AWS ホワイトペーパー

SageMaker Studio 管理のベストプラクティス

SageMaker Studio 管理のベストプラクティス: AWS ホワイトペーパー

Table of Contents

SageMaker Studio 管理のベストプラクティス

発行日: 2023 年 4 月 25 日 (ドキュメントの改訂)

要約

Amazon SageMaker Studio では、すべての機械学習 (ML) 開発ステップを単一のウェブベースのビジュアルインターフェイスで実行できるため、データサイエンスチームの生産性が向上します。SageMaker Studio は、モデルの構築、トレーニング、評価に必要な各ステップに対する完全なアクセス、制御、可視化を提供します。

このホワイトペーパーでは、運用モデル、ドメイン管理、ID 管理、アクセス許可管理、ネットワーク管理、ログ記録、モニタリング、カスタマイズなどのテーマに関するベストプラクティスについて説明します。ここで説明するベストプラクティスは、マルチテナントデプロイを含む企業向けの SageMaker Studio デプロイを意図しています。このドキュメントは、ML プラットフォーム管理者、ML エンジニア、ML アーキテクトを対象としています。

Well-Architected の実現状況の確認

AWS Well-Architected フレームワークは、クラウド内でのシステム構築に伴う意思決定の長所と短所を理解するのに役立ちます。このフレームワークの 6 つの柱により、信頼性、安全性、効率、費用対効果、持続可能性の高いシステムを設計および運用するための、アーキテクチャのベストプラクティスを確認できます。AWS Management Console に無料で提供されている AWS Well-Architected Tool を使用すると、各柱に関する一連の質問に答えることで、これらのベストプラクティスに照らしてワークロードをレビューできます。

機械学習レンズでは、AWS クラウド で機械学習ワークロードを設計、デプロイ、構築する方法に焦点を当てます。このレンズは、Well-Architected Framework で説明されているベストプラクティスを発展させます。

序章

SageMaker Studio を ML プラットフォームとして管理する場合、ワークロードの増大に合わせて ML プラットフォームをスケールできるように、十分な情報に基づいた意思決定を行うためのベストプラクティスのガイダンスが必要です。ML プラットフォームのプロビジョニング、運用、スケーリングを行う場合は、以下の点を考慮してください。

- 適切な運用モデルを選択し、ビジネス目標を達成できるように ML 環境を整えます。
- ユーザー ID に対する SageMaker Studio ドメイン認証の設定方法を選択し、ドメインレベルの制限を考慮します。
- きめ細かなアクセス制御と監査のために、ユーザーの ID と承認を ML プラットフォームにフェデレーションする方法を決定します。
- ML ペルソナのさまざまなロールに対してアクセス許可とガードレールを設定することを検討します。
- ML ワークロードの感度、ユーザー数、インスタンスタイプ、アプリ、起動されるジョブを考慮して、仮想プライベートクラウド (VPC) ネットワークトポロジを計画します。
- 保存中と転送中のデータを分類し、暗号化して保護します。
- コンプライアンスを確保するために、さまざまなアプリケーションプログラミングインターフェイス (API) やユーザーアクティビティをログに記録してモニタリングする方法を検討します。
- 独自のイメージとライフサイクル設定スクリプトを使用して SageMaker Studio ノートブックのエクスペリエンスをカスタマイズします。

運用モデル

運用モデルは、人材、プロセス、テクノロジーを結び付けて、組織がスケーラブルで一貫性のある効率的な方法でビジネス価値を実現できるようにするフレームワークです。ML 運用モデルは、組織全体のチームに標準的な製品開発プロセスを提供します。運用モデルの実装には、規模、複雑さ、ビジネス推進要因に応じて、次の 3 つのモデルがあります。

- 一元化されたデータサイエンスチーム — このモデルでは、すべてのデータサイエンス活動が単一のチームまたは組織内に一元化されます。これは、センターオブエクセレンス (COE) モデルに似ており、データサイエンスプロジェクトに関してはすべてのビジネスユニットがこのチームに参加します。
- 分散型データサイエンスチーム — このモデルでは、データサイエンス活動が複数の異なるビジネス機能や部門に分散されるか、異なる製品ライン別に分散されます。
- フェデレーションデータサイエンスチーム — このモデルの場合、コードリポジトリ、継続的インテグレーション/継続的デリバリー (CI/CD) パイプラインなどの共有サービス機能は一元化されたチームによって管理され、各ビジネスユニットや製品レベルの機能は分散型チームによって管理されます。これは、ハブアンドスポークモデルに似ており、ビジネスユニットごとに独自のデータサイエンスチームがありますが、これらのビジネスユニットチームは、一元化されたチームとの間で活動を調整します。

最初の Studio ドメインを本番環境のユースケースで立ち上げるかどうかを決める前に、運用モデルと環境整理の AWS ベストプラクティスを検討します。詳細については、「複数のアカウントを使用して AWS 環境を整理する」を参照してください。

次のセクションでは、運用モデルごとにアカウント構造を整理するためのガイダンスを提供します。

推奨されるアカウント構造

このセクションでは、運用モデルのアカウント構造を簡単に紹介します。この構造から始めて、組織の運用要件に応じて変更できます。どの運用モデルを選択するかにかかわらず、以下の一般的なベストプラクティスを実装することをお勧めします。

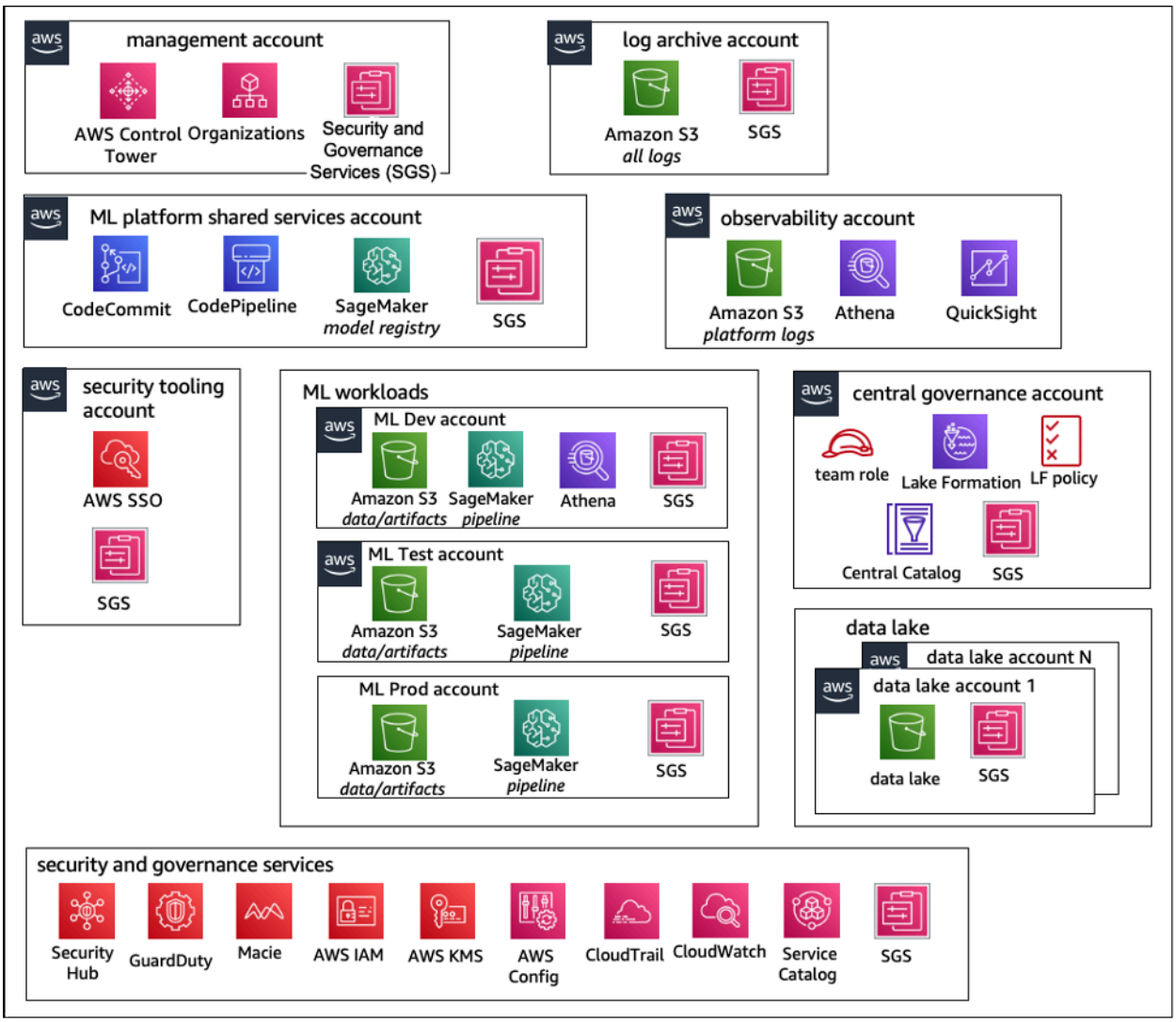
- AWS Control Tower を使用してアカウントの設定、管理、ガバナンスを行います。
- ID プロバイダー (IdP) と AWS IAM アイデンティティセンターを使用して ID を一元管理します。そのために、委任された管理者のセキュリティツールアカウントを使用し、ワークロードへの安全なアクセスを可能にします。

- 開発、テスト、本番環境のワークロードにわたってアカウントレベルの分離を使用して ML ワークロードを実行します。
- ML ワークロードログをログアーカイブアカウントにストリーミングし、オブザーバビリティアカウントでログ分析をフィルタリングして適用します。
- データアクセスをプロビジョニング、制御、監査するための一元化されたガバナンスアカウントを実行します。
- 組織やワークロードの要件に応じて、セキュリティとコンプライアンスを確保するために、セキュリティとガバナンスサービス (SGS) を適切な予防および検出ガードレールと共に各アカウントに組み込みます。

一元化されたモデルアカウント構造

このモデルでは、ML プラットフォームチームが以下を提供する責任があります。

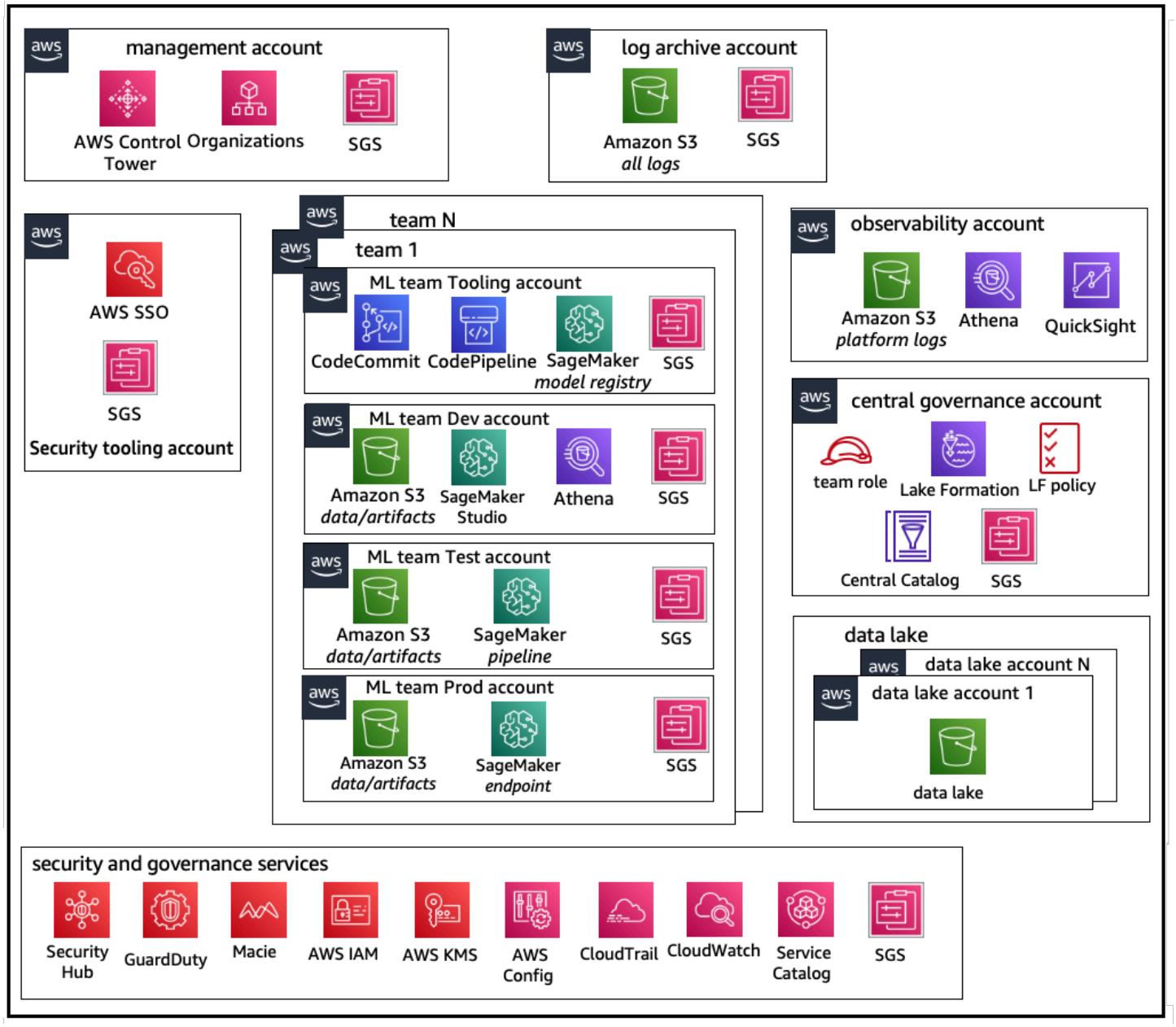
- データサイエンスチーム間で機械学習オペレーション (MLOps) の要件に対処する共有サービスツールアカウント。
- データサイエンスチーム間で共有される ML ワークロードの開発、テスト、および本番環境のアカウント。
- データサイエンスチームのワークロードごとに分離して実行するためのガバナンスポリシー。
- 一般的なベストプラクティス。



一元化された運用モデルのアカウント構造

分散型モデルのアカウント構造

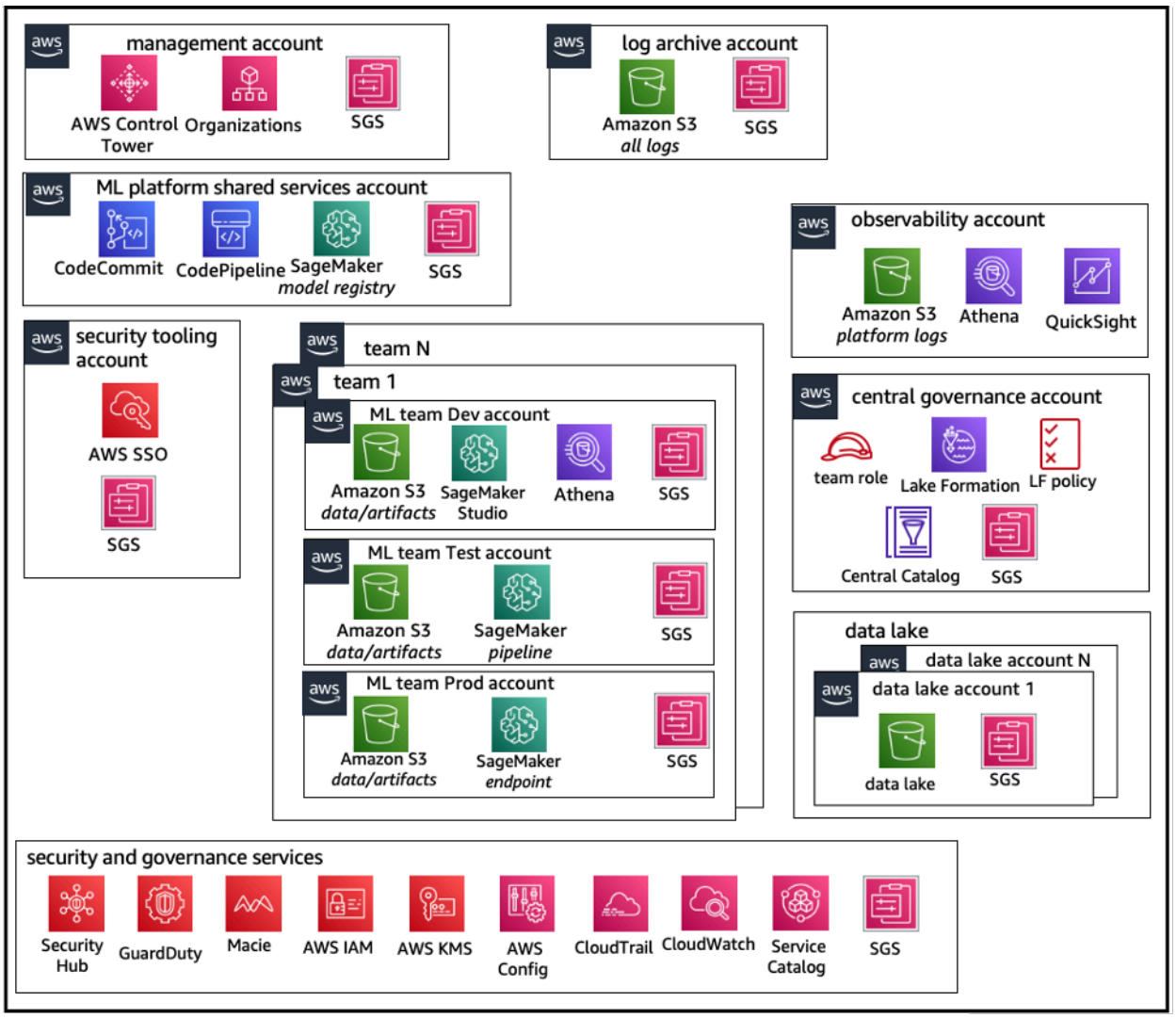
このモデルでは、各 ML チームが独立して ML アカウントとリソースのプロビジョニング、管理、ガバナンスを行います。ただし、機械学習チームには、データガバナンスと監査管理を簡素化するために、一元化されたオブザーバビリティとデータガバナンスモデルのアプローチを使用することをお勧めします。



