IAM Identity Center를 클릭합니다. 데이터 과학자에게는 사용자를 위한 Identity Center 포털이 제공됩니다.
2. 데이터 과학자는 SageMaker AI 실행 역할과 연결된 IdC 사용자 프로필에서 생성된 SageMaker AI Studio 앱 링크를 클릭합니다.
3. 사용자 프로필의 SageMaker AI 실행 역할 권한을 가정하여 SageMaker AI Studio IDE 서비스가 시작됩니다. 이 역할에는 SageMaker AI 실행 역할이 SageMaker AI 훈련 서비스에 역할을 전달할 ARN 수 있는 `iam:PassRole` API 권한이 있습니다.
4. 데이터 과학자가 원격 컴퓨팅 노드(들)에서 훈련 작업을 시작하면 SageMaker AI 실행 역할ARN이 SageMaker AI 훈련 서비스로 전달됩니다. 실행 역할은이를 사용하여 새 역할 세션을 ARN 생성하고 훈련 작업을 ARN실행합니다. 훈련 작업에 대한 권한을 추가로 축소해야 하는 경우 훈련 관련 역할을 생성하고 훈련를 호출할 ARN 때 해당 역할을 전달할 수 있습니다API.

배포된 환경: SageMaker AI 훈련 워크플로

시스템 테스트 및 프로덕션과 같은 배포된 환경에서 작업은 자동 스케줄러 및 이벤트 트리거를 통해 실행되며, 이러한 환경에 대한 인적 액세스는 SageMaker AI Studio 노트북에서 제한됩니다. 이 섹션에서는 배포된 환경에서 IAM 역할이 SageMaker AI 훈련 파이프라인과 작동하는 방법을 설명합니다.



SageMaker 관리형 프로덕션 환경의 AI 훈련 워크플로

1. Amazon EventBridge 스케줄러는 SageMaker AI 훈련 파이프라인 작업을 트리거합니다.
2. SageMaker AI 훈련 파이프라인 작업은 SageMaker AI 훈련 파이프라인 역할을 수임하여 모델을 훈련합니다.

3. 훈련된 SageMaker AI 모델은 SageMaker AI 모델 레지스트리에 등록됩니다.
4. ML 엔지니어는 ML 엔지니어 사용자 역할을 수임하여 훈련 파이프라인 및 SageMaker AI 모델을 관리합니다.

데이터 권한

SageMaker AI Studio 사용자가 모든 데이터 소스에 액세스할 수 있는 기능은 SageMaker AI IAM 실행 역할과 연결된 권한에 의해 제어됩니다. 연결된 정책은 특정 Amazon S3 버킷 또는 접두사에서 읽기, 쓰기 또는 삭제하고 Amazon RDS 데이터베이스에 연결할 수 있는 권한을 부여할 수 있습니다.

AWS Lake Formation 데이터 액세스

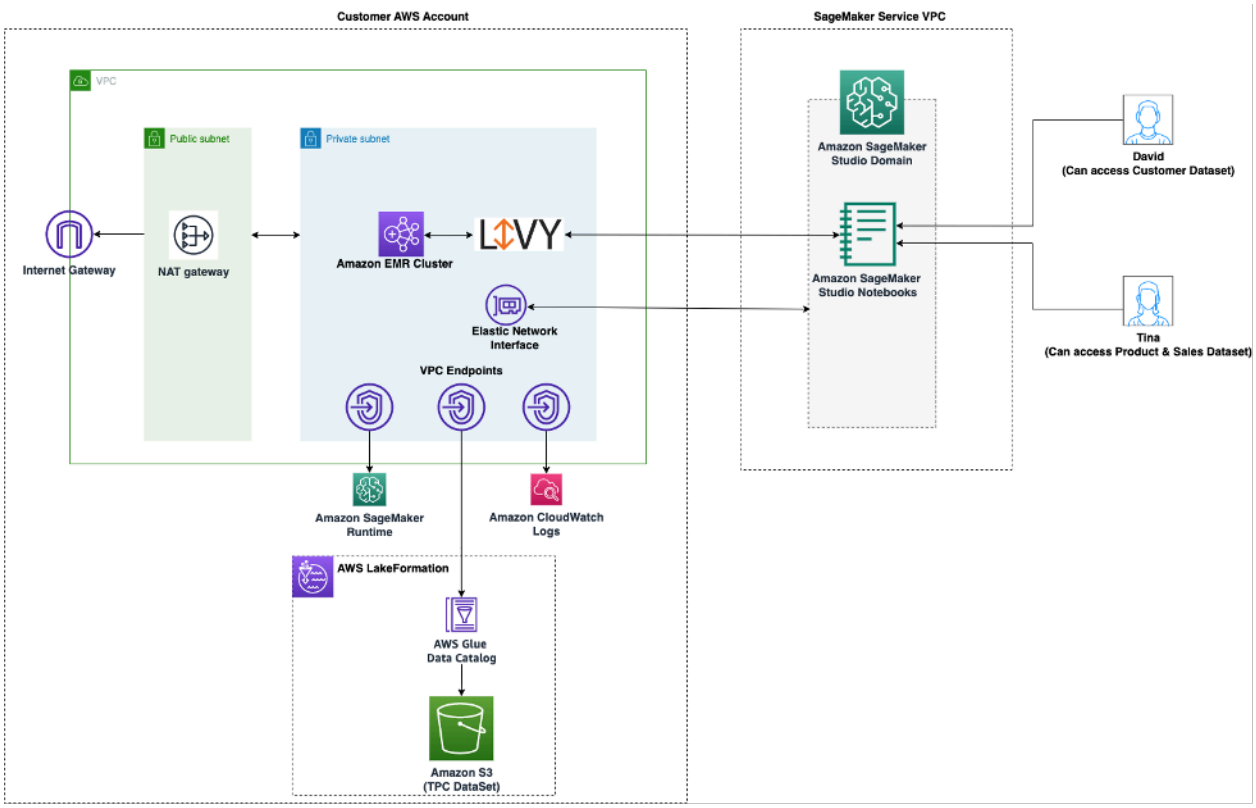
많은 기업에서 사용자가 세밀하게 데이터에 액세스할 수 있도록 AWS Lake Formation에 의해 관리되는 데이터 레이크를 사용하기 시작했습니다. 이러한 관리 대상 데이터의 예로, 관리자는 일부 사용자의 민감한 열을 마스킹하면서도 동일한 기본 테이블에 대한 쿼리를 계속 사용할 수 있습니다.

SageMaker AI Studio에서 Lake Formation을 활용하기 위해 관리자는 SageMaker AI IAM 실행 역할을 로 등록할 수 있습니다`DataLakePrincipals`. 자세한 내용은 Lake Formation 권한 참조 섹션을 참조하세요. 권한이 부여되면 SageMaker AI Studio에서 관리되는 데이터에 액세스하고 쓰는 세 가지 기본 방법이 있습니다.

1. SageMaker AI Studio 노트북에서 사용자는 Amazon Athena와 같은 쿼리 엔진 또는 boto3 위에 빌드된 라이브러리를 활용하여 데이터를 노트북으로 직접 가져올 수 있습니다. for AWS Pandas(이전 SDK에는 awswrangler라고 함)는 널리 사용되는 라이브러리입니다. 다음은 얼마나 원활할 수 있는지를 보여주는 코드 예제입니다.

```
transaction_id = wr.lakeformation.start_transaction(read_only=True)
df = wr.lakeformation.read_sql_query(
    sql=f"SELECT * FROM {table};",
    database=database,
    transaction_id=transaction_id
)
```

2. Amazon에 대한 SageMaker AI Studio 기본 연결을 사용하여 대규모EMR로 데이터를 읽고 씁니다. Apache Livy 및 Amazon EMR 런타임 역할을 사용하여 SageMaker AI Studio는 데이터 액세스 및 처리를 위해 SageMaker AI 실행 IAM 역할(또는 기타 승인된 역할)을 Amazon EMR 클러스터에 전달할 수 있는 네이티브 연결을 구축했습니다. 지침은 Studio에서 Amazon EMR 클러스터에 연결을 참조하세요 up-to-date.



SageMaker Studio에서 Lake Formation에서 관리하는 데이터에 액세스하기 위한 아키텍처

3. AWS Glue 대화형 세션에 대한 SageMaker AI Studio 기본 연결을 사용하여 대규모로 데이터를 읽고 쓸 수 있습니다. SageMaker AI Studio 노트북에는 사용자가에서 대화형으로 명령을 실행할 수 있는 커널이 내장되어 있습니다AWS Glue. 이를 통해 관리되는 데이터 소스에서 대규모로 데이터를 원활하게 읽고 쓸 수 있는 Python, Spark 또는 Ray 백엔드를 확장 가능하게 사용할 수 있습니다. 커널을 사용하면 사용자가 SageMaker 실행 또는 기타 승인된 IAM 역할을 전달할 수 있습니다. 자세한 내용은 AWS Glue 대화형 세션을 사용하여 데이터 준비를 참조하세요.

일반 가드레일

이 섹션에서는 IAM 정책, 리소스 정책, VPC 엔드포인트 정책 및 서비스 제어 정책()을 사용하여 ML 리소스에 거버넌스를 적용하는 데 가장 일반적으로 사용되는 가드레일에 대해 설명합니다SCPs.

