AWS 白皮书

SageMaker 工作室管理最佳实践

SageMaker 工作室管理最佳实践: AWS 白皮书

Table of Contents

SageMaker 工作室管理最佳实践

发布日期：2023 年 4 月 25 日 (文档修订)

摘要

Amazon SageMaker AI Studio 提供了一个基于 Web 的可视化界面，您可以在其中执行所有机器学习 (ML) 开发步骤，从而提高数据科学团队的工作效率。 SageMaker AI Studio 让您可以完全访问、控制和了解构建、训练和评估模型所需的每个步骤。

本白皮书讨论了运营模式、域管理、身份管理、权限管理、网络管理、日志记录、监控和自定义等主题的最佳实践。此处讨论的最佳实践适用于企业 SageMaker AI Studio 部署，包括多租户部署。本文档适用于机器学习平台管理员、机器学习工程师和机器学习架构师。

您使用 Well-Architected 了吗？

当您在云端构建系统时，AWS Well-Architected Framework 可帮助您了解所做决策的利弊。利用此框架的六个支柱，您可以了解到设计和运行可靠、安全、高效、经济有效且可持续的系统的架构最佳实践。您可以使用AWS Management Console免费提供的 AWS Well-Architected Tool，回答与每个支柱相关的一组问题，即可根据这些最佳实践检查自己的工作负载。

在 Machine Learning Lens 中，我们重点介绍了如何在 AWS 云中设计、部署和构建机器学习工作负载。此剖析对 Well-Architected Framework 所述最佳实践进行补充说明。

简介

在将 SageMaker AI Studio 作为机器学习平台进行管理时，您需要最佳实践指导来做出明智的决策，以帮助您在工作负载增长时扩展机器学习平台。如需预置、操作并扩展机器学习平台，请考虑以下事项：

- 选择正确的运营模式并规划机器学习环境，以实现业务目标。
- 选择如何为用户身份设置 SageMaker AI Studio 域身份验证，并考虑域级别限制。
- 确定将用户身份与授权联合到机器学习平台的方法，以实现精细访问控制和审计。
- 考虑为机器学习角色的不同身份设置权限和防护机制。
- 根据机器学习工作负载的敏感度、用户数量、实例类型、应用程序和已启动的作业，规划您的虚拟私有云 (VPC) 网络拓扑。
- 使用加密手段对静态数据和传输中数据进行分类和保护。

- 考虑如何记录和监控各种应用程序编程接口 (APIs) 和用户活动以确保合规性。
- 使用您自己的图像和生命周期配置脚本自定义 SageMaker AI Studio 笔记本体验。

运营模式

运营模式是一种融合了人员、流程和技术的框架，有助于组织以可扩展、一致、高效的方式实现业务价值。机器学习运营模式为组织内的各团队提供了标准的产品开发流程。根据规模、复杂性和业务驱动因素，有三种实施运营模式的模型：

- 集中式数据科学团队 — 在此模型中，所有数据科学活动都集中发生在单个团队或组织中。这类似于卓越中心 (COE) 模型，即所有业务部门都交给该团队进行数据科学项目。
- 分布式数据科学团队 — 在此模型中，数据科学活动分布在不同的业务职能或部门，或者基于不同的产品线。
- 联合数据科学团队 — 在此模型中，集中式团队负责管理代码存储库、持续集成和持续交付 (CI/CD) 管道等共享服务功能，而分布式团队负责管理各业务部门或产品级功能。这类似于星型拓扑连接模型，即每个业务部门都有专门的数据科学团队，但这些团队会与集中式团队协调活动。