分散型運用モデルのアカウント構造

フェデレーションモデルのアカウント構造

このモデルは一元化されたモデルに似ていますが、主な違いは、データサイエンス/ML チームごとに独自の開発/テスト/本番環境のワークロードアカウントを持つことです。これにより、ML リソースの堅牢な物理的分離を可能にし、各チームが他のチームに影響を与えることなく独立してスケールできます。



フェデレーション運用モデルのアカウント構造

ML プラットフォームのマルチテナンシー

マルチテナンシーは、1 つのソフトウェアインスタンスが複数の異なるユーザーグループにサービスを提供できるソフトウェアアーキテクチャです。テナントは、ソフトウェアインスタンスへの特定の権限を持つ、共通のアクセスを共有するユーザーのグループです。例えば、複数の ML 製品を構築している場合、同様のアクセス要件を持つ各製品チームはテナントまたはチームと見なすことができます。

SageMaker Studio インスタンス (SageMaker ドメインなど) 内に複数のチームを実装することは可能ですが、複数のチームを 1 つの SageMaker Studio ドメインにまとめることに伴う利点とトレードオフ (影響範囲、コスト属性、アカウントレベルの制限など) を比較検討します。これらのトレードオフとベストプラクティスの詳細については、以下のセクションを参照してください。

リソースを完全に分離する必要がある場合は、異なるアカウントのテナントごとに SageMaker Studio ドメインを実装することを検討します。分離要件によっては、複数の事業部門 (LOB) を単一のアカウントおよびリージョン内の複数のドメインとして実装できます。共有スペースを使用すると、同じチーム/LOB のメンバー間でほぼリアルタイムのコラボレーションが可能になります。ドメインが複数ある場合でも、ID アクセス管理 (IAM) ポリシーとアクセス許可を使用してリソースを確実に分離します。

ドメインから作成した SageMaker リソースには、ドメインの Amazon リソースネーム (ARN) とユーザープロファイルまたはスペースの ARN が自動的にタグ付けされるため、リソースを簡単に分離できます。ポリシーの例については、ドメインリソース分離に関するドキュメントを参照してください。このドキュメントでは、マルチアカウント戦略またはマルチドメイン戦略をいつ使用するかについての詳細なリファレンスと、機能の比較を確認できます。また、GitHub リポジトリで既存のドメインのタグをバックフィルするサンプルスクリプトも確認できます

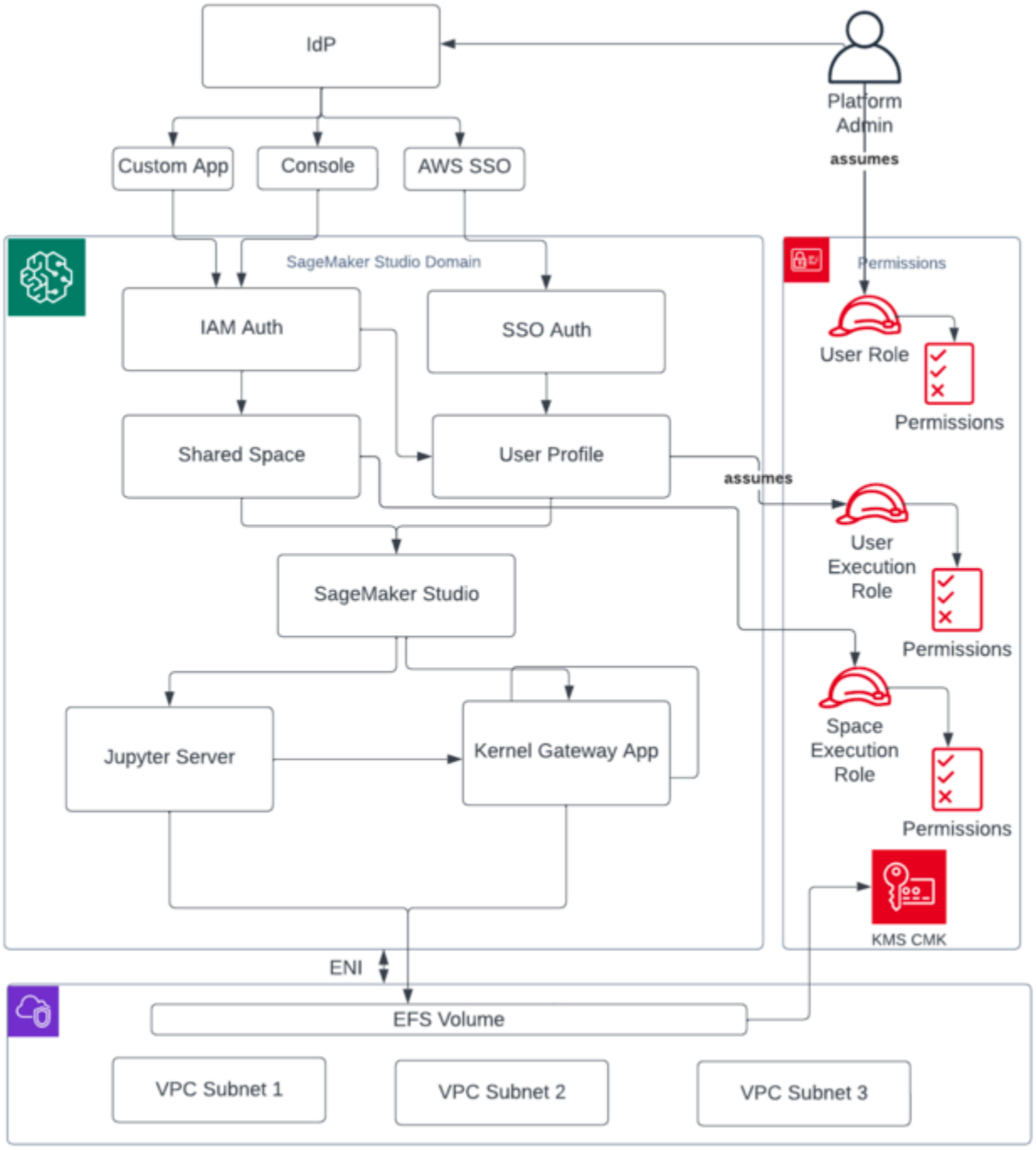
最後に、AWS Service Catalog を使用して複数のアカウントに SageMaker Studio リソースのセルフサービスデプロイを実装できます。詳細については、「複数の AWS アカウントと AWS リージョンで AWS Service Catalog 製品を管理する」を参照してください。

ドメイン管理

Amazon SageMaker ドメインは、以下で構成されます。

- 関連する Amazon Elastic File System (Amazon EFS) ボリューム
- 承認されたユーザーのリスト
- セキュリティ、アプリケーション、ポリシー、Amazon Virtual Private Cloud (Amazon VPC) に関するさまざまな設定

次の図は、SageMakerStudio ドメインを構成するさまざまなコンポーネントの概要を示しています。

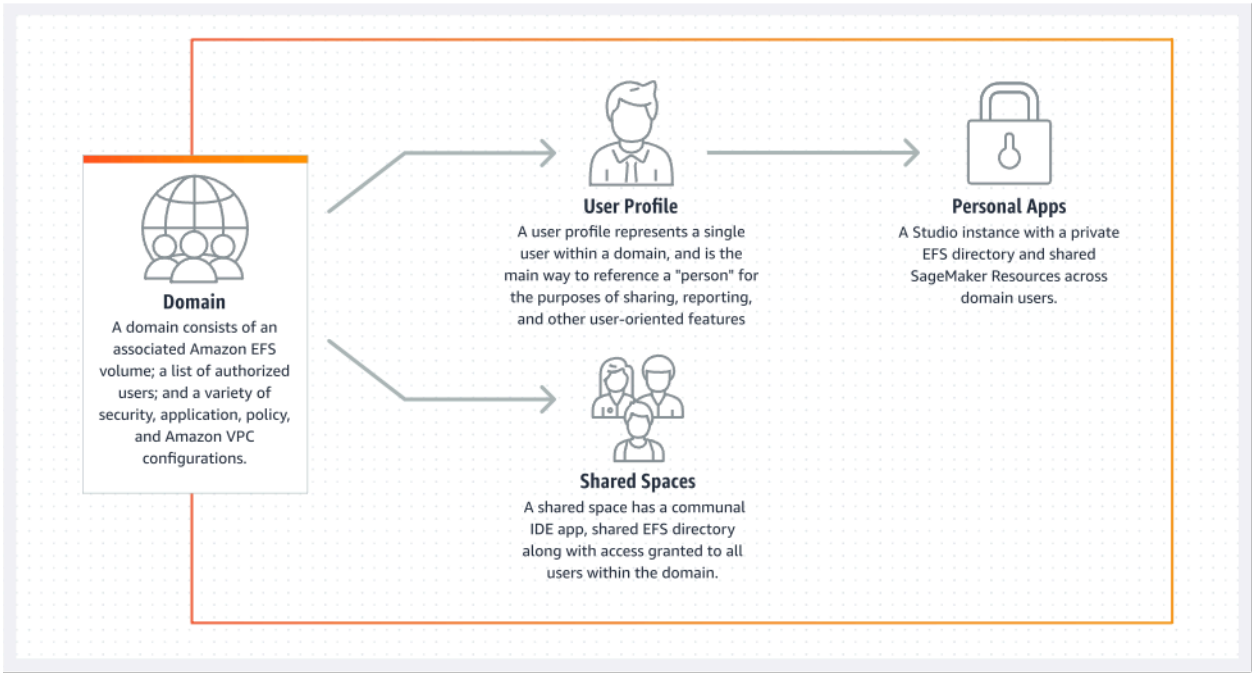


SageMaker Studio ドメインを構成するさまざまなコンポーネントの概要図

複数のドメインと共有スペース

Amazon SageMaker では、アカウントごとに複数の SageMaker ドメインを 1 つの AWS リージョンで作成できるようになりました。ドメインごとに、認証モードなどの独自のドメイン設定と、VPC やサブネットなどのネットワーク設定を構成できます。ユーザープロファイルをドメイン間で共有することはできません。1 人のユーザーがドメインで区切られた複数のチームに属している場合は、ドメインごとにユーザーのユーザープロファイルを作成します。既存のドメインでのタグのバックフィルについては、「複数ドメインの概要」を参照してください。

IAM 認証モードで設定した各ドメインでは、共有スペースを利用してユーザー間のほぼリアルタイムのコラボレーションを行うことができます。共有スペースでは、ユーザーは共有 Amazon EFS ディレクトリとユーザーインターフェイスの共有 JupyterServer アプリにアクセスし、ほぼリアルタイムで共同編集できます。共有スペースで作成されたリソースの自動タグ付けにより、管理者はプロジェクトレベルでコストを追跡できます。また、共有 JupyterServer UI では、実験やモデルレジストリエントリなどのリソースをフィルタリングして、共有 ML の取り組みに関連する項目のみが表示されるようにします。次の図は、各ドメイン内のプライベートアプリと共有スペースの概要を示しています。



単一ドメイン内のプライベートアプリと共有スペースの概要

ドメインに共有スペースを設定する

共有スペースは通常、特定の ML の取り組みやプロジェクトで、単一ドメインのメンバーが同じ基盤のファイルストレージや IDE にほぼリアルタイムでアクセスする必要がある場合に作成します。ユーザーはノートブックのアクセス、閲覧、編集、共有をほぼリアルタイムで行うことができるため、同僚との反復作業を最も早く開始できます。

共有スペースを作成するには、まずスペースを使用するすべてのユーザーのアクセス許可を管理する、スペースのデフォルトの実行ロールを指定する必要があります。この記事の執筆時点では、ドメイン内のすべてのユーザーがドメイン内のすべての共有スペースにアクセスできます。既存のドメインに対する共有スペースの追加に関する最新のドキュメントについては、「共有スペースの作成」を参照してください。

IAM フェデレーション用にドメインを設定する

SageMaker Studio ドメインの AWS Identity and Access Management (IAM) フェデレーションを設定する前に、「ID 管理」セクションで説明しているように、IdP に IAM フェデレーションユーザーロール (プラットフォーム管理者など) を設定する必要があります。

IAM オプションを使用して SageMaker Studio を設定する詳細な手順については、「IAM アイデンティティセンターを使用して Amazon SageMaker ドメインにオンボーディングする」を参照してください。

シングルサインオン (SSO) フェデレーション用にドメインを設定する

シングルサインオン (SSO) フェデレーションを使用するには、SageMaker Studio を実行する必要があるリージョンと同じリージョンにある AWS Organizations 管理アカウントで AWS IAM Identity Center を有効にする必要があります。ドメインの設定手順は、[認証] セクションで [AWS IAM Identity Center (IdC)] を選択することを除いて、IAM フェデレーションの手順と同じです。

詳細な手順については、「IAM アイデンティティセンターを使用して Amazon SageMaker ドメインにオンボーディングする」を参照してください。

SageMaker Studio ユーザープロファイル

ユーザープロファイルは、ドメイン内の単一のユーザーを表し、共有、レポート、その他のユーザー指向機能を実行する際の「個人」を参照する主な方法です。このエンティティは、ユーザーが SageMaker Studio にオンボードするときに作成されます。管理者が E メールでユーザーを招待したり、IdC からインポートしたりすると、ユーザープロファイルが自動的に作成されます。ユーザープロファイルは、各ユーザーの設定を主に保持し、ユーザーのプライベートな Amazon Elastic File System (Amazon EFS) ホームディレクトリへの参照を含みます。SageMaker Studio アプリケーションの物理ユーザーごとにユーザープロファイルを作成することをお勧めします。ユーザーごとに専用のディレクトリが Amazon EFS にあり、ユーザープロファイルを同じアカウント内のドメイン間で共有することはできません。

SageMaker Studio ドメインを共有するユーザープロファイルごとに、ノートブックを実行するための専用のコンピューティングリソース (SageMaker の Amazon Elastic Compute Cloud (Amazon EC2) インスタンスなど) が割り当てられます。ユーザー 1 に割り当てられたコンピューティングインスタンスは、ユーザー 2 に割り当てられたコンピューティングインスタンスから完全に分離されます。同様に、1 つの AWS アカウント内のユーザーに割り当てられたコンピューティングリソースは、別のアカウント内のユーザーに割り当てられたコンピューティングリソースから完全に分離されます。各ユーザーは、分離された Docker コンテナ内の最大 4 つのアプリケーション (アプリ) を実行できます。または、同じインスタンスタイプのイメージを実行できます。

Jupyter Server アプリ

署名付き URL にアクセスするか、AWS IAM IdC を通じてログインしてユーザーの Amazon SageMaker Studio ノートブックを起動すると、Jupyter Server アプリが SageMaker サービスマネージドの VPC インスタンスで起動されます。各ユーザーは、プライベートアプリで専用の Jupyter Server アプリを個別に取得します。デフォルトでは、SageMaker Studio ノートブック用の Jupyter Server アプリは、専用の `ml.t3.medium` インスタンス (システムインスタンスタイプとして予約済み) 上で実行されます。このインスタンスのコンピューティング料金は、お客様に請求されません。