특정 인스턴스로 노트북 액세스 제한

이 서비스 제어 정책은 Studio 노트북을 만드는 동안 데이터 과학자가 액세스할 수 있는 인스턴스 유형을 제한하는 데 사용할 수 있습니다. 모든 사용자는 SageMaker AI Studio를 호스팅하는 기본 Jupyter Server 앱을 생성할 수 있는 “시스템” 인스턴스가 필요합니다.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Sid": "LimitInstanceTypesforNotebooks",
             "Effect": "Deny",
             "Action": [
                 "sagemaker:CreateApp"
             ],
             "Resource": "*",
             "Condition": {
                 "ForAnyValue:StringNotLike": {
                     "sagemaker:InstanceTypes": [
                         "ml.c5.large",
                         "ml.m5.large",
                         "ml.t3.medium",
                         "system"
                     ]
                 }
             }
         }
     ]
 }
```

규정 미준수 SageMaker AI Studio 도메인 제한

SageMaker AI Studio 도메인의 경우 다음 서비스 제어 정책을 사용하여 트래픽을 적용해 퍼블릭 인터넷을 통하지 않고 고객의를 통해 고객 리소스에 액세스할 수 있습니다. VPC

```
{
     "Version": "2012-10-17",
     "Statement": [
         {   "Sid": "LockDownStudioDomain",
             "Effect": "Deny",
             "Action": [
                 "sagemaker:CreateDomain"
             ],
             "Resource": "*",
             "Condition": {
                          "StringNotEquals": {"sagemaker:AppNetworkAccessType":
 "VpcOnly"
                 },
```

```
                "Null": {
                        "sagemaker:VpcSubnets": "true",
                        "sagemaker:VpcSecurityGroupIds": "true"
                }
            }
        }
    ]
}
```

승인되지 않은 SageMaker AI 이미지 시작 제한

다음 정책은 사용자가 도메인 내에서 승인되지 않은 SageMaker AI 이미지를 시작하지 못하도록 합니다.f

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "sagemaker:CreateApp"
             ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "ForAllValues:StringNotLike": {
                    "sagemaker:ImageArns":
                        [
                        "arn:aws:sagemaker:*:*:image/{ImageName}"
                        ]
                }
            }
        }
    ]
}
```

SageMaker AI VPC 엔드포인트를 통해서만 노트북 시작

SageMaker AI 컨트롤 플레인의 VPC 엔드포인트 외에도 SageMaker AI는 사용자가 AI SageMaker Studio 노트북 또는 AI SageMaker 노트북 인스턴스에 연결할 수 있도록 VPC 엔드포인트를 지원합니다. SageMaker AI Studio/노트북 인스턴스에 대한 VPC 엔드포인트를 이미 설정한 경우 다음 IAM 조건 키는 SageMaker AI Studio 엔드포인트 또는 SageMaker AI VPC 엔드포인트를 통해 이루어진 경우에만 SageMaker AI Studio 노트북에 대한 연결을 허용합니다API.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Sid": "EnableSageMakerStudioAccessviaVPCEndpoint",
             "Effect": "Allow",
             "Action": [
                 "sagemaker:CreatePresignedDomainUrl",
                 "sagemaker:DescribeUserProfile"
             ],
             "Resource": "*",
             "Condition": {
                 "ForAnyValue:StringEquals": {
                     "aws:sourceVpce": [
                         "vpce-111bbccc",
                         "vpce-111bbddd"
                     ]
                 }
             }
         }
     ]
 }
```

SageMaker AI Studio 노트북 액세스를 제한된 IP 범위로 제한

기업은 SageMaker AI Studio 액세스를 허용되는 특정 기업 IP 범위로 제한하는 경우가 많습니다. `SourceIP` 조건 키가 있는 다음 IAM 정책은 이를 제한할 수 있습니다.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Sid": "EnableSageMakerStudioAccess",
             "Effect": "Allow",
             "Action": [
                 "sagemaker:CreatePresignedDomainUrl",
                 "sagemaker:DescribeUserProfile"
             ],
             "Resource": "*",
             "Condition": {
                 "IpAddress": {
                     "aws:SourceIp": [
```

```
                        "192.0.2.0/24",
                        "203.0.113.0/24"
                    ]
                }
            }
        }
    ]
}
```

SageMaker AI Studio 사용자가 다른 사용자 프로필에 액세스하지 못하도록 방지

관리자는 사용자 프로필을 생성할 때 프로필에 SageMaker AI Studio 사용자 이름에 태그 키가 지정되어 있는지 확인합니다`studiouserid`. 보안 주체(사용자에게 연결된 사용자 또는 역할)도 `studiouserid` 키가 있는 태그를 가져야 합니다(이 태그의 이름은 무엇이든 지정할 수 있으며 `studiouserid`에 국한되지 않음).

그런 다음 SageMaker AI Studio를 시작할 때 사용자가 수임할 역할에 다음 정책을 연결합니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AmazonSageMakerPresignedUrlPolicy",
            "Effect": "Allow",
            "Action": [
                "sagemaker:CreatePresignedDomainUrl"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "sagemaker:ResourceTag/studiouserid": "${aws:PrincipalTag/
studiouserid}"
                }
            }
        }
    ]
}
```

태그 지정 적용

데이터 과학자는 SageMaker AI Studio 노트북을 사용하여 데이터를 탐색하고 모델을 빌드 및 훈련해야 합니다. 노트북에 태그를 적용하면 사용량을 모니터링하고 비용을 제어할 수 있을 뿐만 아니라 소유권 및 감사 가능성을 보장하는 데도 도움이 됩니다.

SageMaker AI Studio 앱의 경우 사용자 프로필에 태그가 지정되어 있는지 확인합니다. 태그는 사용자 프로필에서 앱에 자동으로 전파됩니다. 태그(CLI 및를 통해 지원됨SDK)를 사용하여 사용자 프로필 생성을 적용하려면 관리자 역할에이 정책을 추가하는 것이 좋습니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "EnforceUserProfileTags",
            "Effect": "Allow",
            "Action": "sagemaker:CreateUserProfile",
            "Resource": "*",
            "Condition": {
                "ForAnyValue:StringEquals": {
                    "aws:TagKeys": [
                        "studiouserid"
                    ]
                }
            }
        }
    ]
}
```

훈련 작업 및 처리 작업과 같은 기타 리소스의 경우 다음 정책을 사용하여 태그를 필수로 설정할 수 있습니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "EnforceTagsForJobs",
            "Effect": "Allow",
            "Action": [
                "sagemaker:CreateTrainingJob",
                "sagemaker:CreateProcessingJob",
            ],
```

```
            "Resource": "*",
            "Condition": {
                "ForAnyValue:StringEquals": {
                    "aws:TagKeys": [
                        "studiouserid"
                    ]
                }
            }
        }
    ]
}
```

SageMaker AI Studio의 루트 액세스

SageMaker AI Studio에서 노트북은 기본적으로 호스트 인스턴스에 대한 루트 액세스 권한이 없는 Docker 컨테이너에서 실행됩니다. 마찬가지로 기본 사용자로 실행을 제외하고 컨테이너 내의 다른 모든 사용자 ID 범위는 호스트 인스턴스 자체에서 권한이 없는 사용자IDs로 다시 매핑됩니다. 따라서 권한 상승 위협은 노트북 컨테이너 자체로 제한됩니다.

사용자 지정 이미지를 만들 때는 루트 이외의 권한을 사용자에게 제공하여 더 엄격한 제어(예: 루트 권한으로 원치 않는 프로세스를 실행하지 않도록 하거나 공개적으로 사용 가능한 패키지를 설치하는 등)를 하는 것이 좋습니다. 이 경우 Dockerfile 내에서 루트 사용자가 아닌 사용자로 실행할 이미지를 만들 수 있습니다. 사용자를 루트로 생성하든 루트가 아니든 관계없이 사용자 지정 앱AppImageConfig의 UID/GID of the user is identical to the UID/GID에가 있는지 확인해야 합니다. 그러면 사용자 지정 이미지를 사용하여 앱을 실행하기 위한 SageMaker AI 구성이 생성됩니다. 예를 들어, Dockerfile이 다음과 같이 루트 사용자가 아닌 사용자를 위해 구축된 경우:

```
ARG NB_UID="1000"
ARG NB_GID="100"
...
USER $NB_UID
```

`AppImageConfig` 파일은 GID에서 동일한 UID 및를 언급해야 합니다. `KernelGatewayConfig`

```
{
    "KernelGatewayImageConfig": {
        "FileSystemConfig": {
            "DefaultUid": 1000,
            "DefaultGid": 100
        }
    }
```