在决定为制作用例启动您的第一个工作室域名之前，请考虑您的运营模式和组织环境 AWS 的最佳实践。有关更多信息，请参阅使用多个帐户组织您的 AWS 环境。

下一节将指导如何针对各种运营模式来组织账户结构。

推荐的账户结构

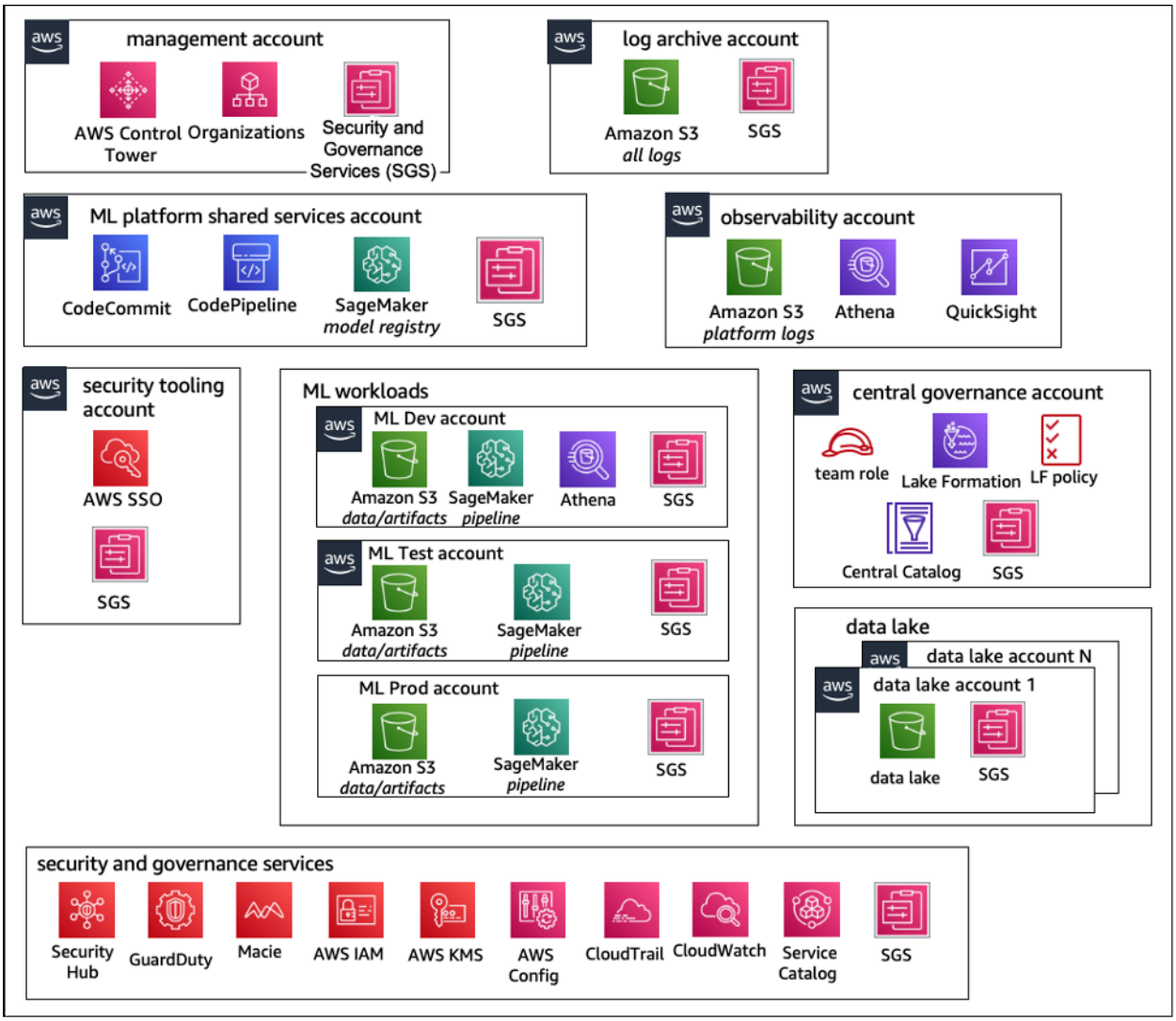
本节简要介绍了一种运营模式账户结构，以便您根据组织的运营要求初步应用并进行修改。无论您选择哪种运营模式，亚马逊都建议您实施以下常见的最佳实践：

- 使用 AWS Control Tower 设置、管理并监管账户。
- 使用您的身份提供商 (IdP) 集中您的身份，AWS IAM使用委派的管理员 Securitiy Tooling 帐户集中身份中心，并启用对工作负载的安全访问。
- 使用跨开发、测试和生产工作负载的账户级隔离，运行机器学习工作负载。
- 将机器学习工作负载日志流式传输到日志存档账户，然后在可观测性账户中筛选并应用日志分析。
- 运行集中式监管账户，用于预置、控制并审核数据访问权限。
- 根据您的组织和工作负载要求，在每个账户中嵌入带有适当预防和侦查防护栏的安全和治理服务 (SGS)，以确保安全性和合规性。