Jupyter カーネルゲートウェイアプリ

カーネルゲートウェイアプリは、API または SageMaker Studio インターフェイスを介して作成でき、選択したインスタンスタイプで実行されます。このアプリは、一般的なデータサイエンスや深層学習パッケージ (TensorFlow、Apache MXNet、PyTorch など) で事前に設定された組み込みの SageMaker Studio イメージのいずれかを使用して実行できます。

ユーザーは、同じ SageMaker Studio イメージ/カーネルゲートウェイアプリ内で、複数の Jupyter Notebook カーネル、ターミナルセッション、インタラクティブコンソールを起動して実行できます。また、同じ物理インスタンスで最大 4 つのカーネルゲートウェイアプリまたはイメージを実行できます (それぞれがコンテナ/イメージによって分離されます)。

追加のアプリを作成するには、別のインスタンスタイプを使用する必要があります。ユーザープロファイルでは、インスタンスタイプを問わず、1 つのインスタンスのみ実行できます。例えば、同じインスタンスで、SageMaker Studio の組み込みデータサイエンスイメージを使用する単純なノートブックと、組み込みの TensorFlow イメージを使用する別のノートブックの両方を実行できます。ユーザーは、インスタンスを実行した時間分の料金を請求されます。ユーザーが SageMaker Studio をアクティブに実行していないときのコストを回避するには、インスタンスをシャットダウンする必要があります。詳細については、「Studio アプリのシャットダウンと更新」を参照してください。

SageMaker Studio インターフェイスからカーネルゲートウェイアプリをシャットダウンして再度開くたびに、アプリは新しいインスタンスで起動されます。つまり、同じアプリを再起動しても、パッケージのインストールは保持されません。同様に、ユーザーがノートブックのインスタンスタイプを変更すると、インストール済みのパッケージとセッション変数は失われます。ただし、独自のイメージの導入やライフサイクルスクリプトなどの機能を使用してユーザー独自のパッケージを SageMaker Studio に持ち込むと、インスタンスを切り替えたり、新しいインスタンスを起動したりしても、パッケージは保持されます。

Amazon Elastic File System ボリューム

ドメインを作成すると、単一の Amazon Elastic File System (Amazon EFS)ボリュームが作成され、ドメイン内のすべてのユーザーが使用できるようになります。各ユーザープロファイルには、ユーザーのノートブック、GitHub リポジトリ、データファイルを保存するためのプライベートホームディレクトリが、Amazon EFS ボリューム内に割り当てられます。ドメイン内の各スペースには、複数のユーザープロファイルからアクセスできるプライベートディレクトリが、Amazon EFS ボリューム内に割り当てられます。フォルダへのアクセスは、ファイルシステムのアクセス許可を通じて、ユーザーごとに分離されます。SageMaker Studio は、ユーザープロファイルやスペースごとにグローバルな一意のユーザー ID を作成し、これを EFS でユーザーのホームディレクトリに対するポータブルオペレーティングシステムインターフェイス (POSIX) のユーザー/グループ ID として適用します。これにより、他のユーザーやスペースはそのデータにアクセスできなくなります。

バックアップとリカバリ

既存の EFS ボリュームを新しい SageMaker ドメインにアタッチすることはできません。本番環境の設定で、Amazon EFS ボリュームが別の EFS ボリュームや Amazon Simple Storage Service (Amazon S3) にバックアップされていることを確認します。EFS ボリュームを誤って削除した場合、管理者は SageMaker Studio ドメインを削除して再作成する必要があります。手順は次のとおりです。

ユーザープロファイル、スペース、および関連する EFS ユーザー ID (UID) のリストをバックアップするには、`ListUserProfiles`、`DescribeUserProfile`、`ListSpaces`、`DescribeSpace` API コールを使用します。

1. 新しい SageMaker Studio ドメインを作成します。
2. ユーザープロファイルとスペースを作成します。
3. ユーザープロファイルごとに、バックアップから EFS/Amazon S3 にファイルをコピーします。
4. 必要に応じて、古い SageMaker Studio ドメインのすべてのアプリとユーザープロファイルを削除します。

詳細な手順については、付録の「SageMaker Studio ドメインのバックアップとリカバリ」セクションを参照してください。。

> **Note**
>
> これを行う別の方法としては、ユーザーがアプリを起動するたびに S3 とやり取りするデータを `LifecycleConfigurations` でバックアップします。

Amazon EBS ボリューム

Amazon Elastic Block Store (Amazon EBS) ストレージボリュームも、各 SageMaker Studio ノートブックインスタンスにアタッチします。これは、インスタンスで実行しているコンテナまたはイメージのルートボリュームとして使用されます。Amazon EFS ストレージは永続的ですが、コンテナにアタッチした Amazon EBS ボリュームは一時的なものです。Amazon EBS ボリュームにローカルに保存したデータは、お客様がアプリを削除した場合に保持されません。

署名付き URL へのアクセスの保護

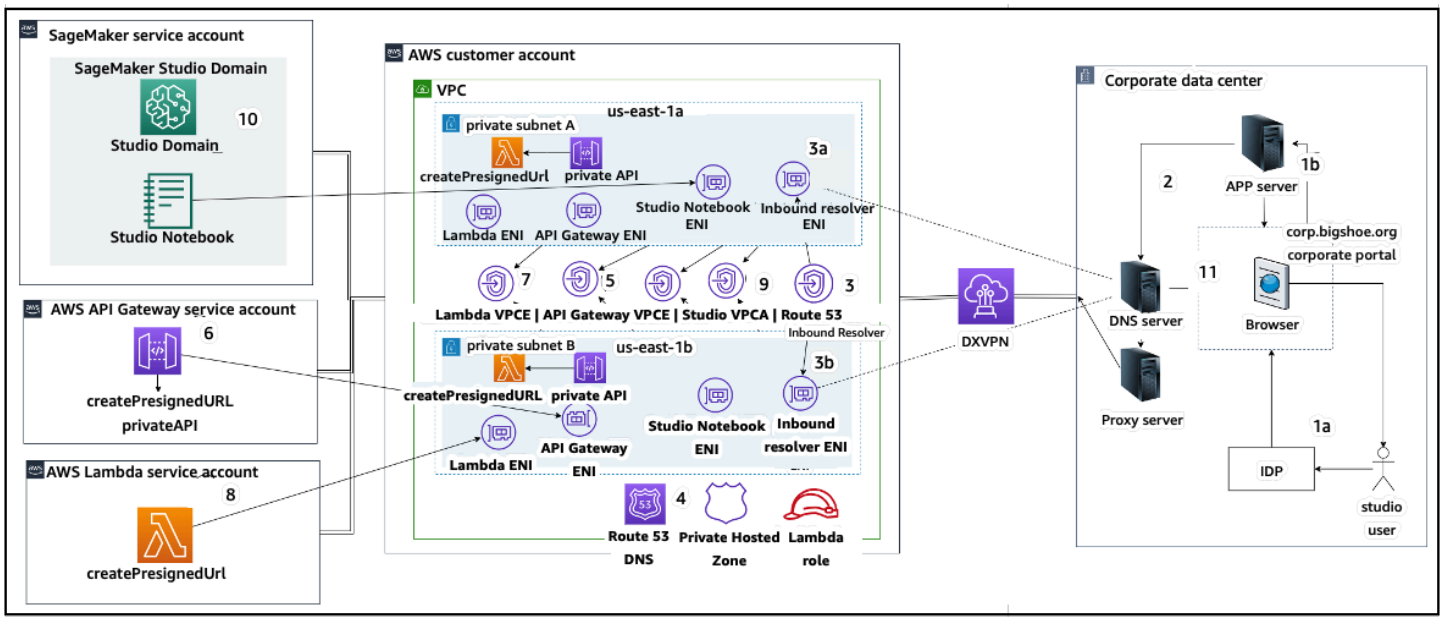
SageMaker Studio ユーザーがノートブックのリンクを開くと、SageMaker Studio はフェデレーションユーザーの IAM ポリシーを検証してアクセスを許可し、ユーザーの署名付き URL を生成して解決します。SageMaker コンソールはインターネットドメイン上で動作するため、この生成された署名付き URL はブラウザセッションに表示されます。これは、適切なアクセス制御が適用されていない場合に、データの盗難やお客様のデータへのアクセスという望ましくない脅威ベクトルとなります。

Studio は、署名付き URL のデータ盗難に対してアクセス制御を適用する方法をいくつかサポートしています。

- IAM ポリシー条件 `aws:sourceIp` を使用したクライアント IP の検証
- IAM 条件 `aws:sourceVpc` を使用したクライアント VPC の検証
- IAM ポリシー条件 `aws:sourceVpce` を使用したクライアント VPC エンドポイントの検証

SageMaker コンソールから SageMaker Studio ノートブックにアクセスする場合、利用できる唯一のオプションは IAM ポリシー条件 `aws:sourceIp` を使用したクライアント IP の検証となります。ただし、Zscaler などのブラウザトラフィックルーティング製品を使用すると、従業員のインターネットアクセスのスケールとコンプライアンスを確保できます。これらのトラフィックルーティング製品は、独自のソース IP を生成しますが、その IP 範囲は企業のお客様によって制御されません。そのため、このような企業のお客様は `aws:sourceIp` 条件を使用できなくなります。

IAM ポリシー条件 `aws:sourceVpce` を使用したクライアント VPC エンドポイントの検証を使用するには、署名付き URL の作成元が SageMaker Studio のデプロイ先と同じお客様の VPC である必要があり、署名付き URL の解決はお客様の VPC の SageMaker Studio VPC エンドポイント経由で行う必要があります。このような企業ネットワークユーザーのアクセス時間中における署名付き URL の解決は、次のアーキテクチャに示すように、DNS 転送ルールを (Zscaler と企業 DNS の両方で) 使用して実現できます。解決されると、Amazon Route 53 のインバウンドリゾルバーを使用してお客様の VPC エンドポイントに送信されます。



企業ネットワーク経由で VPC エンドポイントを使用した Studio の署名付き URL へのアクセス

上記のアーキテクチャを設定するステップバイステップのガイダンスについては、「Amazon SageMaker Studio の署名付き URL のセキュリティ保護パート 1: 基盤インフラストラクチャ」を参照してください。

SageMaker ドメインのクォータと制限

- SageMaker Studio のドメイン SSO フェデレーションは、リージョンに限り、AWS アイデンティティセンターがプロビジョニングされている AWS 組織のメンバーアカウント間でのみサポートされます。
- 共有スペースは、AWS アイデンティティセンターで設定したドメインでは現在サポートされていません。
- VPC やサブネットの設定は、ドメインを作成した後では変更できません。ただし、VPC やサブネットの設定が異なる新しいドメインを作成することはできます。
- ドメインを作成した後は、ドメインアクセスを IAM モードと SSO モードの間で切り替えることはできません。別の認証モードを使用して新しいドメインを作成できます。
- カーネルゲートウェイアプリの数は、各ユーザーに起動したインスタンスタイプごとに 4 つに制限されます。
- 各ユーザーは、各インスタンスタイプのインスタンスを 1 つだけ起動できます。

- ドメイン内で消費されるリソースには、インスタンスタイプごとに起動するインスタンスの数、作成できるユーザープロファイルの数などに関する制限があります。サービス制限の詳細なリストについては、サービスクォータのページを参照してください。
- お客様は、ビジネス上の正当な理由を付してエンタープライズサポートケースを提出し、アカウントレベルのガードレールに応じたドメイン数やユーザープロファイル数などのデフォルトのリソース制限を引き上げることができます。
- アカウントごとの同時実行アプリの数に対するハード制限は 2,500 です。ドメインとユーザープロファイルの制限は、このハード制限に応じて異なります。例えば、アカウントは 1,000 のユーザープロファイルを含む 1 つのドメインを持つことも、それぞれ 50 のユーザープロファイルを含む 20 のドメインを持つこともできます。

ID 管理

このセクションでは、企業ディレクトリ内の従業員ユーザーが AWS アカウントへのフェデレーションを通じて SageMaker Studio にアクセスする方法について説明します。まず、ユーザー、グループ、ロールのマッピング方法とユーザーフェデレーションの仕組みについて簡単に説明します。

ユーザー、グループ、ロール

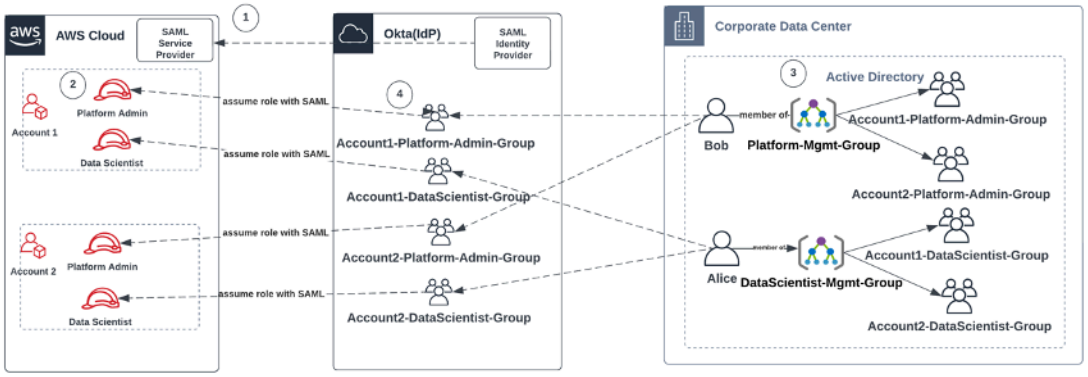
AWS では、ユーザー、グループ、ロールを使用してリソースのアクセス許可を管理します。お客様は、ユーザーとグループを管理するために、IAM を使用するか、Active Directory (AD) などの企業ディレクトリを使用できます。企業ディレクトリは、Okta などの外部 IdP を介して有効にすることで、クラウドやオンプレミスで実行しているさまざまなアプリケーションに対してユーザーを認証できます。

「AWS セキュリティの柱」の「ID 管理」セクションで説明しているように、一元的な IdP でユーザー ID を管理するのがベストプラクティスです。これにより、バックエンドの人事プロセスとの統合や、従業員ユーザーへのアクセス管理が容易になります。

Okta などの IdP を使用すると、エンドユーザーは 1 つ以上の AWS アカウントに対して認証し、Security Assertion Markup Language (SAML) による SSO を使用して特定のロールにアクセスできます。IdP 管理者は、AWS アカウントから IdP にロールをダウンロードしてユーザーに割り当てることができます。エンドユーザーが AWS にログインすると、1 つ以上の AWS アカウントでエンドユーザーに割り当てられている AWS ロールのリストが AWS 画面に表示されます。エンドユーザーは、ログイン時に引き受けるロールを選択できます。このロールは、認証されたセッション中におけるアクセス許可を定義します。

アクセスを許可する特定のアカウントとロールの組み合わせごとに、グループが IdP に存在する必要があります。これは、AWS ロール別のグループと考えることができます。ロール別のグループのメンバーであるユーザーには、1 つの特定の AWS アカウントで 1 つの特定のロールにアクセスするための 1 つの権限が付与されます。ただし、この 1 つの権限を付与するプロセスは、各ユーザーを特定の AWS ロールグループに割り当ててユーザーアクセスを管理するようにスケールすることはできません。管理を簡略化するには、さまざまな AWS 権限のセットを必要とする組織内のすべてのユーザーセットごとに、複数のグループを個別に作成することをお勧めします。

一元的な IdP 設定の例として、AD を設定している企業を考えてみます。この設定では、ユーザーとグループを IdP ディレクトリに同期します。AWS では、これらの AD グループは IAM ロールにマップされます。ワークフローの主なステップは、以下のとおりです。



AD ユーザー、AD グループ、IAM ロールをオンボーディングするためのワークフロー

1. AWS で、各 AWS アカウントと IdP との SAML 統合を設定します。
2. AWS で、AWS アカウントごとにロールを設定し、IdP に同期します。
3. 企業の AD システムで以下の操作を行います。
   a. アカウントロールごとに AD グループを作成し、IdP に同期します (例: `Account1-Platform-Admin-Group` を AWS ロールグループとして作成します)。
   b. 各ペルソナレベルで管理グループ (`Platform-Mgmt-Group` など) を作成し、AWS ロールグループをメンバーとして割り当てます。
   c. 管理グループにユーザーを割り当て、AWS アカウントロールへのアクセスを許可します。
4. IdP で、AWS ロールグループ (`Account1-Platform-Admin-Group` など) を AWS アカウントロール (Account1 のプラットフォーム管理者など) にマップします。
5. データサイエンティストの Alice が IdP にログインすると、AWS フェデレーションアプリの UI に「アカウント 1 データサイエンティスト」と「アカウント 2 データサイエンティスト」の 2 つのオプションが表示されます。
6. Alice が「アカウント 1 データサイエンティスト」オプションを選択すると、AWS アカウント 1 (SageMaker コンソール) の承認されたアプリケーションに接続されます。