```
}
```

사용자 지정 이미지에 허용되는 UID/GID 값은 Studio 이미지의 경우 0/0 및 1000/100입니다. 사용자 지정 이미지 작성 및 관련된 `AppImageConfig` 설정의 예는 이 Github 리포지토리를 참조하세요.

사용자가 이를 조작하지 않도록 `CreateAppImageConfig`하려면 SageMaker AI Studio 노트북 사용자에게 `UpdateAppImageConfig`, 또는 `DeleteAppImageConfig` 권한을 부여하지 마세요.

네트워크 관리

SageMaker AI Studio 도메인을 설정하려면 VPC 네트워크, 서브넷 및 보안 그룹을 지정해야 합니다. VPC 및 서브넷을 지정할 때는 다음 섹션에서 설명하는 사용량과 예상 증가를 IPs 고려하여 할당해야 합니다.

VPC 네트워크 계획

SageMaker AI Studio 도메인에 연결된 고객 VPC 서브넷은 다음 요인에 따라 적절한 클래스리스 도메인 간 라우팅(CIDR) 범위로 생성해야 합니다.

- 사용자 수.
- 사용자당 앱 수.
- 사용자당 고유 인스턴스 유형 수.
- 사용자당 평균 훈련 인스턴스 수.
- 예상 성장률.

SageMaker AI 및 참여 AWS 서비스는 다음과 같은 사용 사례에 대해 탄력적 네트워크 인터페이스(ENI)를 고객 VPC 서브넷에 삽입합니다.

- Amazon은 SageMaker AI 도메인ENI의 EFS 탑재 대상(SageMaker AI 도메인에 연결된 서브넷/가용성 영역당 IP 1개)에 대해를 EFS 주입합니다.
- SageMaker AI Studio는 사용자 프로필 또는 공유 공간에서 사용하는 모든 고유한 인스턴스에 ENI 대해를 주입합니다. 다음과 같은 예가 있습니다.
  - 사용자 프로필이 기본 Jupyter 서버 앱('시스템' 인스턴스 1개), 데이터 과학 앱 및 기본 Python 앱(둘 다 `ml.t3.medium` 인스턴스에서 실행)을 실행하는 경우에 Studio는 두 개의 IP 주소를 주입합니다.
  - 사용자 프로파일이 기본 Jupyter 서버 앱(하나의 '시스템' 인스턴스), Tensorflow GPU 앱(인`ml.g4dn.xlarge`스턴스) 및 데이터 랭글러 앱(인`ml.m5.4xlarge`스턴스)을 실행하는 경우 Studio는 세 개의 IP 주소를 주입합니다.
- 도메인 VPC 서브넷/가용성 영역에 걸쳐 각 VPC 엔드포인트ENI에 대한가 주입됩니다(SageMaker AI VPC 엔드포인트IPs의 경우 4개ECR, S3, 및 같은 참여 서비스 VPC 엔드포인트IPs의 경우 ~6개 CloudWatch).
- SageMaker AI 훈련 및 처리 작업이 동일한 VPC 구성으로 시작되는 경우 각 작업에는 인스턴스당 두 개의 IP 주소가 필요합니다.

> **Note**
>
> VPC 서브넷 및 VPC전용 트래픽과 같은 SageMaker AI Studio에 대한 설정은 SageMaker AI Studio에서 생성된 훈련/처리 작업에 자동으로 전달되지 않습니다. 사용자는 Create*Job을 호출할 때 필요에 따라 VPC 설정 및 네트워크 격리를 설정해야 합니다APIs. 자세한 내용은 인터넷이 연결되지 않은 모드에서 훈련 및 추론 컨테이너 실행을 참조하세요.

시나리오: 데이터 과학자들이 서로 다른 두 가지 인스턴스 유형에 대해 실험을 실행

이 시나리오에서는 SageMaker AI 도메인이 VPC전용 트래픽 모드로 설정되어 있다고 가정합니다. SageMaker AI, SageMaker AI 런타임API, Amazon S3, Amazon와 같은 VPC 엔드포인트가 설정되어 있습니다ECR.

한 데이터 과학자가 서로 다른 두 가지 인스턴스 유형(예: `ml.t3.medium` 및 `ml.m5.large`)에서 실행 중인 Studio 노트북으로 각 인스턴스 유형에서 두 개의 앱을 실행하는 실험을 진행하고 있습니다.

또한 데이터 과학자가 `ml.m5.4xlarge` 인스턴스에서 동일한 VPC 구성으로 훈련 작업을 동시에 실행하고 있다고 가정합니다.

이 시나리오에서 SageMaker AI Studio 서비스는 다음과 ENIs 같이 주입합니다.

표 1 - 실험 시나리오를 VPC 위해 고객에게 ENIs 주입

개체	대상	ENI 주입됨	참고	수준
EFS 탑재 대상	VPC 서브넷	3	AZs/서브넷 3개	도메인
VPC 엔드포인트	VPC 서브넷	30	VPCE 각각 10개가 있는 AZs/서브넷 3개	도메인
Jupyter 서버	VPC 서브넷	1개	인스턴스당 IP 1개	User
KernelGateway 앱	VPC 서브넷	2개	인스턴스 유형당 IP 1개	User
학습	VPC 서브넷	2	훈련 인스턴스 IPs당 2개	User

개체	대상	ENI 주입됨	참고	수준
			EFA를 사용하는 경우 훈련 인스턴스IPs당 5개	

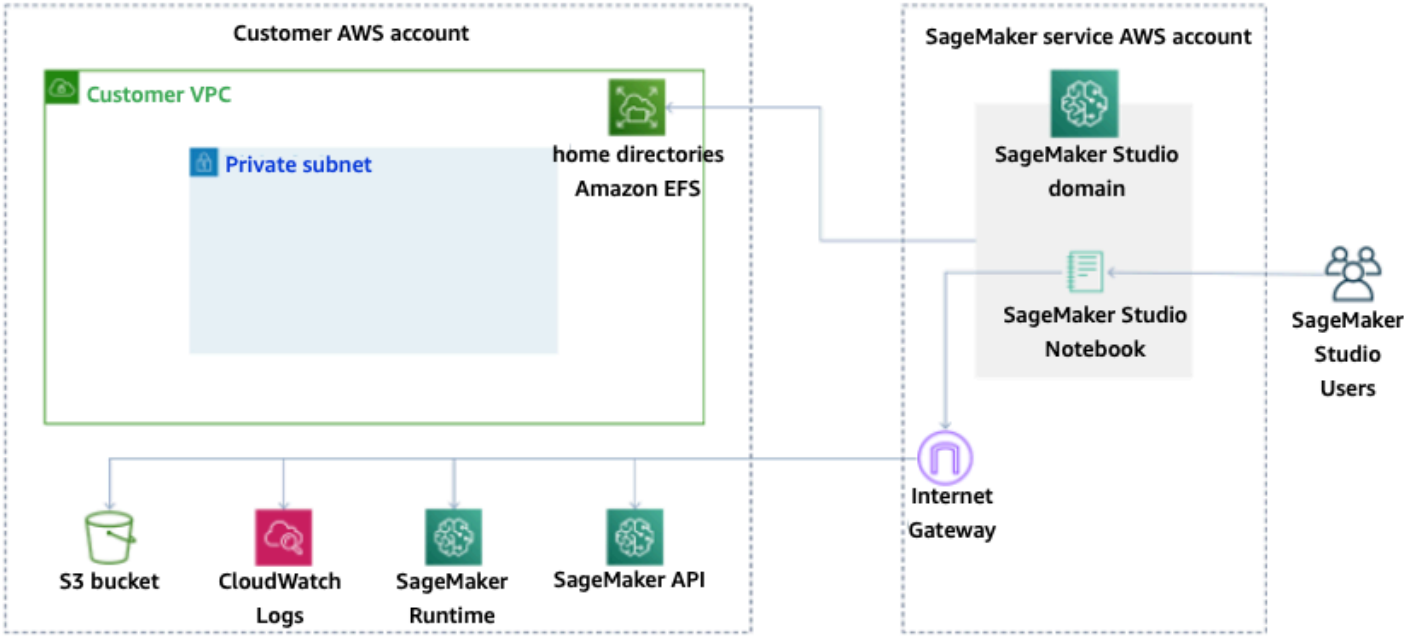
이 시나리오의 경우, 고객에서 총 38개가 IPs 소비되며, VPC 여기서 33IPs개는 도메인 수준에서 사용자 간에 공유되고 5개IPs는 사용자 수준에서 소비됩니다. 이 도메인에 이러한 활동을 동시에 수행하는 유사한 사용자 프로필이 있는 사용자가 100명 있는 경우 도메인 수준 IP 소비 외에도 IPs 사용자 수준에서 5 x 100 = 500을 소비합니다. 즉, 서브넷IPs당 11개로 총 511개 입니다IPs. 이 시나리오에서는 1024개의 IP 주소를 할당할 /22CIDR로 VPC 서브넷을 생성해야 하며, 확장할 공간이 있습니다.

VPC 네트워크 옵션

SageMaker AI Studio 도메인은 다음 옵션 중 하나를 사용하여 VPC 네트워크 구성을 지원합니다.