SAML アカウントフェデレーションの詳細な設定手順については、Okta の「AWS アカウントフェデレーション用に SAML 2.0 を設定する方法」を参照してください。

ユーザーフェデレーション

SageMaker Studio の認証は IAM または IAM IdC のいずれかを使用して行うことができます。ユーザーを IAM を介して管理する場合は、IAM モードを選択できます。企業が外部 IdP を使用する場合

は、IAM または IAM IdC を介してフェデレーションできます。認証モードは、既存の SageMaker Studio ドメインでは更新できないため、本番環境の SageMaker Studio ドメインを作成する前に認証モードを決める必要があります。

SageMaker Studio を IAM モードで設定した場合、SageMaker Studio ユーザーは署名付き URL を介してアプリケーションにアクセスします。この URL にブラウザからアクセスすると、ユーザーは SageMaker Studio アプリに自動的にサインインされます。

IAM ユーザー

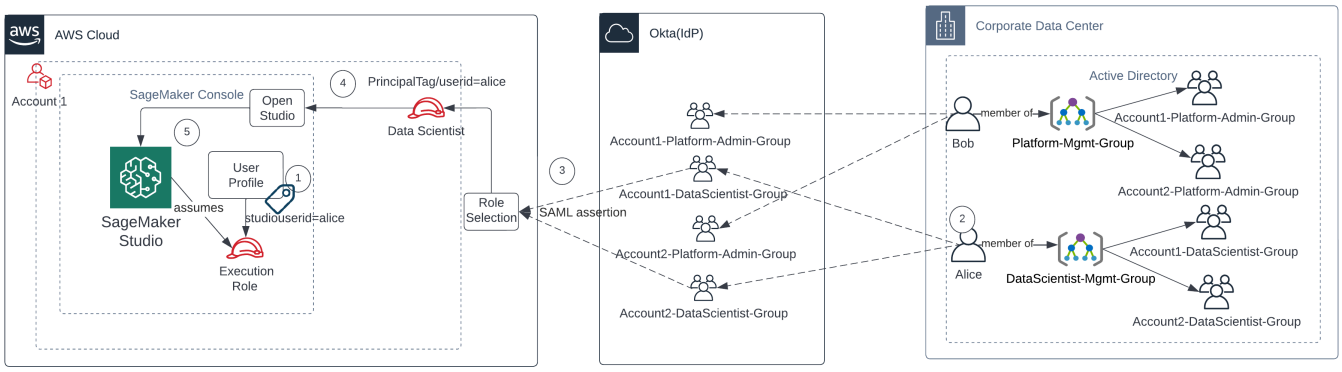
IAM ユーザーの場合、管理者はユーザーごとに SageMaker Studio ユーザープロファイルを作成して IAM ロールに関連付けます。このロールは、ユーザーが Studio 内から実行する必要があるアクションを許可します。AWS ユーザーが各自の SageMaker Studio ユーザープロファイルのみにアクセスできるように制限する場合、管理者は SageMaker Studio ユーザープロファイルにタグを付け、そのタグ値が AWS ユーザー名と一致したときにのみアクセスを許可する IAM ポリシーをユーザーにアタッチする必要があります。ポリシーステートメントは次のようになります。

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

AWS IAM またはアカウントフェデレーション

AWS アカウント フェデレーション方式により、お客様は Okta などの SAML IdP から SageMaker コンソールにフェデレーションできます。ユーザーが各自のユーザープロファイルのみにアクセス

できるように制限する場合、管理者は SageMaker Studio ユーザープロファイルにタグを付け、IdP に `PrincipalTags` を追加して一時的なタグとして設定します。次の図は、フェデレーションユーザー (データサイエンティストの Alice) が自分の SageMaker Studio ユーザープロファイルへのアクセスについて、どのように承認されるかを示しています。



IAM フェデレーションモードでの SageMaker Studio へのアクセス

1. SageMaker Studio の Alice のユーザープロファイルは、ユーザー ID がタグ付けされ、実行ロールに関連付けられています。
2. Alice が IdP (Okta) に対して認証を行います。
3. IdP は Alice を認証し、Alice がメンバーである 2 つのロール (アカウント 1 とアカウント 2 のデータサイエンティスト) で SAML アサーションをポストします。Alice は、アカウント 1 のデータサイエンティストのロールを選択します。
4. Alice はアカウント 1 の SageMaker コンソールにログインし、データサイエンティストのロールを引き受けます。Alice は、Studio アプリインスタンスのリストから自分の Studio アプリインスタンスを開きます。
5. 引き受けたロールセッションの Alice プリンシパルタグが、選択した SageMaker Studio アプリインスタンスのユーザープロファイルタグと照合して検証されます。プロファイルタグが有効な場合、SageMaker Studio アプリインスタンスが起動し、実行ロールを引き受けます。

ユーザーオンボーディングの一環として SageMaker Execution のロールとポリシーの作成を自動化する場合、これを実現する方法の 1 つは次のとおりです。

1. `SageMaker-Account1-Group` などの AD グループを各アカウントレベルと各 Studio ドメインレベルで設定します。
2. ユーザーを SageMaker Studio にオンボーディングする必要がある場合は、SageMaker-Account1-Group をユーザーのグループメンバーシップに追加します。

SageMaker-Account1-Group メンバーシップイベントをリッスンする自動プロセスを設定し、AWS API を使用して AD グループメンバーシップに基づくロール、ポリシー、タグ、SageMaker Studio ユーザープロファイルを作成します。ロールをユーザープロファイルにアタッチします。サンプルポリシーについては、「SageMaker Studio ユーザーが他のユーザープロファイルにアクセスできないようにする」を参照してください。

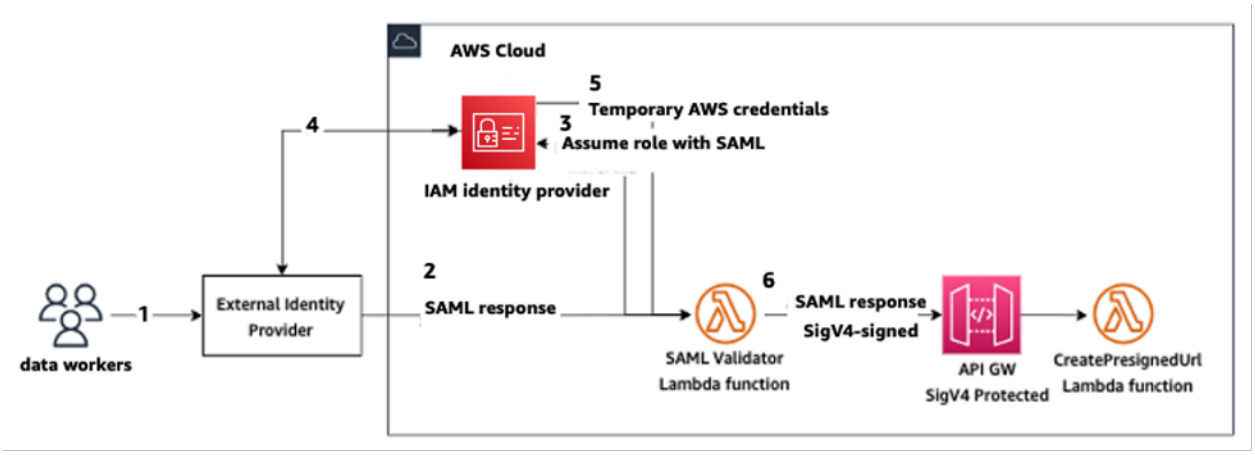
AWS Lambda を使用した SAML 認証

IAM モードでは、SAML アサーションを使用してユーザーを SageMaker Studio に対して認証することもできます。このアーキテクチャでは、お客様に既存の IdP があるため、ユーザーが (AWS ID フェデレーションアプリケーションの代わりに) Studio にアクセスするための SAML アプリケーションを作成できます。お客様の IdP が IAM に追加されます。AWS Lambda 関数は、IAM と STS を使用して SAML アサーションを検証し、API ゲートウェイまたは Lambda 関数を直接呼び出して署名付きドメイン URL を作成するのに役立ちます。

このソリューションの利点は、Lambda 関数で SageMaker Studio にアクセスするためのロジックをカスタマイズできることです。次に例を示します。

- ユーザープロファイルを自動的に作成します (まだ存在していない場合)。
- SAML 属性を解析して SageMaker Studio の実行ロールに対してロールやポリシードキュメントをアタッチまたは削除します。
- ライフサイクル設定 (LCC) やタグを追加して、ユーザープロファイルをカスタマイズします。

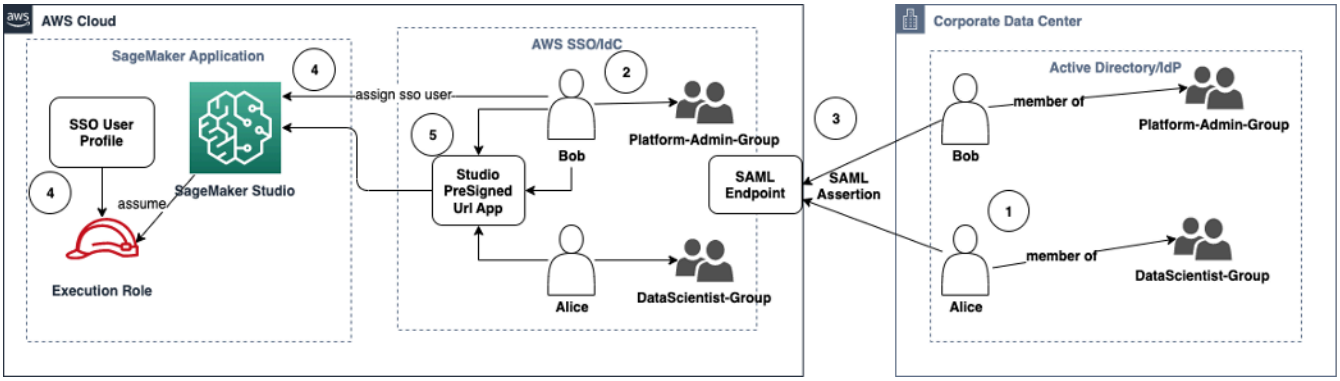
要約すると、このソリューションは、認証と承認のためのカスタムロジックを備えた SAML2.0 アプリケーションとして SageMaker Studio を公開します。実装の詳細については、付録の「SAML アサーションを使用した SageMaker Studio へのアクセス」セクションを参照してください。



カスタム SAML アプリケーションを使用した SageMaker Studio へのアクセス

AWS IAM IdC フェデレーション

IdC フェデレーション方式を使用すると、お客様は SAML IdP (Okta など) から SageMaker Studio アプリケーションに直接フェデレーションできます。次の図は、フェデレーションユーザーが各自の SageMaker Studio インスタンスへのアクセスをどのように承認されるかを示しています。



IAM IdC モードでの SageMaker Studio へのアクセス

1. 企業 AD の場合、ユーザーはプラットフォーム管理者グループやデータサイエンティストグループなどの AD グループのメンバーです。
2. ID プロバイダー (IdP) の AD ユーザーおよび AD グループは、AWS IAM アイデンティティセンターに同期され、それぞれシングルサインオンユーザーおよびグループとして割り当てることができます。
3. IdP は、SAML アサーションを AWS IdC SAML エンドポイントにポストします。
4. SageMaker Studio では、IdC ユーザーが SageMaker Studio アプリケーションに割り当てられます。この割り当ては、IdC グループを使用して行うことができ、SageMaker Studio は各 IdC ユーザーレベルに適用されます。この割り当てを作成すると、SageMaker Studio は IdC ユーザープロファイルを作成し、ドメイン実行ロールをアタッチします。
5. ユーザーは、IdC からクラウドアプリケーションとしてホストされている安全な署名付き URL を使用して SageMaker Studio アプリケーションにアクセスします。SageMaker Studio は、IdC ユーザープロファイルにアタッチされた実行ロールを引き受けます。

ドメインの認証に関するガイダンス

ドメインの認証モードを選択する場合の考慮事項は次のとおりです。

1. ユーザーが AWS Management Consoleにアクセスせずに SageMaker Studio UI を直接表示できるようにする場合は、AWS IAM IdC でシングルサインオンモードを使用します。

2. ユーザーが AWS Management Consoleにアクセスせずに SageMaker Studio UI を IAM モードで直接表示できるようにする場合は、バックエンドで Lambda 関数を使用してユーザープロファイル用の署名付き URL を生成し、ユーザーを SageMaker Studio UI にリダイレクトできます。
3. IdC モードでは、各ユーザーは 1 つのユーザープロファイルにマップされます。
4. IdC モードでは、すべてのユーザープロファイルに自動的にデフォルトの実行ロールが割り当てられます。ユーザーに異なる実行ロールを割り当てる場合は、UpdateUserProfile API を使用してユーザープロファイルを更新する必要があります。
5. インターネットを経由せずに (生成した署名付き URL を使用して) IAM モードで VPC エンドポイントへの SageMaker Studio UI アクセスを制限する場合は、カスタム DNS リゾルバーを使用できます。ブログ記事「Amazon SageMaker Studio の署名付き URL のセキュリティ保護パート 1: 基盤インフラストラクチャ」を参照してください。

アクセス許可の管理

このセクションでは、SageMaker Studio ドメインのプロビジョニングと運用によく使用される IAM ロール、ポリシー、およびガードレールを設定するためのベストプラクティスについて説明します。

IAM ロールとポリシー

ベストプラクティスとして、ML ライフサイクルに関わる関係者やアプリケーション (プリンシパルと呼ばれます) を最初に特定し、どの AWS アクセス許可をプリンシパルに付与する必要があるかを見極めます。SageMaker はマネージドサービスであるため、ユーザーに代わって API コールを実行できるサービスプリンシパル (AWS のサービス) を考慮する必要もあります。次の図は、組織内のさまざまなペルソナに応じて作成するさまざまな IAM ロールを示しています。

SageMaker の IAM ロール

これらのロールについて詳しく説明し、必要となる特定の IAM アクセス許可の例を示します。

- ML 管理者ユーザーロール — これは、データサイエンティストのために環境をプロビジョニングするプリンシパルであり、Studio ドメインとユーザープロファイルの作成 (`sagemaker:CreateDomain`、`sagemaker:CreateUserProfile`)、ユーザー用の AWS Key Management Service (AWS KMS) キーの作成、データサイエンティスト用の S3 バケットの作成、コンテナを格納するための Amazon ECR リポジトリの作成を行います。また、ユーザー用のデフォルト設定とライフサイクルスクリプトを設定したり、カスタムイメージを作成して SageMaker Studio ドメインにアタッチしたり、カスタムプロジェクトや Amazon EMR テンプレートなどのサービスカタログ製品を提供したりすることもできます。

  例えば、このプリンシパルはトレーニングジョブを実行しないため、SageMaker のトレーニングジョブや処理ジョブを起動するアクセス許可は不要です。CloudFormation や Terraform などのコードテンプレートとしてインフラストラクチャを使用してドメインやユーザーをプロビジョニングする場合、プロビジョニングサービスは、このロールを引き受けて管理者に代わってリソースを作成します。このロールには、AWS Management Consoleを使用して SageMaker への読み取り専用アクセス権を付与できます。

このユーザーロールには、プライベート VPC 内でドメインを起動するための特定の EC2 アクセス許可、EFS ボリュームを暗号化するための KMS アクセス許可、Studio のためにサービスにリンクされたロールを作成するアクセス許可 (`iam:CreateServiceLinkedRole`) も必要です。これらのアクセス許可の詳細については、このドキュメントの後半で説明します。

- データサイエンティストのユーザーロール — このプリンシパルは、SageMaker Studio にログインし、データを探索したり、処理/トレーニングジョブやパイプラインを作成したりするユーザーです。ユーザーが必要とする主なアクセス許可は、SageMaker Studio を起動するアクセス許可で、残りのポリシーは SageMaker 実行サービスロールで管理できます。
- SageMaker 実行サービスロール — SageMaker はマネージドサービスであるため、ユーザーに代わってジョブを起動します。このロールには、多くの場合、最も広範なアクセス許可が付与されます。多くのお客様は、トレーニングジョブ、処理ジョブ、またはモデルホスティングジョブを実行するために 1 つの実行ロールを使用することを選択するためです。これは簡単な開始方法ですが、お客様はジャーニー中に成熟すると、ノートブックの実行ロールを API アクション別に複数のロールに分割する場合があります。特に、デプロイされた環境でジョブを実行する場合はそうです。