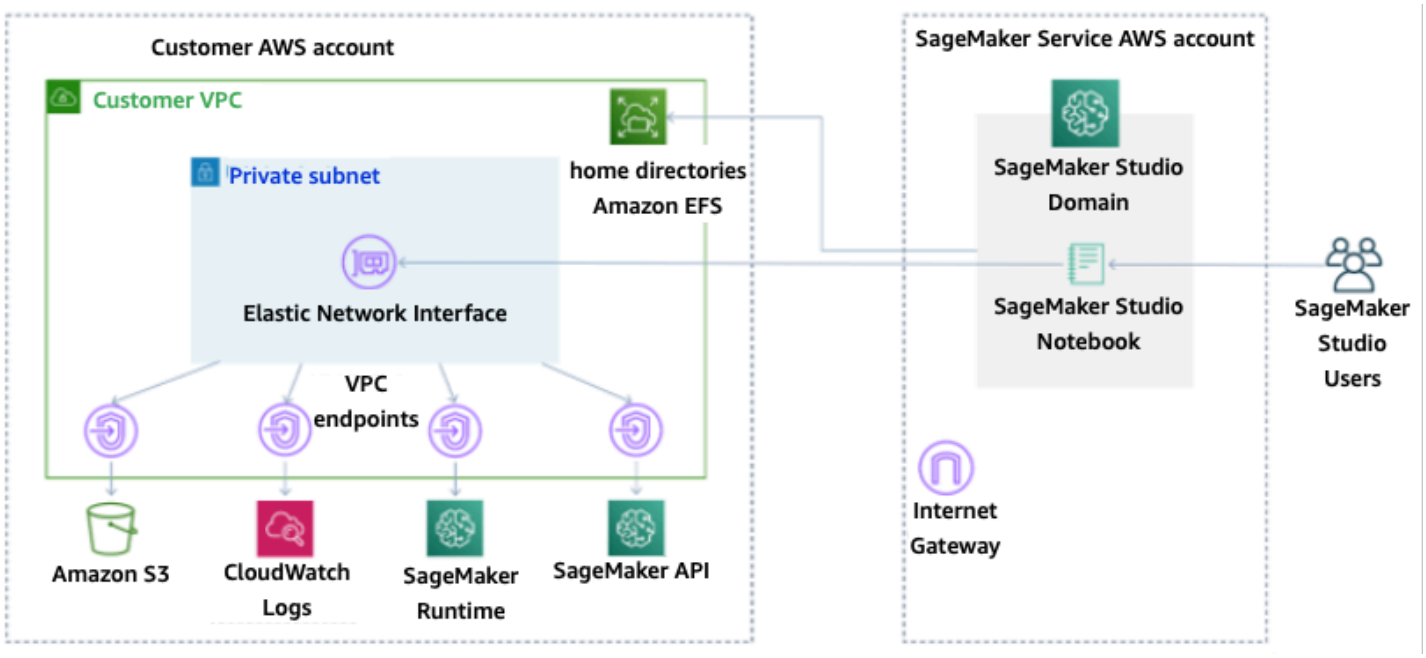
- 퍼블릭 인터넷 전용
- VPC 전용

퍼블릭 인터넷 전용 옵션을 사용하면 다음 다이어그램과 같이 SageMaker AI API 서비스가 SageMaker AI 서비스 계정에서 VPC관리하는에 프로비저닝된 인터넷 게이트웨이를 통해 퍼블릭 인터넷을 사용할 수 있습니다.



기본 모드: SageMaker AI 서비스 계정을 통한 인터넷 액세스

VPC 유일한 옵션은 SageMaker AI 서비스 계정에서 VPC 관리하는의 인터넷 라우팅을 비활성화하고 다음 다이어그램과 같이 고객이 VPC 엔드포인트를 통해 라우팅되도록 트래픽을 구성할 수 있도록 허용합니다.



VPC 전용 모드: SageMaker AI 서비스 계정을 통한 인터넷 액세스 없음

VPC 전용 모드로 설정된 도메인의 경우 기본 인스턴스의 완전한 격리를 위해 사용자 프로필당 보안 그룹을 설정합니다. AWS 계정의 각 도메인에는 자체 VPC 구성 및 인터넷 모드가 있을 수 있습니다. VPC 네트워크 구성 설정에 대한 자세한 내용은 의 SageMaker AI Studio 노트북을 외부 리소스에 연결을 참조VPC하세요.

제한 사항

- SageMaker AI Studio 도메인을 생성한 후에는 새 서브넷을 도메인에 연결할 수 없습니다.
- VPC 네트워크 유형(퍼블릭 인터넷 전용 또는 VPC 전용)은 변경할 수 없습니다.

데이터 보호

ML 워크로드를 설계하기 전에, 보안에 영향을 미치는 기본적인 방법들을 준비해야 합니다. 예를 들어, 데이터 분류는 민감도 수준에 따라 데이터를 분류하는 방법을 제공하며, 암호화는 무단 액세스를 할 경우 데이터를 이해할 수 없게 만들어 데이터를 보호합니다. 이러한 방법이 중요한 이유는, 조작 부주의 방지 또는 규제 의무 사항 준수와 같은 목표의 달성을 지원하기 때문입니다.

SageMaker AI Studio는 저장 및 전송 중 데이터를 보호하기 위한 여러 기능을 제공합니다. 그러나 AWS 공동 책임 모델에 설명된 대로 고객은 AWS 글로벌 인프라에서 호스팅되는 콘텐츠에 대한 제어를 유지할 책임이 있습니다. 이 섹션에서는 고객이 이러한 기능을 사용하여 데이터를 보호하는 방법에 대해 설명합니다.

저장 데이터 보호

모델 구축 데이터 및 모델 아티팩트와 함께 SageMaker AI Studio 노트북을 보호하기 위해 SageMaker AI는 노트북과 훈련 및 배치 변환 작업의 출력을 암호화합니다. SageMaker AI는 기본적으로 AWS Amazon S3용 관리형 키를 사용하여 노트북을 암호화합니다. 이 Amazon S3용 AWS 관리형 키는 교차 계정 액세스를 위해 공유할 수 없습니다. 교차 계정 액세스의 경우 교차 계정 액세스를 위해 공유할 수 있도록 SageMaker AI 리소스를 생성하는 동안 고객 관리형 키를 지정합니다.

SageMaker AI Studio를 사용하면 데이터를 다음 위치에 저장할 수 있습니다.

- S3 버킷 - 공유 가능한 노트북이 활성화되면 SageMaker AI Studio는 S3 버킷에서 노트북 스냅샷과 메타데이터를 공유합니다.
- EFS 볼륨 - SageMaker AI Studio는 노트북 및 데이터 파일을 저장하기 위해 도메인에 EFS 볼륨을 연결합니다. 이 EFS 볼륨은 도메인이 삭제된 후에도 유지됩니다.
- EBS 볼륨 - 노트북이 실행되는 인스턴스에 EBS 연결됩니다. 이 볼륨은 인스턴스 기간 동안 지속됩니다.

를 사용한 유휴 암호화 AWS KMS

- AWS KMS 키를 전달하여 노트북, 훈련, 튜닝, 배치 변환 작업 및 엔드포인트에 연결된 EBS 볼륨을 암호화할 수 있습니다.
- KMS 키를 지정하지 않으면 SageMaker AI는 시스템 관리형 KMS 키로 운영 체제(OS) 볼륨과 ML 데이터 볼륨을 모두 암호화합니다.

- 규정 준수를 위해 KMS 키로 암호화해야 하는 민감한 데이터는 ML 스토리지 볼륨 또는 Amazon S3에 저장해야 하며, 둘 다 지정한 KMS 키를 사용하여 암호화할 수 있습니다.

전송 중 데이터 보호

SageMaker AI Studio는 ML 모델 아티팩트 및 기타 시스템 아티팩트가 전송 중 및 저장 중에 암호화되도록 합니다. SageMaker AI API 및 콘솔에 대한 요청은 보안(SSL) 연결을 통해 이루어집니다. 일부 인트라 네트워크의 전송 중 데이터(서비스 플랫폼 내)는 암호화되지 않습니다. 여기에는 다음이 포함됩니다.

- 서비스 컨트롤 플레인과 훈련 작업 인스턴스(고객 데이터 아님) 간의 명령 및 제어 통신.
- 분산 처리 및 훈련 작업(인트라 네크워크)의 노드 간 통신.

하지만 훈련 클러스터의 노드 간 통신을 암호화하도록 선택할 수도 있습니다. 컨테이너 간 트래픽 암호화를 활성화하면 훈련 시간이 늘어날 수 있는데, 분산된 딥 러닝 알고리즘을 사용하는 경우 특히 더 그렇습니다.

기본적으로 Amazon SageMaker AI는 Amazon에서 훈련 작업을 실행VPC하여 데이터를 안전하게 유지합니다. 프라이빗을 구성하여 훈련 컨테이너와 데이터를 보호하기 위해 다른 수준의 보안을 추가할 수 있습니다VPC. 또한 SageMaker AI Studio 도메인이 VPC 전용 모드로 실행되도록 구성하고 인터넷을 통해 트래픽을 내보내지 않고 프라이빗 네트워크를 통해 트래픽을 라우팅하도록 VPC 엔드포인트를 설정할 수 있습니다.

데이터 보호 가드레일

유휴 SageMaker AI 호스팅 볼륨 암호화

온라인 추론을 위해 SageMaker AI 엔드포인트를 호스팅하는 동안 암호화를 적용하려면 다음 정책을 사용합니다.