  ロールは、作成時に SageMaker Studio ドメインに関連付けます。ただし、お客様は (職務などに基づいて) ドメイン内のユーザープロファイルごとに異なるロールを関連付けるという柔軟性を必要とする場合があるため、ユーザープロファイルごとに別個の IAM ロールを関連付けることもできます。1 人の物理ユーザーを 1 つのユーザープロファイルにマップすることをお勧めします。ユーザープロファイルの作成時にロールをアタッチしない場合、デフォルトの動作では、SageMakerStudio ドメインの実行ロールがユーザープロファイルに関連付けられます。

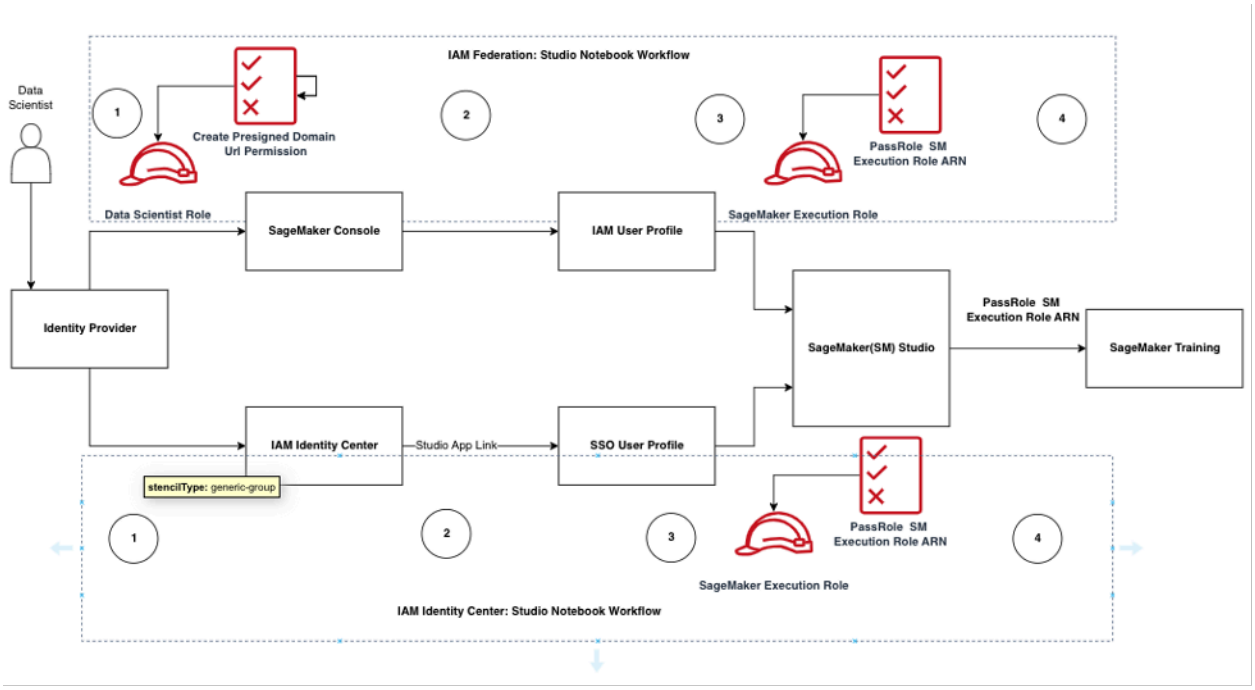
  複数のデータサイエンティストと ML エンジニアが 1 つのプロジェクトで協働し、リソースにアクセスするためのアクセス許可モデルを共有する必要がある場合は、チームレベルの SageMaker サービス実行ロールを作成してチームメンバー間で IAM アクセス許可を共有することをお勧めします。各ユーザーレベルでアクセス許可をロックダウンする必要がある場合は、ユーザーレベルの SageMaker サービス実行ロールを個別に作成できます。ただし、サービスの制限に注意する必要があります。

SageMaker Studio ノートブックの承認ワークフロー

このセクションでは、データサイエンティストが SageMaker Studio ノートブックから直接モデルを構築およびトレーニングするためにさまざまなアクティビティを実行する際に、各アクティビティに対して SageMaker Studio ノートブックの承認がどのように機能するかについて説明します。SageMaker ドメインは次の 2 つの承認モードをサポートしています。

- IAM フェデレーション
- IAM アイデンティティセンター

次に、このホワイトペーパーでは、モード別のデータサイエンティスト承認ワークフローを詳しく紹介します。



Studio ユーザー向けの認証と承認ワークフロー

IAM フェデレーション: SageMaker Studio ノートブックワークフロー

1. データサイエンティストは、企業 ID プロバイダーの認証を受け、SageMaker コンソールでデータサイエンティストのユーザーロール (ユーザーフェデレーションロール) を引き受けます。このフェデレーションロールには、ロールの Amazon リソースネーム (ARN) を SageMaker Studio に渡すための `iam:PassRole` API アクセス許可が SageMaker 実行ロールにあります。
2. データサイエンティストは、SageMaker 実行ロールに関連付けられている Studio IAM ユーザープロファイルから Open Studio リンクを選択します。
3. SageMaker Studio IDE サービスが起動し、ユーザープロファイルの SageMaker 実行ロールのアクセス許可を引き受けます。このロールには、ロールの ARN を SageMaker トレーニングサービスに渡すための `iam:PassRole` API アクセス許可が SageMaker 実行ロールにあります。
4. データサイエンティストがリモートコンピューティングノードでトレーニングジョブを起動すると、SageMaker 実行ロールの ARN が SageMaker トレーニングサービスに渡されます。これによ

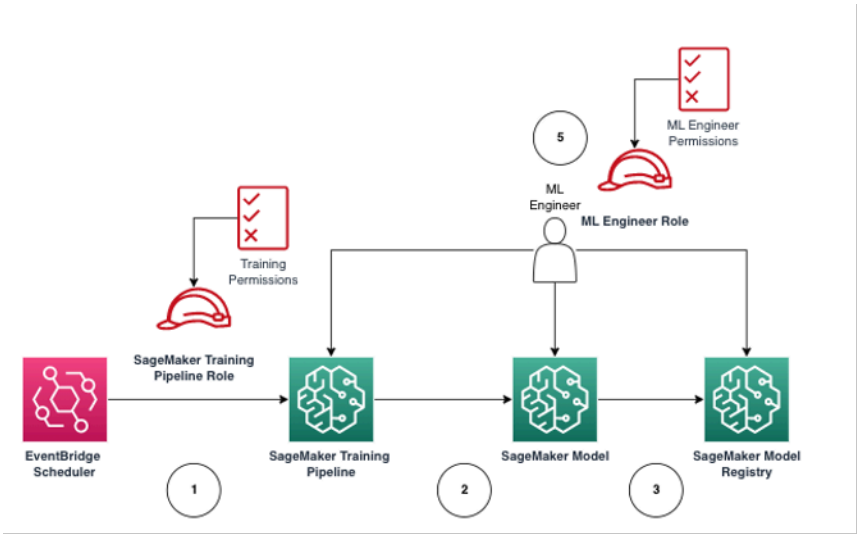
り、この ARN を使用して新しいロールセッションが作成され、トレーニングジョブが実行されます。トレーニングジョブのアクセス許可をさらに絞り込む必要がある場合は、トレーニング専用のロールを作成し、このロールの ARN をトレーニング API を呼び出すときに渡すことができます。

IAM アイデンティティセンター: SageMaker Studio ノートブックワークフロー

1. データサイエンティストは、企業 ID プロバイダーの認証を受け、AWS IAM アイデンティティセンターをクリックします。データサイエンティストには、ユーザー用のアイデンティティセンターポータルが表示されます。
2. データサイエンティストは、SageMaker 実行ロールに関連付けられた IdC ユーザープロファイルから作成された SageMaker Studio アプリリンクをクリックします。
3. SageMaker Studio IDE サービスが起動し、ユーザープロファイルの SageMaker 実行ロールのアクセス許可を引き受けます。このロールには、ロールの ARN を SageMaker トレーニングサービスに渡すための `iam:PassRole` API アクセス許可が SageMaker 実行ロールにあります。
4. データサイエンティストがリモートコンピューティングノードでトレーニングジョブを起動すると、SageMaker 実行ロールの ARN が SageMaker トレーニングサービスに渡されます。実行ロールの ARN は、この ARN を使用して新しいロールセッションを作成し、トレーニングジョブを実行します。トレーニングジョブのアクセス許可をさらに絞り込む必要がある場合は、トレーニング専用のロールを作成し、このロールの ARN をトレーニング API を呼び出すときに渡すことができます。

デプロイされた環境: SageMaker トレーニングワークフロー

システムテストや本番稼働などのデプロイされた環境では、ジョブは自動スケジューラとイベントトリガーを介して実行され、これらの環境に対する人間のアクセスは SageMaker Studio ノートブックからのものに制限されます。このセクションでは、デプロイされた環境で IAM ロールが SageMaker トレーニングパイプラインとどのように連携するかについて説明します。



マネージド本番環境での SageMaker トレーニングワークフロー

1. Amazon EventBridge スケジューラは SageMaker トレーニングパイプラインジョブをトリガーします。
2. SageMaker トレーニングパイプラインジョブは、モデルをトレーニングするための SageMaker トレーニングパイプラインロールを引き受けます。
3. トレーニング済みの SageMaker モデルは SageMaker モデルレジストリに登録されます。
4. ML エンジニアは ML エンジニアのユーザーロールを引き受け、トレーニングパイプラインと SageMaker モデルを管理します。

データのアクセス許可

SageMaker Studio ユーザーが任意のデータソースにアクセスできるかどうかは、SageMaker IAM 実行ロールに関連付けられたアクセス許可によって決まります。アタッチされたポリシーにより、特定の Amazon S3 バケットまたはプレフィックスの読み取り、書き込み、削除と、Amazon RDS データベースへの接続をユーザーに許可できます。

AWS Lake Formation データへのアクセス

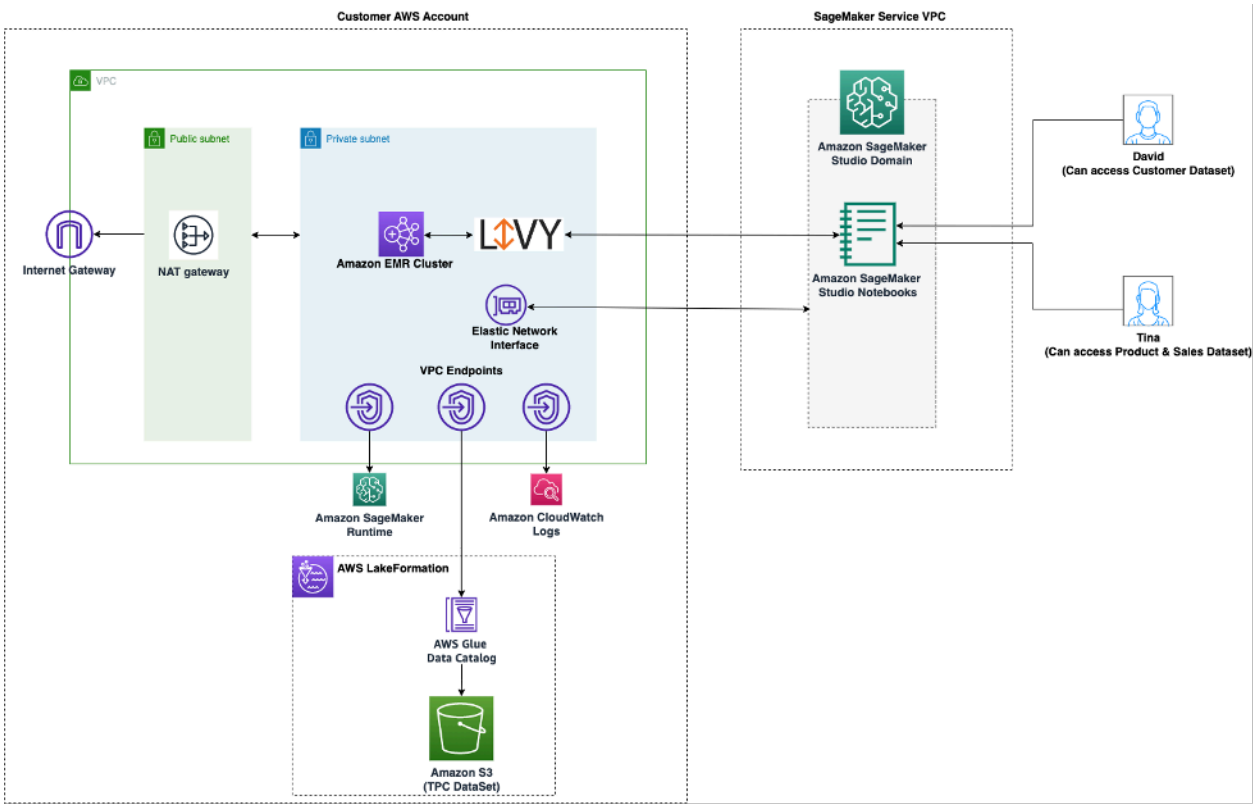
多くの企業は、AWS Lake Formation が管理するデータレイクの使用を開始し、ユーザー向けにきめ細かいデータアクセスを可能にしています。このような管理されたデータの例として、管理者は一部のユーザーに対して機密性の高い列をマスクしながら、同じ基になるテーブルのクエリを有効にすることができます。

SageMaker Studio から Lake Formation を利用する場合、管理者は SageMaker IAM 実行ロールを `DataLakePrincipals` として登録できます。詳細については、「Lake Formation 許可のリファレンス」を参照してください。承認された場合、SageMaker Studio の管理されたデータにアクセスして書き込むための方法が、次に示すように主に 3 つあります。

1. ユーザーは、SageMaker Studio ノートブックから、Amazon Athena などのクエリエンジンや boto3 上に構築されるライブラリを利用して、データをノートブックに直接取り込むことができます。AWS SDK for Pandas (旧 awswrangler) は人気のあるライブラリです。次のコード例は、これがどれほどシームレスになれるかを示しています。

```
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. Amazon EMR に対する SageMaker Studio のネイティブ接続を使用して、データの読み取りと書き込みを大規模に行います。Apache Livy と Amazon EMR のランタイムロールを使用することで、SageMaker Studio はネイティブ接続を構築しています。これにより、SageMaker 実行 IAM ロール (または他の承認されたロール) を Amazon EMR クラスターに渡して、データアクセスと処理を行うことができます。最新の手順については、「Studio から Amazon EMR クラスターに接続する」を参照してください。



Lake Formation が管理するデータに SageMaker Studio からアクセスするためのアーキテクチャ

3. AWS Glue インタラクティブセッションに対する SageMaker Studio のネイティブ接続を使用して大規模にデータを読み書きします。SageMaker Studio ノートブックには、ユーザーが AWS Glue に対してコマンドをインタラクティブに実行できるカーネルが組み込まれています。これにより、管理されたデータソースに対してデータのシームレスな読み書きを大規模に行うことができる Python、Spark、または Ray バックエンドのスケーラブルな使用が可能になります。カーネルを使用すると、ユーザーは SageMaker の実行ロールや他の承認された IAM ロールを渡すことができます。詳細については、「AWS Glue のインタラクティブセッションを使用してデータを準備する」を参照してください。

一般的なガードレール

このセクションでは、IAM ポリシー、リソースポリシー、VPC エンドポイントポリシー、サービスコントロールポリシー (SCP) を使用して ML リソースにガバナンスを適用する際に最もよく使用されるガードレールについて説明します。

ノートブックへのアクセスを特定のインスタンスに制限する

このサービスコントロールポリシーを使用すると、Studio ノートブックの作成中にデータサイエンティストがアクセスできるインスタンスタイプを制限できます。どのユーザーにも、SageMaker Studio をホストするデフォルトの Jupyter Server アプリの作成を許可された「システム」インスタンスが必要であることに注意してください。

```
{
     "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"
                     ]
                 }
             }
         }
     ]
 }
```

非準拠の SageMaker Studio ドメインを制限する

SageMaker Studio ドメインでは、次のサービスコントロールポリシーを使用して、お客様のリソースにアクセスするトラフィックがパブリックインターネットを経由せずに、お客様の VPC を経由するように強制できます。