```
{
   "Version": "2012-10-17",
   "Statement": [
     {
         "Sid": "Encryption",
         "Effect": "Allow",
         "Action": [
```

```
            "sagemaker:CreateEndpointConfig"
        ],
        "Resource": "*",
        "Condition": {
            "Null": {
                "sagemaker:VolumeKmsKey": "false"
            }
        }
    }
  ]
}
```

모델 모니터링 중에 사용되는 S3 버킷 암호화하기

모델 모니터링은 SageMaker AI 엔드포인트로 전송된 데이터를 캡처하여 S3 버킷에 저장합니다. 데이터 캡처 구성을 설정할 때, S3 버킷을 암호화해야 합니다. 현재 이에 대한 보완 제어 기능은 없습니다.

모델 모니터링 서비스는 엔드포인트 출력을 캡처하는 것 외에도, 사전 지정된 기준선을 기준으로 드리프트를 확인합니다. 드리프트를 모니터링하는 데 사용되는 출력 및 중간 스토리지 볼륨을 암호화해야 합니다.

```
{
   "Version": "2012-10-17",
   "Statement": [
     {
         "Sid": "Encryption",
         "Effect": "Allow",
         "Action": [
             "sagemaker:CreateMonitoringSchedule",
             "sagemaker:UpdateMonitoringSchedule"
         ],
         "Resource": "*",
         "Condition": {
             "Null": {
                 "sagemaker:VolumeKmsKey": "false",
                 "sagemaker:OutputKmsKey": "false"
             }
         }
     }
   ]
 }
```

SageMaker AI Studio 도메인 스토리지 볼륨 암호화

Studio 도메인에 연결된 스토리지 볼륨에 암호화를 적용합니다. 이 정책은 사용자가 스튜디오 도메인에 연결된 스토리지 볼륨을 암호화CMK하기 위해를 제공해야 합니다.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Sid": "EncryptDomainStorage",
             "Effect": "Allow",
             "Action": [
                 "sagemaker:CreateDomain"
             ],
             "Resource": "*",
             "Condition": {
                 "Null": {
                     "sagemaker:VolumeKmsKey": "false"
                 }
             }
         }
     ]
 }
```

노트북을 공유하는 데 사용되는 S3에 저장된 데이터를 암호화합니다.

이는 SageMaker AI Studio 도메인의 사용자 간에 노트북을 공유하는 데 사용되는 버킷에 저장된 모든 데이터를 암호화하는 정책입니다.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Sid": "EncryptDomainSharingS3Bucket",
             "Effect": "Allow",
             "Action": [
                 "sagemaker:CreateDomain",
                 "sagemaker:UpdateDomain"
             ],
             "Resource": "*",
             "Condition": {
                 "Null": {
```

```
                        "sagemaker:DomainSharingOutputKmsKey": "false"
                    }
                }
            }
        ]
}
```

제한 사항

- 도메인이 생성되면 연결된 EFS 볼륨 스토리지를 사용자 지정 AWS KMS 키로 업데이트할 수 없습니다.
- 키가 생성된 후에는 KMS 키로 훈련/처리 작업 또는 엔드포인트 구성을 업데이트할 수 없습니다.

로깅 및 모니터링

컴파일 작업, 처리 작업, 훈련 작업, 엔드포인트, 변환 작업, 노트북 인스턴스 및 노트북 인스턴스 수명 주기 구성을 디버깅하는 데 도움이 되도록 알고리즘 컨테이너, 모델 컨테이너 또는 노트북 인스턴스 수명 주기 구성이 stdout 또는 stderr로 전송하는 모든 항목도 Amazon CloudWatch Logs로 전송됩니다. 원시 데이터를 CloudWatch수집하고 읽기 가능하며 실시간에 가까운 지표로 처리하는 Amazon을 사용하여 SageMaker AI Studio를 모니터링할 수 있습니다. 이러한 통계는 15개월 동안 보관되므로, 기록 정보에 액세스하고 웹 애플리케이션 또는 서비스가 어떻게 실행되고 있는지 전체적으로 더 잘 파악할 수 있습니다.

를 사용하여 로깅 CloudWatch

데이터 과학 프로세스는 기본적으로 실험적이고 반복적이므로 노트북 사용, 훈련/처리 작업 실행 시간, 훈련 지표, 간접 호출 대기 시간과 같은 엔드포인트 서비스 지표와 같은 활동을 기록하는 것이 필수적입니다. 기본적으로 SageMaker AI는 지표를 CloudWatch 로그에 게시하며, 이러한 로그는를 사용하여 고객 관리형 키로 암호화할 수 있습니다 AWS KMS.

또한 VPC 엔드포인트를 사용하여 퍼블릭 인터넷을 사용하지 CloudWatch 않고 로 로그를 보낼 수 있습니다. 특정 임곗값을 주시하다가 해당 임곗값이 충족될 때 알림을 전송하거나 조치를 취하도록 경보를 설정할 수도 있습니다. 자세한 내용은 Amazon CloudWatch 사용 설명서를 참조하세요.

SageMaker AI는 아래에 Studio에 대한 단일 로그 그룹을 생성합니다`/aws/sagemaker/studio`. 각 사용자 프로필 및 앱은 이 로그 그룹 아래에 고유한 로그 스트림을 가지며 생명 주기 구성 스크립트에도 자체 로그 스트림이 있습니다. 예를 들어 Jupyter 서버 앱이 있고 수명 주기 스크립트가 첨부된 'studio-user'라는 사용자 프로필과 데이터 과학 커널 게이트웨이 앱의 로그 스트림은 다음과 같습니다.

```
/aws/sagemaker/studio/<domain-id>/studio-user/JupyterServer/default
```

```
/aws/sagemaker/studio/<domain-id>/studio-user/JupyterServer/default/
LifecycleConfigOnStart
```

```
/aws/sagemaker/studio/<domain-id>/studio-user/KernelGateway/datascience-app
```

SageMaker AI가 사용자를 대신하여 CloudWatch 에 로그를 전송하려면 Training/Processing/Transform 작업 호출자에게 다음 권한이 필요합니다APIs.

```
{
     "Version": "2012-10-17",
     "Statement": [
```

```
        {
            "Action": [
                "logs:CreateLogDelivery",
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:DeleteLogDelivery",
                "logs:Describe*",
                "logs:GetLogEvents",
                "logs:GetLogDelivery",
                "logs:ListLogDeliveries",
                "logs:PutLogEvents",
                "logs:PutResourcePolicy",
                "logs:UpdateLogDelivery"
            ],
            "Resource": "*",
            "Effect": "Allow"
        }
    ]
}
```

사용자 지정 AWS KMS 키로 이러한 로그를 암호화하려면 먼저 CloudWatch 서비스가 키를 암호화하고 복호화할 수 있도록 키 정책을 수정해야 합니다. 로그 암호화 AWS KMS 키를 생성한 후 다음을 포함하도록 키 정책을 수정합니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "logs.region.amazonaws.com"
            },
            "Action": [
                "kms:Encrypt*",
                "kms:Decrypt*",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:Describe*"
            ],
            "Resource": "*",
            "Condition": {
                "ArnLike": {
```

```
                    "kms:EncryptionContext:aws:logs:arn": "arn:aws:logs:region:account-
id:*"
                }
            }
        }
    ]
}
```

언제든지 ArnEquals를 사용하고 암호화하려는 CloudWatch 로그에 대한 특정 Amazon 리소스 이름(ARN)을 제공할 수 있습니다. 여기서는 간소화를 위해 이 키를 사용하여 계정의 모든 로그를 암호화할 수 있음을 보여줍니다. 또한 훈련, 처리 및 모델 엔드포인트는 인스턴스 CPU 및 메모리 사용률, 호스팅 호출 지연 시간 등에 대한 지표를 게시합니다. 특정 임계값을 초과할 때 이벤트를 관리자에게 알리SNS도록 Amazon을 추가로 구성할 수 있습니다. 훈련 및 처리의 소비자는 다음 권한을가APIs져야 합니다.

```
{
     "Version": "2012-10-17",
     "Statement": [
         {
             "Action": [
                 "cloudwatch:DeleteAlarms",
                 "cloudwatch:DescribeAlarms",
                 "cloudwatch:GetMetricData",
                 "cloudwatch:GetMetricStatistics",
                 "cloudwatch:ListMetrics",
                 "cloudwatch:PutMetricAlarm",
                 "cloudwatch:PutMetricData",
                 "sns:ListTopics"
             ],
             "Resource": "*",
             "Effect": "Allow",
             "Condition": {
                 "StringLike": {
                     "cloudwatch:namespace": "aws/sagemaker/*"
                 }
             }

         },
         {
             "Action": [
                 "sns:Subscribe",
                 "sns:CreateTopic"
```

```
            ],
            "Resource": [
                "arn:aws:sns:*:*:*SageMaker*",
                "arn:aws:sns:*:*:*Sagemaker*",
                "arn:aws:sns:*:*:*sagemaker*"
            ],
            "Effect": "Allow"
        }
    ]
}
```

를 사용한 감사 AWS CloudTrail

규정 준수 태세를 개선하려면 모든 APIs를 감사하십시오 AWS CloudTrail. 기본적으로 모든 SageMaker AIAPIs는 로 로깅됩니다AWS CloudTrail. 활성화하는 데 추가 IAM 권한이 필요하지 않습니다 CloudTrail.

`InvokeEndpoint` 및를 제외한 모든 SageMaker AI 작업은에 의해 `InvokeEndpointAsync`로깅 CloudTrail 되고 작업에 문서화됩니다. 예를 들어 `CreateTrainingJob`, `CreateEndpoint`및 `CreateNotebookInstance` 작업에 대한 호출은 CloudTrail 로그 파일에 항목을 생성합니다.

모든 CloudTrail 이벤트 항목에는 요청을 생성한 사람에 대한 정보가 포함됩니다. 보안 인증 정보를 이용하면 다음을 쉽게 판단할 수 있습니다.

- 요청을 루트로 했는지 아니면 AWS IAM 사용자 보안 인증으로 했는지 여부.
- 역할 또는 페더레이션 사용자에 대한 임시 보안 인증을 사용하여 요청이 생성되었는지 여부.
- 요청이 다른 AWS 서비스에 의해 이루어졌는지 여부. 예제 이벤트는 Log SageMaker AI API Calls with CloudTrail documentation를 참조하세요.

기본적으로는 사용자 프로필의 Studio 실행 역할 이름을 각 이벤트의 식별자로 CloudTrail 로깅합니다. 이는 각 사용자에게 고유한 실행 역할이 있는 경우 작동합니다. 여러 사용자가 동일한 실행 역할을 공유하는 경우 `sourceIdentity` 구성을 사용하여 Studio 사용자 프로필 이름을에 전파할 수 있습니다 CloudTrail. Amazon SageMaker AI Studio에서 사용자 리소스 액세스 모니터링을 참조하여 `sourceIdentity` 기능을 활성화합니다. 공유 공간에서 모든 작업은 공간을 소스ARN로 지칭하며를 통해 감사할 수 없습니다`sourceIdentity`.

비용 분담

SageMaker AI Studio는 관리자가 개별 도메인, 공유 공간 및 사용자의 지출을 추적하는 데 도움이 되는 기능을 내장했습니다.

자동 태그 지정

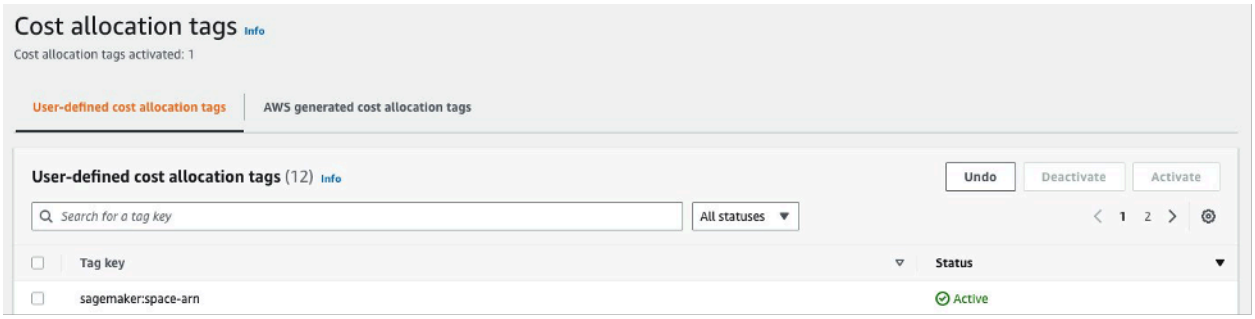
SageMaker 이제 AI Studio는 훈련 작업, 처리 작업 및 커널 앱과 같은 새 SageMaker 리소스에 해당를 사용하여 자동으로 태그를 지정합니다`sagemaker:domain-arn`. 보다 세분화된 수준에서 SageMaker AI는 리소스의 보안 주체 생성자를 지정`sagemaker:space-arn`하기 위해 `sagemaker:user-profile-arn` 또는 로 리소스에 태그를 지정합니다.

SageMaker AI 도메인 EFS 볼륨에는 도메인 값이 `ManagedByAmazonSageMakerResource` 인 키로 태그가 지정됩니다ARN. 여기에는 사용자 레벨별 스페이스 사용량을 파악할 수 있는 세분화된 태그가 없습니다. 관리자는 사용자 지정 모니터링을 위해 EFS 볼륨을 EC2 인스턴스에 연결할 수 있습니다.

비용 모니터링

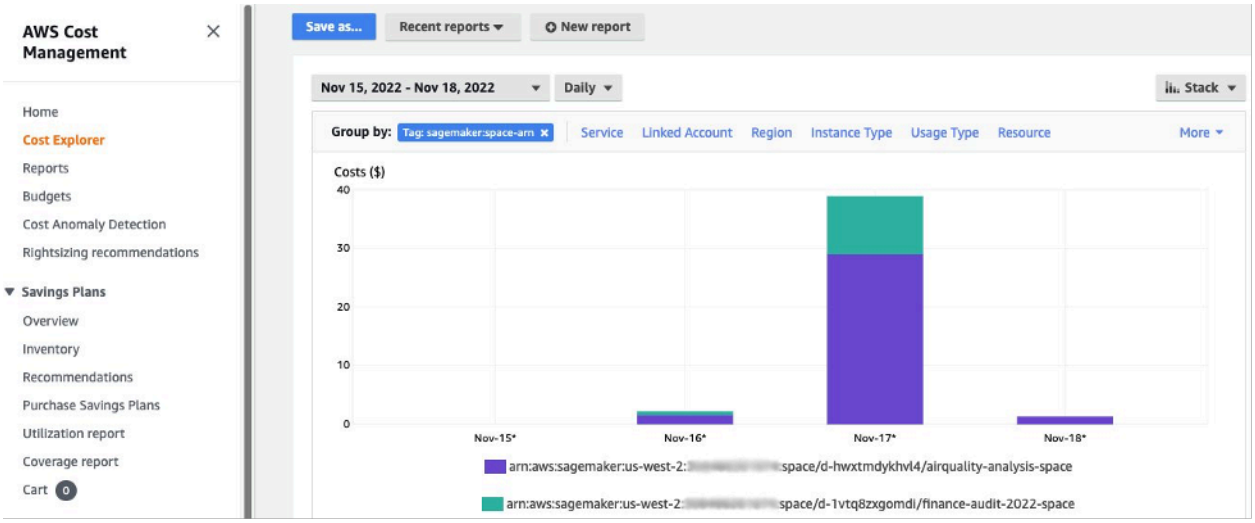
자동 태그를 사용하면 관리자가 AWS Cost Explorer 및와 같은 out-of-the-box 솔루션과 비용 및 사용 보고서()의 데이터를 기반으로 구축된 사용자 지정 솔루션을 통해 ML 지출을 추적AWS Budgets, 보고 및 모니터링할 수 있습니다CURs. AWS

첨부된 태그를 비용 분석에 사용하려면 먼저 AWS Billing 콘솔의 비용 할당 태그 섹션에서 태그를 활성화해야 합니다. 태그가 비용 할당 태그 패널에 표시되는 데 최대 24시간이 걸릴 수 있으므로 활성화하기 전에 SageMaker AI 리소스를 생성해야 합니다.



Cost Explorer에서 비용 할당 태그로 ARN 활성화된 공간

비용 할당 태그를 활성화하면 AWS 가 태그가 지정된 리소스를 추적하기 시작하고 24~48시간 후에는 태그가 비용 탐색기에 선택 가능한 필터로 표시됩니다.



샘플 도메인의 공유 스페이스별로 그룹화된 비용

비용 관리

첫 번째 SageMaker AI Studio 사용자가 온보딩되면 SageMaker AI는 도메인에 대한 EFS 볼륨을 생성합니다. 노트북과 데이터 파일이 사용자의 홈 디렉터리에 저장되면이 EFS 볼륨에 대한 스토리지 비용이 발생합니다. 사용자가 Studio 노트북을 시작하면, Studio 노트북은 노트북을 실행하는 컴퓨팅 인스턴스용으로 시작됩니다. 비용에 대한 자세한 설명은 Amazon SageMaker AI 요금을 참조하세요.

관리자는 공통 가드레일 섹션에 언급된 IAM 정책을 사용하여 사용자가 실행할 수 있는 인스턴스 목록을 지정하여 컴퓨팅 비용을 제어할 수 있습니다. 또한 고객이 SageMaker AI Studio 자동 종료 확장을 사용하여 유휴 앱을 자동으로 종료하여 비용을 절감하는 것이 좋습니다. 이 서버 확장 프로그램은 사용자 프로필별로 실행 중인 앱을 정기적으로 폴링하고 관리자가 설정한 제한 시간에 따라 유휴 앱을 종료합니다.

도메인의 모든 사용자에 대해 이 확장을 설정하려면 사용자 지정 섹션에 설명된 대로 수명 주기 구성을 사용할 수 있습니다. 또한 확장 검사기를 사용하여 모든 도메인 사용자에게 확장이 설치되어 있는지 확인할 수도 있습니다.

사용자 지정

수명 주기 구성

수명 주기 구성은 새 SageMaker AI Studio 노트북 시작과 같은 SageMaker AI Studio 수명 주기 이벤트에서 시작하는 쉘 스크립트입니다. 이러한 쉘 스크립트를 사용하여 사용자 지정 패키지 설치, 비활성 노트북 앱의 자동 종료를 위한 Jupyter 확장, Git 구성 설정과 같은 SageMaker AI Studio 환경에 대한 사용자 지정을 자동화할 수 있습니다. 수명 주기 구성을 빌드하는 방법에 대한 자세한 지침은이 블로그: 수명 주기 구성을 사용하여 Amazon SageMaker AI Studio 사용자 지정을 참조하세요.

SageMaker AI Studio 노트북용 사용자 지정 이미지

Studio 노트북에는 Amazon SageMaker AI PythonSDK과 최신 버전의 IPython 런타임 또는 커널로 구성된 미리 빌드된 이미지 세트가 함께 제공됩니다. 이 기능을 사용하면 사용자 지정 이미지를 Amazon SageMaker AI 노트북으로 가져올 수 있습니다. 이렇게 하면 도메인에서 인증된 모든 사용자가 해당 이미지를 사용할 수 있습니다.

개발자와 데이터 과학자는 아래의 다양한 사용 사례에 맞는 사용자 지정 이미지를 요구할 수 있습니다.