```
{
     "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"
                }
            }
        }
    ]
}
```

未承認の SageMaker イメージの起動を制限する

次のポリシーは、ユーザーが自分のドメイン内で未承認の SageMaker イメージを起動できないようにします。

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

SageMaker VPC エンドポイント経由でのみノートブックを起動する

SageMaker コントロールプレーンの VPC エンドポイントに加えて、SageMaker はユーザーが SageMaker Studio ノートブックまたは SageMaker ノートブックインスタンスに接続するための VPC エンドポイントをサポートしています。SageMaker Studio/ノートブックインスタンスの VPC エンドポイントを既に設定している場合、次の IAM 条件キーは、SageMaker Studio VPC エンドポイント経由または SageMaker API エンドポイント経由での SageMaker Studio ノートブックへの接続のみを許可します。

```
{
     "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 Studio ノートブックへのアクセスを特定の IP 範囲に制限する

多くの場合、企業は SageMaker Studio へのアクセスを特定の許可された企業 IP 範囲に制限します。次の IAM ポリシーで SourceIP 条件キーを使用すると、これを制限できます。

```
{
     "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 Studio ユーザーが他のユーザープロファイルにアクセスできないようにする

管理者は、ユーザープロファイルを作成するときに、必ずタグキー studiouserid を使用してプロファイルに SageMaker Studio ユーザー名をタグ付けしてください。プリンシパル (ユーザーまたはユーザーにアタッチされたロール) にもキー studiouserid でタグ付けする必要があります (このタグには任意の名前を使用可能であり、studiouserid に限定されません)。

次に、SageMaker Studio の起動時にユーザーが引き受けるロールに次のポリシーをアタッチします。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AmazonSageMakerPresignedUrlPolicy",
            "Effect": "Allow",
            "Action": [
                "sagemaker:CreatePresignedDomainUrl"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
```

```
                    "sagemaker:ResourceTag/studiouserid": "${aws:PrincipalTag/
studiouserid}"
                }
            }
        }
    ]
}
```

タグ付けを強制する

データサイエンティストは、SageMaker Studio ノートブックを使用してデータを探索し、モデルを構築およびトレーニングする必要があります。ノートブックにタグを付けると、使用状況のモニタリングやコストの管理に加えて、所有権や監査可能性の確認にも役立ちます。

SageMaker Studio アプリの場合は、ユーザープロファイルがタグ付けされていることを確認してください。タグは、ユーザープロファイルからアプリに自動的に伝達されます。ユーザープロファイルの作成時にタグの使用を強制する (CLI および SDK でサポート) には、次のポリシーを管理者ロールに追加することを検討してください。

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

トレーニングジョブや処理ジョブなどの他のリソースでは、次のポリシーを使用してタグを必須にすることができます。

```
{
```

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

SageMaker Studio のルートアクセス

SageMaker Studio の場合、ノートブックは Docker コンテナで実行されます。デフォルトでは、このコンテナにはホストインスタンスへのルートアクセス権がありません。同様に、デフォルトのユーザーとして実行する場合を除き、コンテナ内の他のすべてのユーザー ID 範囲は、ホストインスタンス自体に非特権ユーザー ID として再マップされます。そのため、権限昇格の脅威はノートブックコンテナ自体に限定されます。

カスタムイメージを作成する場合、より厳密な制御を行うために、ユーザーにルート以外のアクセス許可を付与できます。例えば、望ましくないプロセスをルートとして実行しないようにしたり、公的に入手可能なパッケージをインストールしたりできるようにします。このような場合は、Dockerfile 内でルート以外のユーザーとして実行するイメージを作成できます。ユーザーをルートとして作成するか、ルート以外として作成するかにかかわらず、ユーザーの UID/GID が、カスタムアプリの AppImageConfig の UID/GID と同一であることを確認する必要があります。これにより、SageMaker がカスタムイメージを使用してアプリを実行するための設定が作成されます。例えば、Dockerfile が次のように非ルートユーザー向けにビルドされているとします。

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

```
USER $NB_UID
```

AppImageConfig ファイルでは、次のように KernelGatewayConfig に同じ UID と GID を記述する必要があります。

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

カスタムイメージで使用できる UID/GID 値は 0/0、Studio イメージの場合は 1000/100 です。カスタムイメージの作成例および関連する AppImageConfig 設定については、こちらの Github リポジトリを参照してください。

ユーザーがこれを改ざんしないように、SageMaker Studio ノートブックユーザーに CreateAppImageConfig、UpdateAppImageConfig、または DeleteAppImageConfig のアクセス許可を付与しないでください。

ネットワーク管理

SageMaker Studio ドメインを設定するには、VPC ネットワーク、サブネット、セキュリティグループを指定する必要があります。VPC とサブネットを指定するときは、以下のセクションで説明する使用量と予想される増加率を考慮して IP を割り当ててください。

VPC ネットワークの計画

SageMaker Studio ドメインに関連するお客様の VPC サブネットは、以下の要因に応じて、適切な Classless Inter-Domain Routing (CIDR) 範囲を使用して作成する必要があります。

- ユーザー数。
- ユーザーごとのアプリの数。
- ユーザーごとの一意なインスタンスタイプの数。
- ユーザーごとのトレーニングインスタンスの平均数。
- 予想される増加率。

以下のユースケースにおいて、SageMaker および AWS の参加サービスは、Elastic Network Interface (ENI) をお客様の VPC サブネット内に挿入します。

- Amazon EFS は SageMaker ドメインの EFS マウントターゲットに ENI を挿入します (SageMaker ドメインにアタッチされたサブネット/アベイラビリティーゾーンごとに 1 つの IP)。
- SageMaker Studio は、ユーザープロファイルまたは共有スペースで使用する一意のインスタンスごとに ENI を挿入します。次に例を示します。
  - ユーザープロファイルがデフォルトの Jupyter サーバーアプリ (1 つの「システム」インスタンス)、データサイエンスアプリと Base Python アプリ (両方とも `ml.t3.medium` インスタンス上で実行) を実行する場合、Studio は 2 つの IP アドレスを挿入します。
  - ユーザープロファイルがデフォルトの Jupyter サーバーアプリ (1 つの「システム」インスタンス)、Tensorflow GPU アプリ (`ml.g4dn.xlarge` インスタンス上)、データラングラーアプリ (`ml.m5.4xlarge` インスタンス上) を実行する場合、Studio は 3 つの IP アドレスを挿入します。
- ドメインの VPC サブネット/アベイラビリティーゾーンにわたって VPC エンドポイントごとに ENI が挿入されます (SageMaker VPC エンドポイントには 4 つの IP が挿入され、S3、ECR、CloudWatch などの参加サービスには最大 6 つの IP が挿入されます)。

- SageMaker のトレーニングジョブと処理ジョブを同じ VPC 設定で起動する場合、各ジョブにはインスタンスごとに 2 つの IP アドレスが必要です。

> **Note**
>
> サブネットや VPC 専用トラフィックなどの SageMaker Studio の VPC 設定は、SageMaker Studio で作成したトレーニング/処理ジョブには自動的に渡されません。ユーザーは Create*Job API を呼び出すときに、状況に応じて VPC 設定とネットワーク分離を設定する必要があります。詳細については、「トレーニングおよび推論コンテナをインターネット無料モードで実行する」を参照してください。

シナリオ: データサイエンティストが 2 つの異なるインスタンスタイプで実験を行う

このシナリオでは、SageMaker ドメインが VPC 専用トラフィックモードで設定されていると仮定します。SageMaker API、SageMaker ランタイム、Amazon S3、Amazon ECR などの VPC エンドポイントが設定されています。

データサイエンティストが Studio ノートブックで実験を行い、 2 つの異なるインスタンスタイプ (`ml.t3.medium` と `ml.m5.large` など) で実行し、インスタンスタイプごとに 2 つのアプリを起動します。

データサイエンティストが `ml.m5.4xlarge` インスタンスでも同じ VPC 設定で同時にトレーニングジョブを実行していると仮定します。

このシナリオの場合、SageMaker Studio サービスは次のように ENI を挿入します。

表 1 — 実験シナリオでお客様の VPC に挿入された ENI

エンティティ	ターゲット	ENI の挿入数	メモ	レベル
EFS マウントターゲット	VPC サブネット	3	3 つの AZ/サブネット	ドメイン
VPC エンドポイント	VPC サブネット	30	3 つの AZ/サブネット (それぞれに 10 個の VPCE)	ドメイン

エンティティ	ターゲット	ENI の挿入数	メモ	レベル
Jupyter Server	VPC サブネット	1	インスタンスごとに 1 つの IP	ユーザー
KernelGateway アプリ	VPC サブネット	2	インスタンスタイプごとに 1 つの IP	ユーザー
トレーニング	VPC サブネット	2	トレーニングインスタンスごとに 2 つの IP EFA を使用する場合は、トレーニングインスタンスごとに 5 つの IP	ユーザー

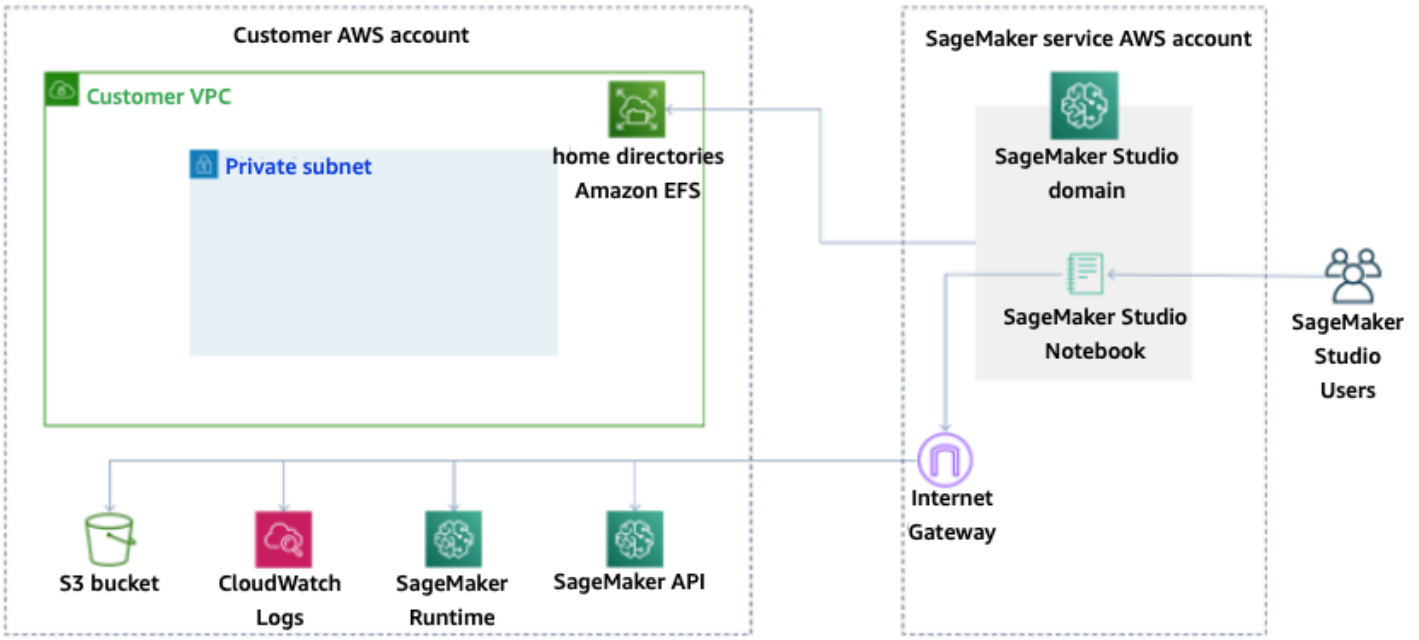
このシナリオでは、お客様の VPC で合計 38 個の IP を消費します。そのうち、33 個の IP はドメインレベルでユーザー間で共有し、5 個の IP はユーザーレベルで消費します。このドメインで同様のユーザープロファイルを持つ 100 人のユーザーがこれらのアクティビティを同時に実行する場合、ユーザーレベルでは 5×100 = 500 個の IP を消費することになります。これとは別に、ドメインレベルではサブネットごとに 11 個の IP を消費するため、合計で 511 個の IP になります。このシナリオの場合、VPC サブネット CIDR を /22 で作成する必要があります。これにより、1024 個の IP アドレスが割り当てられ、拡張の余地が残ります。

VPC ネットワークのオプション

SageMaker Studio ドメインは、以下のオプションのいずれかを使用した VPC ネットワークの設定をサポートしています。

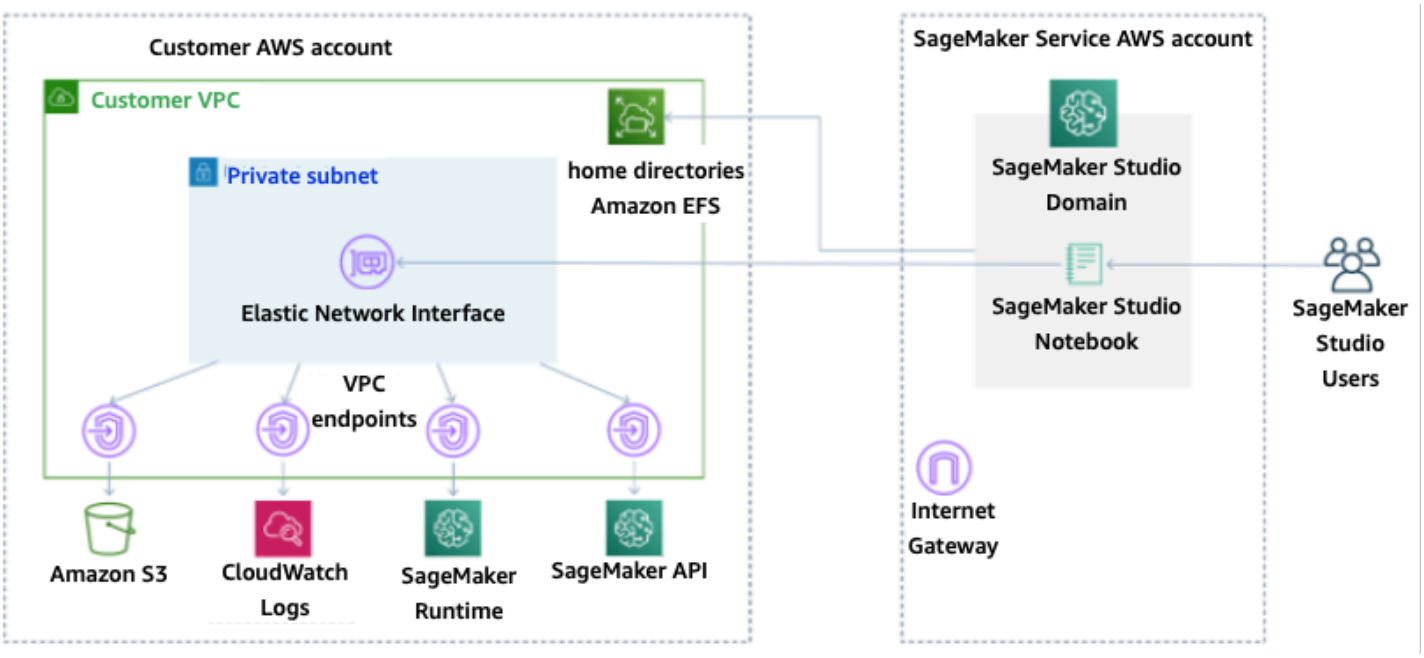
- パブリックインターネット専用
- VPC 専用

パブリックインターネット専用オプションを使用すると、次の図に示すように、SageMaker API サービスは、SageMaker サービスアカウントが管理する VPC でプロビジョニングされたインターネットゲートウェイ経由でパブリックインターネットを使用できます。



デフォルトモード: SageMaker サービスアカウント経由のインターネットアクセス

VPC 専用オプションを使用すると、次の図に示すように、SageMaker サービスアカウントが管理する VPC からのインターネットルーティングが無効になり、お客様は VPC エンドポイント経由のトラフィックルーティングを設定できます。



VPC 専用モード: SageMaker サービスアカウント経由のインターネットアクセスなし

VPC 専用モードで設定したドメインの場合は、ユーザープロファイルごとにセキュリティグループを設定して、基盤となるインスタンスを完全に分離します。AWS アカウント内のドメインごとに独自の VPC 設定とインターネットモードを使用できます。VPC ネットワーク設定の構成の詳細については、「VPC 内の SageMaker Studio ノートブックを外部リソースに接続する」を参照してください。

制限事項

- SageMaker Studio ドメインの作成後に、新しいサブネットを関連付けることはできません。
- VPC ネットワークタイプ (パブリックインターネット専用または VPC 専用) を変更することはできません。

データ保護

ML ワークロードを設計する前に、セキュリティに影響する基本的なプラクティスを整えておく必要があります。例えば、データ分類は機密レベルに基づいてデータを分類する方法を提供し、暗号化は不正アクセスに対してデータを解読できないようにすることでデータを保護します。これらの方法は、誤操作の防止や規制義務の遵守などの目的に役立つため、重要です。

SageMaker Studio には、保管中および転送中のデータを保護するための機能がいくつか用意されています。ただし、「AWS 責任共有モデル」で説明しているように、AWS グローバルインフラストラクチャでホストされているコンテンツを管理する責任はお客様にあります。このセクションでは、お客様がこれらの機能を使用してどのようにデータを保護できるかについて説明します。

保管中のデータの保護

Amazon SageMaker Studio ノートブックとモデル構築データおよびモデルアーティファクトを保護するために、SageMaker ではノートブックと、トレーニングジョブやバッチ変換ジョブの出力を暗号化します。SageMaker は、Amazon S3 の AWS マネージドキーを使用して、これらをデフォルトで暗号化します。この Amazon S3 の AWS マネージドキーは、クロスアカウントアクセスでは共有できません。クロスアカウントアクセスの場合は、SageMaker リソースの作成時にカスタマーマネージドキーを指定してクロスアカウントアクセスで共有できるようにします。

SageMaker Studio では、データを次の場所に保存できます。

- S3 バケット — 共有可能なノートブックが有効になっている場合、SageMaker Studio はノートブックのスナップショットとメタデータを S3 バケットで共有します。
- EFS ボリューム — SageMaker Studio は、EFS ボリュームをドメインにアタッチしてノートブックとデータファイルを保存します。この EFS ボリュームは、ドメインを削除した後も保持されます。
- EBS ボリューム — ノートブックが動作しているインスタンスに EBS をアタッチします。このボリュームは、インスタンスの存続期間中、保持されます。

AWS KMS による保管中の暗号化

- AWS KMS キーを渡して、ノートブック、トレーニング、チューニング、バッチ変換ジョブ、エンドポイントに接続されている EBS ボリュームを暗号化できます。

- KMS キーを指定しない場合、SageMaker はオペレーティングシステム (OS) のボリュームと ML データボリュームの両方をシステムマネージドの KMS キーで暗号化します。
- コンプライアンス上の理由から KMS キーを使用して暗号化する必要がある機密データは、ML ストレージボリュームまたは Amazon S3 に保存する必要があります。どちらの場合も、必要に応じて、指定した KMS キーを使用して暗号化できます。

転送中のデータの保護

Amazon SageMaker は、ML モデルのアーティファクトや他のシステムアーティファクトを転送中と保管中に確実に暗号化します。SageMaker の API とコンソールに対するリクエストには、安全な SSL 接続が使用されます。転送中のネットワーク内データ (サービスプラットフォーム内) の一部は暗号化されません。これには、以下のものが含まれます：

- サービスコントロールプレーンとトレーニングジョブインスタンス (顧客データではない) の間のコマンドとコントロールの通信。
- 分散処理ジョブとトレーニングジョブ (ネットワーク内) のノード間の通信。

ただし、トレーニングクラスター内のノード間の通信は、暗号化することを選択できます。コンテナ間のトラフィック暗号化を有効にすると、特に分散型深層学習アルゴリズムを使用している場合に、トレーニング時間が増える可能性があります。

デフォルトでは、Amazon SageMaker はトレーニングジョブを Amazon VPC で実行し、データのセキュリティを確保します。プライベート VPC を設定することで、別のレベルのセキュリティを追加して、トレーニングコンテナとデータを保護することができます。さらに、SageMaker Studio ドメインを VPC 専用モードで実行するように設定し、インターネット経由でトラフィックを出力せずにプライベートネットワーク経由でトラフィックをルーティングするように VPC エンドポイントを設定できます。

データ保護のガードレール

保管中の SageMaker ホスティングボリュームの暗号化

オンライン推論用の SageMaker エンドポイントのホスティング中に暗号化を適用するには、次のポリシーを使用します。

```
{
```

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

モデルモニタリング中に使用する S3 バケットの暗号化

モデルモニタリング では、SageMaker エンドポイントに送信されたデータをキャプチャし、S3 バケットに保存します。データキャプチャ設定をセットアップするときは、S3 バケットを暗号化する必要があります。現在、これに代わる制御は他にありません。

モデルモニタリングサービスは、エンドポイント出力をキャプチャするだけでなく、事前に指定したベースラインに対するドリフトもチェックします。ドリフトのモニタリングに使用する出力と中間ストレージボリュームを暗号化する必要があります。

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

```
                "sagemaker:OutputKmsKey": "false"
            }
        }
    }
  ]
}
```

SageMaker Studio ドメインのストレージボリュームの暗号化

Studio ドメインにアタッチしたストレージボリュームに暗号化を適用します。このポリシーでは、スタジオドメインにアタッチしたストレージボリュームを暗号化するための CMK をユーザーが提供する必要があります。

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

ノートブックの共有に使用する S3 に保存されているデータの暗号化

これは、SageMaker Studio ドメイン内のユーザー間でノートブックを共有するために使用するバケットに保存されているデータをすべて暗号化するポリシーです。

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

```
            "Sid": "EncryptDomainSharingS3Bucket",
            "Effect": "Allow",
            "Action": [
                "sagemaker:CreateDomain",
                "sagemaker:UpdateDomain"
            ],
            "Resource": "*",
            "Condition": {
                "Null": {
                    "sagemaker:DomainSharingOutputKmsKey": "false"
                }
            }
        }
    ]
}
```

制限事項

- ドメインの作成後は、アタッチした EFS ボリュームストレージをカスタム AWS KMS で更新することはできません。
- 作成後のトレーニング/処理ジョブやエンドポイント設定を KMS キーで更新することはできません。

ログ記録とモニタリング

アルゴリズムコンテナ、モデルコンテナ、またはノートブックインスタンスのライフサイクル設定から stdout または stderr に送信されるものは、すべて Amazon CloudWatch Logs にも送信されて、コンパイルジョブ、処理ジョブ、トレーニングジョブ、エンドポイント、変換ジョブ、ノートブックインスタンス、およびノートブックインスタンスのライフサイクル設定をデバッグするために利用されます。Amazon CloudWatch を使用して SageMaker をモニタリングすることで、raw データを収集し、ほぼリアルタイムで読み取り可能なメトリクスに加工できます。これらの統計は 15 か月間保持されるため、履歴情報にアクセスして、ウェブアプリケーションまたはサービスの動作をより的確に把握できます。

CloudWatch でのログ記録

データサイエンスのプロセスは本質的に実験的で反復的であるため、ノートブックの使用状況、トレーニング/処理ジョブの実行時間、トレーニングのメトリクス、エンドポイントサービスのメトリクス (呼び出しのレイテンシーなど) などのアクティビティを口グに記録することが重要です。デフォルトでは、SageMaker はメトリクスを CloudWatch Logs に公開します。これらの口グは、AWS KMS を使用してカスタマーマネージドキーで暗号化できます。

VPC エンドポイントを使用して、パブリックインターネットを使用せずに CloudWatch にログを送信することもできます。また、特定のしきい値をモニタリングするアラームを設定し、しきい値に達したときに通知を送信したりアクションを実行したりできます。詳細については、「Amazon CloudWatch ユーザーガイド」を参照してください。

SageMaker は、Studio 用に 1 つのロググループを `/aws/sagemaker/studio` に作成します。ユーザープロファイルとアプリごとに、このロググループの下に独自のログストリームがあり、ライフサイクル設定スクリプトにも独自のログストリームがあります。例えば、「studio-user」という名前のユーザープロファイルと Jupyter Server アプリ、アタッチされたライフサイクルスクリプト、およびデータサイエンスカーネルゲートウェイアプリには、以下のログストリームがあります。

```
/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 がユーザーに代わって CloudWatch にログを送信するには、トレーニング/処理/変換ジョブ API の呼び出し元に次のアクセス許可が必要です。

```
{
     "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"
         }
     ]
 }
```

これらのログをカスタム AWS KMS キーで暗号化するには、まずキーポリシーを変更して CloudWatch サービスがキーを暗号化および復号できるようにする必要があります。ログ暗号化の AWS KMS キーを作成したら、キーポリシーを変更して以下を含めます。

```
{
    "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:*"
                }
            }
        }
    ]
}
```

暗号化する CloudWatch ログには、いつでも `ArnEquals` を使用して特定の Amazon リソースネーム (ARN) を指定できることに注意してください。ここでは、簡単にするために、このキーを使用してアカウント内のすべてのログを暗号化できることを示しています。さらに、トレーニング、処理、モデルエンドポイントは、インスタンスの CPU とメモリの使用率、ホスティングの呼び出しのレイテンシーなどに関するメトリクスを公開します。さらに、特定のしきい値を超えたときにイベントを管理者に通知するように Amazon SNS を設定することもできます。トレーニングおよび処理 API のコンシューマーには、次のアクセス許可が必要です。

```
{
     "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"
        }
    ]
}
```

AWS CloudTrail での監査

コンプライアンス体制を改善するには、AWS CloudTrail を使用してすべての API を監査します。デフォルトでは、すべての SageMaker API が AWS CloudTrail でログに記録されます。CloudTrail を有効にするために追加の IAM アクセス許可は必要ありません。

`InvokeEndpoint` と `InvokeEndpointAsync` を除くすべての SageMaker のアクションは、CloudTrail によってログに記録され、オペレーションに文書化されます。例えば `CreateTrainingJob`、`CreateEndpoint`、`CreateNotebookInstance` の各アクションに対する呼び出しにより、CloudTrail ログファイルにエントリが生成されます。

CloudTrail の各イベントエントリには、リクエストの生成元に関する情報が含まれます。このアイデンティティ情報は以下のことを判断するのに役立ちます：

- リクエストが、ルートと AWS IAM ユーザー認証情報のどちらを使用して送信されたか。
- リクエストがロールまたはフェデレーションユーザーの一時的なセキュリティ認証情報を使用して行われたかどうか。
- リクエストが、別の AWS のサービスによって送信されたかどうか。イベントの例については、「SageMaker API コールを CloudTrail でログに記録する」ドキュメントを参照してください。

デフォルトでは、CloudTrail は、ユーザープロファイルの Studio 実行ロール名を各イベントの識別子としてログに記録します。これは、各ユーザーが独自の実行ロールを持っている場合に有効です。複数のユーザーが同じ実行ロールを共有している場合は、`sourceIdentity` 設定を使用して Studio のユーザープロファイル名を CloudTrail に伝播できます。`sourceIdentity` 機能を有効に

するには、「Amazon SageMaker Studio からのユーザーリソースアクセスのモニタリング」を参照してください。共有スペースでは、すべてのアクションがスペース ARN をソースとして参照するため、`sourceIdentity` を通じて監査することはできません。

コスト属性

SageMaker Studio には、管理者がドメイン、共有スペース、ユーザーの支出を個別に追跡するのに役立つ機能が組み込まれています。

自動タグ付け

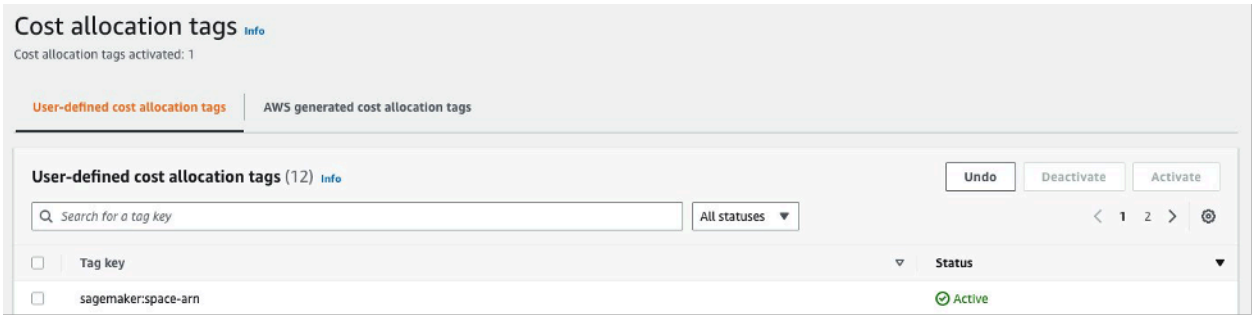
SageMaker Studio では、トレーニングジョブ、処理ジョブ、カーネルアプリなどの新しい SageMaker リソースに、個別の `sagemaker:domain-arn` を自動的にタグ付けするようになりました。より詳細なレベルの場合、SageMaker はリソースの主な作成者を示すためにリソースに `sagemaker:user-profile-arn` または `sagemaker:space-arn` もタグ付けします。

SageMaker ドメインの EFS ボリュームには、`ManagedByAmazonSageMakerResource` という名前のキーとドメイン ARN の値がタグ付けされます。ユーザーごとのスペース使用量を把握するための詳細なタグはありません。ただし、管理者は EFS ボリュームを EC2 インスタンスにアタッチして、モニタリングをカスタマイズできます。

コストモニタリング

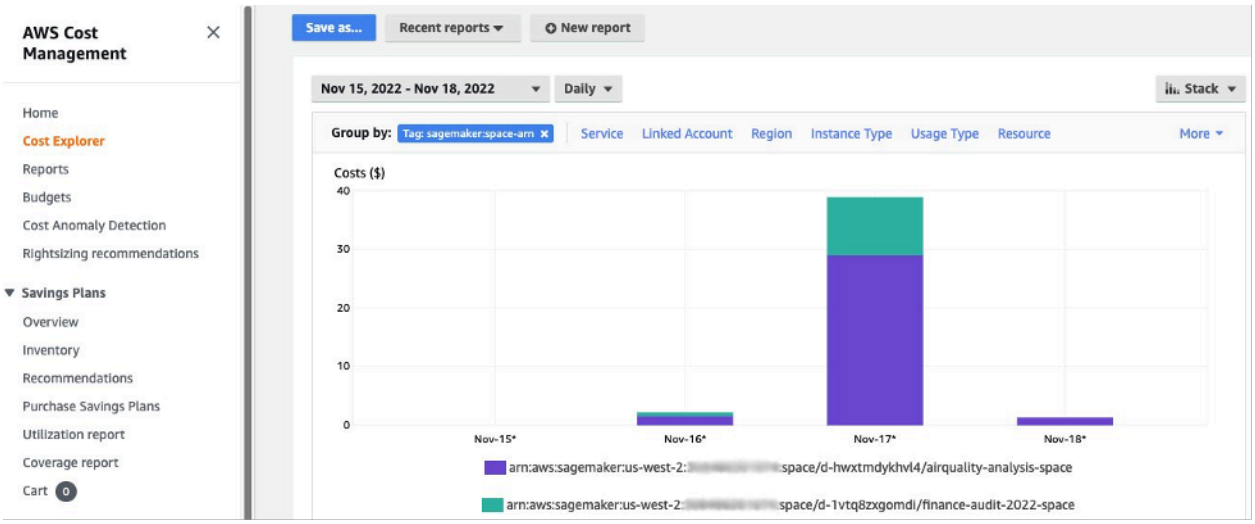
自動タグにより、管理者は ML の支出を追跡、レポート、モニタリングするために、AWS Cost Explorer や AWS Budgets などのすぐに使えるソリューションを使用することも、AWS のコストと使用状況レポート (CURS) のデータに基づいて構築したカスタムソリューションを使用することもできます。

アタッチしたタグをコスト分析に使用するには、まず AWS Billing コンソールの [コスト配分タグ] セクションでタグを有効にする必要があります。タグがコスト配分タグパネルに表示されるまでに最大 24 時間かかる場合があるため、タグを有効にする前に SageMaker リソースを作成する必要があります。



Cost Explorer でコスト配分タグとして有効になっているスペース ARN

コスト配分タグを有効にすると、AWS はタグ付けされたリソースの追跡を開始します。タグは、24～48 時間後に選択可能なフィルターとして Cost Explorer に表示されます。



サンプルドメインの共有スペース別にグループ化したコスト

コスト管理

最初の SageMaker Studio ユーザーをオンボーディングすると、SageMaker はドメインの EFS ボリュームを作成します。ノートブックとデータファイルをユーザーのホームディレクトリに保存すると、この EFS ボリュームにはストレージコストが発生します。ユーザーが Studio ノートブックを起動すると、ノートブックを実行しているコンピューティングインスタンスでノートブックが起動されます。コストの詳細な内訳については、「Amazon SageMaker の料金」を参照してください。

管理者は、コンピューティングコストを管理するために「一般的なガードレール」セクションの説明にあるように IAM ポリシーを使用して、ユーザーがスピンアップできるインスタンスのリストを指定できます。また、お客様がコストを節約するには、SageMaker の Studio の自動シャットダウン拡張機能を使用してアイドル状態のアプリを自動的にシャットダウンすることをお勧めします。このサーバー拡張機能は、ユーザープロファイルごとに実行中のアプリを定期的にポーリングし、管理者が設定したタイムアウトに基づいてアイドル状態のアプリをシャットダウンします。

この拡張機能をドメイン内のすべてのユーザーに設定するには、「カスタマイズ」セクションで説明しているライフサイクル設定を使用できます。さらに、拡張機能チェッカーを使用して、ドメインのすべてのユーザーに拡張機能がインストールされていることを確認することもできます。

カスタマイズ

ライフサイクル設定

ライフサイクル設定は、新しい SageMaker Studio ノートブックの開始などの SageMaker Studio ライフサイクルイベントによって開始されるシェルスクリプトです。これらのシェルスクリプトを使用して、カスタムパッケージのインストール、Jupyter 拡張機能による非アクティブなノートブックアプリの自動シャットダウン、Git 設定のセットアップなど、SageMaker Studio 環境のカスタマイズを自動化できます。ライフサイクル設定の構築方法の詳細については、こちらのブログ「ライフサイクル設定を使用して Amazon SageMaker Studio をカスタマイズする」を参照してください。

SageMaker Studio ノートブックのカスタムイメージ

Studio ノートブックには、Amazon SageMaker Python SDK および最新バージョンの IPython ランタイムまたはカーネルで構成される構築済みイメージのセットが付属しています。この機能を使用すると、独自のカスタムイメージを Amazon SageMaker ノートブックに持ち込むことができます。これにより、これらのイメージは、ドメイン内に認証されたすべてのユーザーが利用できるようになります。

デベロッパーやデータサイエンティストは、以下のようないくつかの異なるユースケースでカスタムイメージが必要になる場合があります。

- TensorFlow、MXNet、PyTorch などの一般的な ML フレームワークの特定バージョンまたは最新バージョンにアクセスします。
- ローカルで開発されたカスタムコードやアルゴリズムを SageMaker Studio ノートブックに持ち込んで、イテレーションやモデルトレーニングを迅速化します。
- API を介してデータレイクやオンプレミスのデータストアにアクセスします。管理者は、対応するドライバーをイメージに含める必要があります。
- IPython 以外の、カーネルと呼ばれるバックエンドランタイム (R、Julia、その他) にアクセスします。また、説明されている方法を使用してカスタムカーネルをインストールすることもできます。

カスタムイメージの作成方法の詳細については、「SageMaker のカスタムイメージの作成」を参照してください。

JupyterLab 拡張機能

SageMaker Studio JuypterLab 3 ノートブックを使用すると、増え続けるオープンソースの JupyterLab 拡張機能のコミュニティを活用できます。このセクションでは、SageMaker のデベロッパーワークフローに該当する数例を取り上げますが、すべての利用可能な拡張機能を参照するか、さらには自分で作成することをお勧めします。

JupyterLab 3 では、拡張機能のパッケージ化とインストールのプロセスが大幅に簡易化されました。前述の拡張機能は、bash スクリプトを使用してインストールできます。例えば、SageMaker Studio では、Studio ランチャーからシステムターミナルを開いて、次のコマンドを実行します。さらに、ライフサイクル設定を使用して、これらの拡張機能のインストールを自動化し、Studio を再起動しても設定が保持されるようにすることができます。これは、ドメイン内のすべてのユーザーに対して設定することも、個々のユーザーレベルで設定することもできます。

例えば、Amazon S3 ファイルブラウザ用の拡張機能をインストールするには、システムターミナルで次のコマンドを実行し、ブラウザを必ず更新します。

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

JupyterLab ノートブックのバージョン 1 と 3 の両方で動作するライフサイクル設定を記述して下位互換性を維持する方法など、拡張機能の管理の詳細については、「JupyterLab および Jupyter Server 拡張機能のインストール」を参照してください。

Git リポジトリ

SageMaker Studio には Jupyter Git 拡張機能がプリインストールされており、ユーザーは Git リポジトリのカスタム URL の入力、EFS ディレクトリへのクローン作成、変更のプッシュ、コミット履歴の表示を行うことができます。管理者は、推奨される git リポジトリをドメインレベルで設定し、エンドユーザーがドロップダウンで選択できるようにすることができます。最新の手順については、「推奨される Git リポジトリを Studio にアタッチする」を参照してください。

リポジトリがプライベートの場合、拡張機能は、標準の git インストールを使用して認証情報をターミナルに入力するようユーザーに求めます。また、ユーザーは個々の EFS ディレクトリに ssh 認証情報を保存して、管理を容易にすることもできます。

Conda 環境

SageMaker Studio ノートブックは Amazon EFS を永続的なストレージレイヤーとして使用します。データサイエンティストは、永続的なストレージを利用してカスタム Conda 環境を作成し、これらの環境を使用してカーネルを作成できます。これらのカーネルは、EFS に支えられており、カーネル、アプリ、または Studio の再起動間で保持されます。Studio は、すべての有効な環境を KernelGateway カーネルとして自動的に選択します。

Conda 環境を作成するプロセスは、データサイエンティストには簡単ですが、カーネルがカーネルセレクターに設定されるまでに約 1 分かかります。環境を作成するには、システムターミナルで次のコマンドを実行します。