- MXNet PyTorch, 등과 같은 인기 있는 ML 프레임워크의 특정 버전 또는 최신 버전 TensorFlow에 대한 액세스.
- 빠른 반복 및 모델 훈련을 위해 로컬에서 개발된 사용자 지정 코드 또는 알고리즘을 SageMaker AI Studio 노트북에 가져옵니다.
- 를 통해 데이터 레이크 또는 온프레미스 데이터 스토어에 액세스합니다APIs. 관리자는 이미지에 해당 드라이버를 포함해야 합니다.
- (R, Julia 또는 기타) 이외의 백엔드 런타임IPython(커널이라고도 함https://github.com/aws-samples/sagemaker-studio-custom-image-samples)에 대한 액세스. 설명된 방법을 사용하여 사용자 지정 커널을 설치할 수도 있습니다.

사용자 지정 이미지를 빌드하는 방법에 대한 자세한 지침은 사용자 지정 SageMaker AI 이미지 생성을 참조하세요.

JupyterLab 확장

SageMaker AI Studio JuypterLab 3 Notebook을 사용하면 계속 증가하는 오픈 소스 JupyterLab 확장 커뮤니티를 활용할 수 있습니다. 이 섹션에서는 SageMaker AI 개발자 워크플로에 자연스럽게 맞는 몇 가지를 강조하지만 사용 가능한 익스텐션을 찾아보거나 자체 확장을 생성하는 것이 좋습니다.

JupyterLab 이제 3을 사용하면 확장을 패키징하고 설치하는 프로세스가 크게 쉬워집니다. bash 스크립트를 통해 앞서 언급한 확장을 설치할 수 있습니다. 예를 들어 SageMaker AI Studio에서 Studio 시작 관리자에서 시스템 터미널을 열고 다음 명령을 실행합니다. 또한 수명 주기 구성을 사용하여 이러한 확장의 설치를 자동화하여 Studio를 다시 시작해도 유지되도록 할 수 있습니다. 도메인의 모든 사용자에 대해, 또는 개별 사용자 수준에서 이를 구성할 수 있습니다.

예를 들어 Amazon S3 파일 브라우저용 확장을 설치하려면 시스템 터미널에서 다음 명령을 실행하고 브라우저를 새로 고쳐야 합니다.

```
conda init
conda activate studio
pip install jupyterlab_s3_browser
jupyter serverextension enable --py jupyterlab_s3_browser
conda deactivate
restart-jupyter-server
```

이전 버전과의 호환성을 위해 JupyterLab 노트북 버전 1과 버전 3 모두에서 작동하는 수명 주기 구성을 작성하는 방법을 포함하여 확장 관리에 대한 자세한 내용은 설치 JupyterLab 및 Jupyter Server 확장을 참조하세요.

Git 리포지토리

SageMaker AI Studio에는 사용자가 Git 리포지토리URL의 사용자 지정을 입력하고, EFS 디렉터리에 복제하고, 변경 사항을 푸시하고, 커밋 기록을 볼 수 있도록 Jupyter Git 확장이 사전 설치되어 있습니다. 관리자는 도메인 수준에서 제안된 Git 리포지토리를 구성하여 최종 사용자에게 드롭다운 선택 항목으로 표시되도록 할 수 있습니다. 지침은 Studio에 제안된 Git 리포지토리 연결을 참조하세요. up-to-date

리포지토리가 비공개인 경우, 이 확장은 표준 Git 설치를 사용하여 터미널에 보안 인증 정보를 입력하도록 사용자에게 요청합니다. 또는 사용자가 개별 EFS 디렉터리에 ssh 자격 증명을 저장하여 더 쉽게 관리할 수 있습니다.

Conda 환경

SageMaker AI Studio 노트북은 Amazon을 영구 스토리지 계층EFS으로 사용합니다. 데이터 과학자는 영구 스토리지를 사용하여 사용자 지정 conda 환경을 만들고 이러한 환경을 사용하여 커널을 생성할 수 있습니다. 이러한 커널은에서 지원되며 커널EFS, 앱 또는 Studio 재시작 간에 영구적입니다. Studio는 유효한 모든 환경을 커널로 KernelGateway 자동으로 선택합니다.

conda 환경을 만드는 과정은 데이터 과학자에게 간단한 일이지만, 커널이 커널 선택기에 채워지는 데는 약 1분이 걸립니다. 환경을 만들려면 시스템 터미널에서 다음을 실행하세요.

```
mkdir -p ~/.conda/envs
conda create --yes -p ~/.conda/envs/custom
conda activate ~/.conda/envs/custom
conda install -y ipykernel
conda config --add envs_dirs ~/.conda/envs
```

자세한 지침은 Amazon Studio 노트북에서 Python 패키지를 관리하는 네 가지 접근 방식의 Studio EFS 볼륨 섹션의 Conda 환경 지속을 참조하세요. SageMaker

결론

이 백서에서는 플랫폼 관리자가 SageMaker AI Studio 플랫폼을 설정하고 관리할 수 있도록 운영 모델, 도메인 관리, 자격 증명 관리, 권한 관리, 네트워크 관리, 로깅, 모니터링 및 사용자 지정과 같은 영역의 몇 가지 모범 사례를 검토했습니다.

부록

멀티테넌시 비교

표 2 - 멀티테넌시 비교

다중 도메인	다중 계정	단일 도메인 내의 속성 기반 액세스 제어(ABAC)
리소스 격리는 태그를 사용하여 이루어집니다. SageMaker AI Studio는 도메인 ARN 및 사용자 프로필/스페이스를 사용하여 모든 리소스에 태그를 자동으로 지정합니다ARN.	각 테넌트는 자체 계정에 있으므로 절대적인 리소스 격리가 있습니다.	태그를 사용하여 리소스 격리가 이루어집니다. 사용자는에 대해 생성된 리소스의 태그 지정을 관리해야 합니다ABAC.
목록은 태그로 제한할 수 APIs 없습니다. 리소스의 UI 필터링은 공유 공간에서 수행되지만 AWS CLI 또는 Boto3를 통해 수행된 API 호출 나열SDK은 리전 전반의 리소스를 나열합니다.	테넌트는 전용 계정에 있으므로 목록 APIs 격리도 가능합니다.	목록은 태그로 제한할 수 APIs 없습니다. AWS CLI 또는 Boto3를 통해 수행된 API 호출을 나열SDK하면 리전 전반의 리소스가 나열됩니다.
SageMaker 도메인를 비용 할당 태그ARN로 사용하여 테넌트당 AI Studio 컴퓨팅 및 스토리지 비용을 쉽게 모니터링할 수 있습니다.	SageMaker 테넌트당 AI Studio 컴퓨팅 및 스토리지 비용은 전용 계정으로 쉽게 모니터링할 수 있습니다.	SageMaker 테넌트당 AI Studio 컴퓨팅 비용은 사용자 지정 태그를 사용하여 계산해야 합니다. SageMaker 모든 테넌트가 동일한 EFS 볼륨을 공유하므로 도메인당 AI Studio 스토리지 비용을 모니터링할 수 없습니다.

다중 도메인	다중 계정	단일 도메인 내의 속성 기반 액세스 제어(ABAC)
서비스 할당량은 계정 수준에서 설정되므로 단일 테넌트가 여전히 모든 리소스를 사용할 수 있습니다.	서비스 할당량은 각 테넌트의 계정 수준에서 설정할 수 있습니다.	서비스 할당량은 계정 수준에서 설정되므로 단일 테넌트가 여전히 모든 리소스를 사용할 수 있습니다.
코드형 인프라(IaC) 또는 서비스 카탈로그를 통해 여러 테넌트로 확장할 수 있습니다.	여러 테넌트로 확장하려면 Organizations와 여러 계정을 벤딩해야 합니다.	크기 조정에는 각 새 테넌트에 대한 테넌트별 역할이 필요하며 사용자 프로필에 테넌트 이름으로 태그를 지정해야 합니다.
테넌트 내 사용자 간의 협업은 공유 공간을 통해 가능합니다.	테넌트 내 사용자 간의 협업은 공유 공간을 통해 가능합니다.	모든 테넌트는 공동 작업을 위해 동일한 공유 공간에 액세스할 수 있습니다.

SageMaker AI Studio 도메인 백업 및 복구

실수로 EFS 삭제되거나 네트워킹 또는 인증 변경으로 인해 도메인을 다시 생성해야 하는 경우 다음 지침을 따르십시오.

옵션 1:를 EFS 사용하여 기존에서 백업 EC2

SageMaker Studio 도메인 백업

1. SageMaker Studio(CLI,)에 사용자 프로필과 공백을 나열합니다SDK.
2. 사용자 프로필/스페이스를의 UIDs에 매핑합니다EFS.
   a. users/spaces, describe the user profile/space (CLI, SDK) 목록의 각 사용자에 대해.
   b. 사용자 프로필/스페이스를 `HomeEfsFileSystemUid`에 매핑합니다.
   c. 사용자에게 고유한 실행 역할이 있다면 `UserSettings['ExecutionRole']`에 사용자 프로필을 매핑합니다.
   d. 기본 스페이스 실행 역할을 식별합니다.
3. 새 도메인을 만들고 기본 스페이스 실행 역할을 지정합니다.
4. 사용자 프로필 및 스페이스를 생성합니다.

- 사용자 목록의 각 사용자에 대해 실행 역할 매핑을 사용하여 사용자 프로필(CLI, SDK)을 생성합니다.