```
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
```

詳細な手順については、「Amazon SageMaker Studio ノートブックで Python パッケージを管理する 4 つのアプローチ」の「Conda 環境を Studio の EFS ボリュームに永続化する」セクションを参照してください。

結論

このホワイトペーパーでは、運用モデル、ドメイン管理、ID 管理、アクセス許可管理、ネットワーク管理、ログ記録、モニタリング、カスタマイズなどの分野にわたるいくつかのベストプラクティスを紹介し、プラットフォーム管理者が SageMaker Studio プラットフォームを設定および管理できるようにします。

付録

マルチテナンシー比較

表2 — マルチテナンシー比較

マルチドメイン	マルチアカウント	単一のドメイン内の属性ベースのアクセス制御 (ABAC)
リソースの分離は、タグを使用して行われます。SageMaker Studio は、すべてのリソースにドメイン ARN およびユーザープロファイル/スペース ARN を自動的にタグ付けします。	各テナントはそれぞれのアカウントにあるため、リソースは完全に分離されます。	リソースの分離は、タグを使用して行われます。ユーザーは、ABAC 用に作成されたリソースのタグ付けを管理する必要があります。
List API をタグで制限することはできません。リソースの UI フィルタリングは共有スペースで行われますが、AWS CLI または Boto3 SDK を介して行われた List API コールでは、リージョン全体のリソースが一覧表示されます。	テナントはそれぞれの専用アカウントにあるため、List API の分離も可能です。	List API をタグで制限することはできません。AWS CLI または Boto3 SDK を介して行われた List API コールでは、リージョン全体のリソースが一覧表示されます。
テナントごとの SageMaker Studio のコンピューティングコストとストレージコストは、ドメイン ARN を使用することで、コスト配分タグとして簡単にモニタリングできます。	テナントごとの SageMaker Studio のコンピューティングコストとストレージコストは、専用アカウントで簡単にモニタリングできます。	テナントあたりの SageMaker Studio コンピューティングコストは、カスタムタグを使用して計算する必要があります。 すべてのテナントが同じ EFS ボリュームを共有するため、SageMaker Studio のストレージコストをドメインごと

マルチドメイン	マルチアカウント	単一のドメイン内の属性ベースのアクセス制御 (ABAC)
		にモニタリングすることはできません。
Service Quotas はアカウントレベルで設定されるため、1 つのテナントですべてのリソースを使用することができます。	Service Quotas は、テナントごとにアカウントレベルで設定できます。	Service Quotas はアカウントレベルで設定されるため、1 つのテナントですべてのリソースを使用することができます。
Infrastructure as Code (IaC) または Service Catalog を使用して、複数のテナントへのスケーリングを実現できます。	複数のテナントへのスケーリングには、Organizations と複数のアカウントの供給が含まれます。	スケーリングでは新しいテナントごとにテナント固有のロールが必要で、ユーザープロファイルではテナント名を手動でタグ付けする必要があります。
テナント内のユーザー間のコラボレーションは、共有スペースを通じて行えます。	テナント内のユーザー間のコラボレーションは、共有スペースを通じて行えます。	すべてのテナントは、コラボレーションのために同じ共有スペースにアクセスできます。

SageMaker Studio ドメインのバックアップとリカバリ

EFS を誤って削除した場合や、ネットワークや認証の変更によりドメインを再作成する必要がある場合は、以下の手順に従ってください。

オプション 1: EC2 を使用して既存の EFS からバックアップする

SageMaker Studio ドメインのバックアップ

1. SageMaker Studio のユーザープロファイルとスペースを一覧表示します (CLI、SDK)。
2. ユーザープロファイル/スペースを EFS の UID にマップします。
   a. ユーザー/スペースのリスト内のユーザーごとに、ユーザープロファイル/スペースを記述します (CLI、SDK)。

   b. ユーザープロファイル/スペースを HomeEfsFileSystemUid にマップします。
   c. ユーザー別に実行ロールが異なる場合は、ユーザープロファイルを UserSettings['ExecutionRole'] にマップします。
   d. スペースのデフォルトの実行ロールを特定します。
3. 新しいドメインを作成し、スペースのデフォルトの実行ロールを指定します。
4. ユーザープロファイルとスペースを作成します。
   - ユーザーリスト内のユーザーごとに、実行ロールのマッピングを使用してユーザープロファイルを作成します (CLI、SDK)。
5. 新しい EFS と UID のマッピングを作成します。
   a. ユーザーリスト内のユーザーごとに、ユーザープロファイルを記述します (CLI、SDK)。
   b. ユーザープロファイルを HomeEfsFileSystemUid にマップします。
6. 必要に応じて、すべてのアプリ、ユーザープロファイル、スペースを削除し、その後にドメインを削除します。

EFS のバックアップ

EFS をバックアップするには、次の手順に従います。

1. EC2 インスタンスを起動し、古い SageMaker Studio ドメインのインバウンド/アウトバウンドセキュリティグループを、新しい EC2 インスタンスにアタッチします (ポート 2049 で TCP 経由の NFS トラフィックを許可します)。「VPC 内の SageMaker Studio ノートブックを外部リソースに接続する」を参照してください。
2. SageMaker Studio の EFS ボリュームを新しい EC2 インスタンスにマウントします。「EFS ファイルシステムをマウントする」を参照してください。
3. ファイルを EBS ローカルストレージにコピーします (>sudo cp -rp /efs /studio-backup:)。
   a. 新しいドメインセキュリティグループを EC2 インスタンスにアタッチします。
   b. 新しい EFS ボリュームを EC2 インスタンスにマウントします。
   c. ファイルを新しい EFS ボリュームにコピーします。
   d. ユーザーのコレクション内のユーザーごとに以下の操作を行います。
      i. ディレクトリを作成します (mkdir new_uid)。
      ii. 古い UID ディレクトリから新しい UID ディレクトリにファイルをコピーします。
      iii. すべてのファイルの所有権を変更します (chown <new_UID>)。

オプション 2: S3 とライフサイクル設定を使用して既存の EFS からバックアップする

1. 「Amazon Linux 2 を使用して Amazon SageMaker ノートブックインスタンスに作業内容を移行する」を参照します。
2. バックアップ用の S3 バケット (`>studio-backup` など) を作成します。
3. 実行ロールを持つすべてのユーザープロファイルを一覧表示します。
4. 現在の SageMaker Studio ドメインで、デフォルトの LCC スクリプトをドメインレベルで設定します。
   - LCC で、`/home/sagemaker-user` のすべてを S3 のユーザープロファイルプレフィックス (`s3://studio-backup/studio-user1` など) にコピーします。
5. デフォルトの Jupyter Server アプリをすべて再起動します (LCC を実行するため)。
6. アプリ、ユーザープロファイル、ドメインをすべて削除します。
7. 新しい SageMaker Studio ドメインを作成します。
8. ユーザープロファイルと実行ロールのリストから新しいユーザープロファイルを作成します。
9. LCC をドメインレベルで設定します。
   - LCC で、S3 のユーザープロファイルプレフィックスのすべてを `/home/sagemaker-user` にコピーします。
10. LCC 設定 を使用してすべてのユーザー向けのデフォルトの Jupyter Server アプリを作成します (CLI、SDK)。

SAML アサーションを使用した SageMaker Studio へのアクセス

ソリューションのセットアップ:

1. 外部 IdP で SAML アプリケーションを作成します。
2. 外部 IdP を IAM の ID プロバイダーとして設定します。
3. IdP からアクセスできる `SAMLValidator` Lambda 関数を (関数 URL または API ゲートウェイを通じて) 作成します。
4. `GeneratePresignedUrl` Lambda 関数と、この関数にアクセスするための API ゲートウェイを作成します。
5. API ゲートウェイを呼び出すためにユーザーが引き受けることができる IAM ロールを作成します。このロールは、次の形式を使用して SAML アサーションに属性として渡す必要があります。

- 属性名: https://aws.amazon.com/SAML/Attributes/Role
- 属性値: <IdentityProviderARN>、<RoleARN>

6. SAML Assertion Consumer Service (ACS) エンドポイントを SAMLValidator 呼び出しの URL に更新します。

SAML バリデーターのサンプルコード:

```
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
```

詳細情報

- AWS で安全かつ管理が行き届いた機械学習環境を設定する (AWS ブログ)
- リソースを完全に分離するようにチームやグループ向けの Amazon SageMaker Studio を設定する (AWS ブログ)
- AWS SSO と Okta Universal Directory を使用した Amazon SageMaker Studio のオンボーディング (AWS ブログ)
- SAML 2.0 を AWS アカウントフェデレーション用に設定する方法 (Okta ドキュメント)
- AWS で安全なエンタープライズ機械学習プラットフォームを構築する (AWS テクニカルガイド)
- ライフサイクル設定を使用して Amazon SageMaker Studio をカスタマイズする (AWS ブログ)
- Amazon SageMaker Studio ノートブックで独自のカスタムコンテナイメージを使用する (AWS ブログ)
- カスタム SageMaker プロジェクトテンプレートの作成 — ベストプラクティス (AWS ブログ)
- Amazon SageMaker Pipelines を使用したマルチアカウントモデルのデプロイ (AWS ブログ)
- パート 1: NatWest Group がスケーラブルで安全かつ持続可能な MLOps プラットフォームを構築した方法 (AWS ブログ)
- Amazon SageMaker Studio の署名付き URL のセキュリティ保護パート 1: 基盤インフラストラクチャ (AWS ブログ)

寄稿者

このドキュメントの寄稿者は次のとおりです。

- Ram Vittal、Amazon Web Services、ML ソリューションアーキテクト
- Sean Morgan、Amazon Web Services、ML ソリューションアーキテクト
- Durga Sury、Amazon Web Services、ML ソリューションアーキテクト

アイデア、改訂、展望を提供してくれた以下の人々に特に感謝します。

- Alessandro Cerè、Amazon Web Services、AI/ML ソリューションアーキテクト
- Sumit Thakur、Amazon Web Services、SageMaker プロダクトリーダー
- Han Zhang、Amazon Web Services、シニアソフトウェア開発エンジニア
- Bhadrinath Pani、Amazon Web Services、ソフトウェア開発エンジニア

ドキュメントの改訂

このホワイトペーパーの更新に関する通知を受け取るには、RSS フィードにサブスクライブしてください。

変更	説明	日付
ホワイトペーパーの更新	壊れたリンクを修正し、全体を通して多くの編集上の変更を行いました。	2023 年 4 月 25 日
初版発行	ホワイトペーパーの発行。	2022 年 10 月 19 日

注意

お客様は、本書に記載されている情報を独自に評価する責任を負うものとします。本書は、(a) 情報提供のみを目的とし、(b) AWS の現行製品と慣行について説明しており、これらは予告なしに変更されることがあり、(c) AWS およびその関連会社、サプライヤー、またはライセンサーからの契約上の義務や保証をもたらすものではありません。AWS の製品やサービスは、明示または黙示を問わず、一切の保証、表明、条件なしに「現状のまま」提供されます。お客様に対する AWS の責任は AWS 契約によって規定されます。本書は、AWS とお客様との間で締結されるいかなる契約の一部でもなく、その内容を修正するものでもありません。

AWS 用語集

AWS の最新の用語については、「AWS の用語集リファレンス」の「AWS 用語集」を参照してください。