5. 새 EFS 및에 대한 매핑을 생성합니다UIDs.
   a. 사용자 목록의 각 사용자에 대해 사용자 프로필(CLI,)을 설명합니다SDK.
   b. `HomeEfsFileSystemUid`에 사용자 프로필을 매핑합니다.
6. 필요한 경우 모든 앱, 사용자 프로필, 스페이스를 삭제한 다음 도메인을 삭제합니다.

EFS 백업

를 백업하려면 다음 지침을 EFS사용합니다.

1. EC2 인스턴스를 시작하고 이전 SageMaker Studio 도메인의 인바운드/아웃바운드 보안 그룹을 새 EC2 인스턴스에 연결합니다(포트 2049TCP에서를 통한 NFS 트래픽 허용. 에서 외부 리소스VPC에 SageMaker Studio 노트북 연결을 참조하세요.
2. SageMaker Studio EFS 볼륨을 새 EC2 인스턴스에 마운트합니다. EFS 파일 시스템 탑재를 참조하세요.
3. 파일을 EBS 로컬 스토리지로 복사합니다. `>sudo cp -rp /efs /studio-backup:`
   a. 새 도메인 보안 그룹을 EC2 인스턴스에 연결합니다.
   b. 새 EFS 볼륨을 EC2 인스턴스에 마운트합니다.
   c. 파일을 새 EFS 볼륨에 복사합니다.
   d. 사용자의 컬렉션에 있는 각 사용자에 대해:
      i. `mkdir new_uid` 디렉터리를 만듭니다.
      ii. 이전 UID 디렉터리의 파일을 새 UID 디렉터리로 복사합니다.
      iii. 모든 파일의 소유권을 변경합니다: 모든 파일에 `chown <new_UID>`를 실행하세요.

옵션 2: S3 및 수명 주기 구성을 EFS 사용하여 기존에서 백업

1. Amazon Linux 2를 사용하여 Amazon SageMaker 노트북 인스턴스로 작업 마이그레이션을 참조하세요.
2. 백업용 S3 버킷을 생성합니다(예:`>studio-backup`).
3. 실행 역할이 있는 모든 사용자 프로필을 나열합니다.
4. 현재 SageMaker Studio 도메인에서 도메인 수준에서 기본 LCC 스크립트를 설정합니다.

- 에서의 모든 `/home/sagemaker-user` 내용을 S3의 사용자 프로필 접두사(예: `s3://studio-backup/studio-user1`)에 LCC복사합니다.

5. 모든 기본 Jupyter Server 앱(실행할 용)LCC을 다시 시작합니다.
6. 모든 앱, 사용자 프로필, 도메인을 삭제합니다.
7. 새 SageMaker Studio 도메인을 생성합니다.
8. 사용자 프로필 및 실행 역할 목록에서 새 사용자 프로필을 생성합니다.
9. 도메인 LCC 수준에서를 설정합니다.
   - 에서 S3의 사용자 프로필 접두사에 있는 모든 항목을에 LCC복사합니다. `/home/sagemaker-user`
10. LCC 구성(CLI,)이 있는 모든 사용자를 위한 기본 Jupyter Server 앱을 생성합니다SDK.

SageMaker SAML어설션을 사용한 Studio 액세스

솔루션 설정:

1. 외부 IdP에서 SAML 애플리케이션을 생성합니다.
2. 에서 외부 IdP를 자격 증명 공급자로 설정합니다IAM.
3. IdP에서 액세스할 수 있는 `SAMLValidator` Lambda 함수를 생성합니다(함수 URL 또는 API 게이트웨이를 통해).
4. `GeneratePresignedUrl` Lambda 함수와 API 게이트웨이를 생성하여 함수에 액세스합니다.
5. 사용자가 API Gateway를 호출하기 위해 맡을 수 있는 IAM 역할을 생성합니다. 이 역할은 어SAML 설션에서 다음 형식의 속성으로 전달되어야 합니다.
   - 속성 이름: https://aws.amazon.com/SAML/속성/역할
   - 속성 값: `<IdentityProviderARN>`, `<RoleARN>`
6. SAML Assertion Consumer Service(ACS) 엔드포인트를 `SAMLValidator` 호출로 업데이트합니다URL.

SAML 검사기 예제 코드:

```
import requests
import os
import boto3
from urllib.parse import urlparse, parse_qs
import base64
```

```
import requests
from aws_requests_auth.aws_auth import AWSRequestsAuth
import json


# Config for calling AssumeRoleWithSAML
idp_arn = "arn:aws:iam::0123456789:saml-provider/MyIdentityProvider"
api_gw_role_arn = 'arn:aws:iam:: 0123456789:role/APIGWAccessRole'
studio_api_url = "abcdef.execute-api.us-east-1.amazonaws.com"
studio_api_gw_path = "https://" + studio_api_url + "/Prod "

# Every customer will need to get SAML Response from the POST call
def get_saml_response(event):
    saml_response_uri = base64.b64decode(event['body']).decode('ascii')
    request_body = parse_qs(saml_response_uri)
    print(f"b64 saml response: {request_body['SAMLResponse'][0]}")
    return request_body['SAMLResponse'][0]


def lambda_handler(event, context):
    sts = boto3.client('sts')

    # get temporary credentials
    response = sts.assume_role_with_saml(
                    RoleArn=api_gw_role_arn,
                    PrincipalArn=durga_idp_arn,
                    SAMLAssertion=get_saml_response(event)
                )
    auth = AWSRequestsAuth(aws_access_key=response['Credentials']['AccessKeyId'],
                      aws_secret_access_key=response['Credentials']['SecretAccessKey'],
                      aws_host=studio_api_url,
                      aws_region='us-west-2',
                      aws_service='execute-api',
                      aws_token=response['Credentials']['SessionToken'])

    presigned_response = requests.post(
        studio_api_gw_path,
        data=saml_response_data,
        auth=auth)

    return presigned_response
```

참조 자료

- 에서 안전하고 잘 관리되는 기계 학습 환경 설정 AWS(AWS 블로그)
- 전체 리소스 격리를 통해 팀 및 그룹에 Amazon SageMaker AI Studio 구성(AWS 블로그)
- 및 Okta Universal Directory를 사용하여 AWS SSO Amazon SageMaker AI Studio 온보딩(AWS 블로그)
- AWS 계정 연동을 위해 SAML 2.0을 구성하는 방법(Okta 설명서)
- AWS에서 안전한 엔터프라이즈 기계 학습 플랫폼 구축(AWS 기술 가이드)
- 수명 주기 구성을 사용하여 Amazon SageMaker AI Studio 사용자 지정(AWS 블로그)
- Amazon SageMaker AI Studio 노트북에 사용자 지정 컨테이너 이미지 가져오기(AWS 블로그)
- 사용자 지정 SageMaker AI 프로젝트 템플릿 빌드 - 모범 사례(AWS 블로그)
- Amazon SageMaker AI Pipelines을 사용한 다중 계정 모델 배포(AWS 블로그)
- 1부: NatWest Group이 확장 가능하고 안전하며 지속 가능한 MLOps 플랫폼을 구축한 방법(AWS 블로그)
- 보안 Amazon SageMaker AI Studio 사전 서명된 1URLs부: 기본 인프라(AWS 블로그)

기여자

다음은 이 문서의 기여자입니다.

- Ram Vittal, ML 솔루션스 아키텍트, Amazon Web Services
- Sean Morgan, ML 솔루션스 아키텍트, Amazon Web Services
- Durga Sury, ML 솔루션스 아키텍트, Amazon Web Services

아이디어, 수정 및 관점을 제공해 주신 다음 분들께 특별히 감사드립니다.

- Alessandro Cerè, AI/ML 솔루션스 아키텍트, Amazon Web Services
- Sumit Thakur, Amazon Web Services SageMaker AI 제품 리더
- Han Zhang, 선임 소프트웨어 개발 엔지니어, Amazon Web Services
- Bhadrinath Pani, 소프트웨어 개발 엔지니어, Amazon Web Services

문서 수정

이 백서에 대한 업데이트 알림을 받으려면 RSS 피드를 구독하면 됩니다.

변경 사항	설명	날짜
백서 업데이트	끊어진 링크가 수정되었고 전체적으로 편집상 변경이 많이 이루어졌습니다.	2023년 4월 25일
최초 게시	백서가 게시되었습니다.	2022년 10월 19일

고지 사항

고객은 본 문서의 정보를 독립적으로 평가할 책임이 있습니다. 본 문서는 (a) 정보 제공의 목적으로만 제공되고, (b) 사전 통지 없이 변경될 수 있는 현재 AWS 제품 및 관행을 나타내고, (c) AWS 및 그 계열사, 공급업체 또는 라이선스 제공자로부터 어떠한 약속이나 보증도 하지 않습니다. AWS 제품 또는 서비스는 명시적이든 묵시적이든 어떠한 종류의 보증, 진술 또는 조건 없이 '있는 그대로' 제공됩니다. 고객에 대한 AWS의 책임 및 채무는 AWS 계약에 준거합니다. 본 문서는 AWS와 고객 간의 어떠한 계약도 구성하지 않으며 이를 변경하지도 않습니다.

© 2022 Amazon Web Services, Inc. 또는 계열사. All rights reserved.

AWS 용어집

최신 AWS 용어는 AWS 용어집 참조의 AWS 용어집을 참조하세요.

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