集中式模型账户结构

在此模型中，机器学习平台团队负责提供：

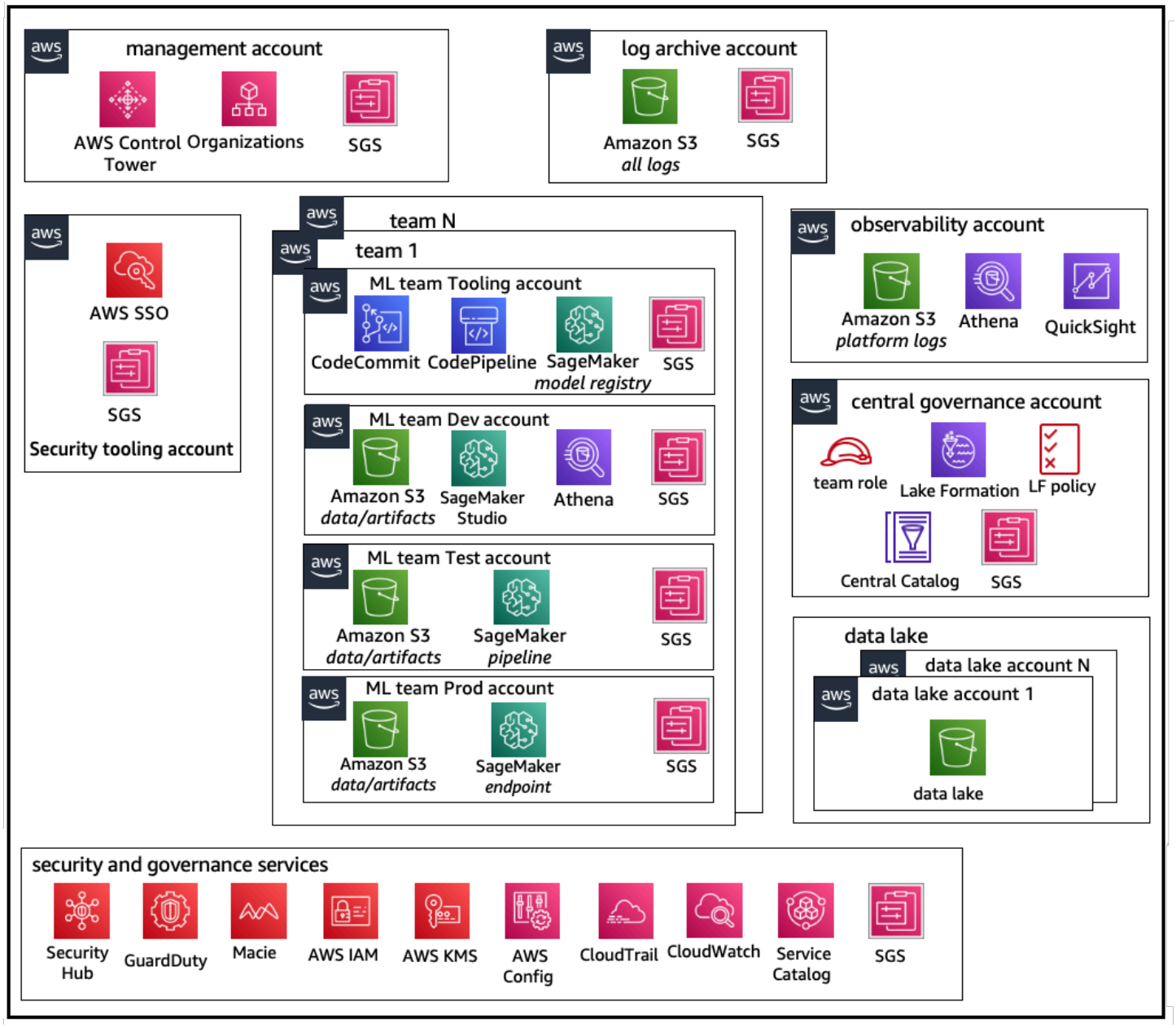
- 一个共享服务工具账户，用于满足数据科学团队的 Machine Learning Operations (MLOps) 要求。
- 跨数据科学团队共享账户，可开发、测试并生产机器学习工作负载。
- 监管策略，可确保独立运行各数据科学团队的工作负载。
- 常见的最佳实践。



集中式运营模式账户结构

分布式模型账户结构

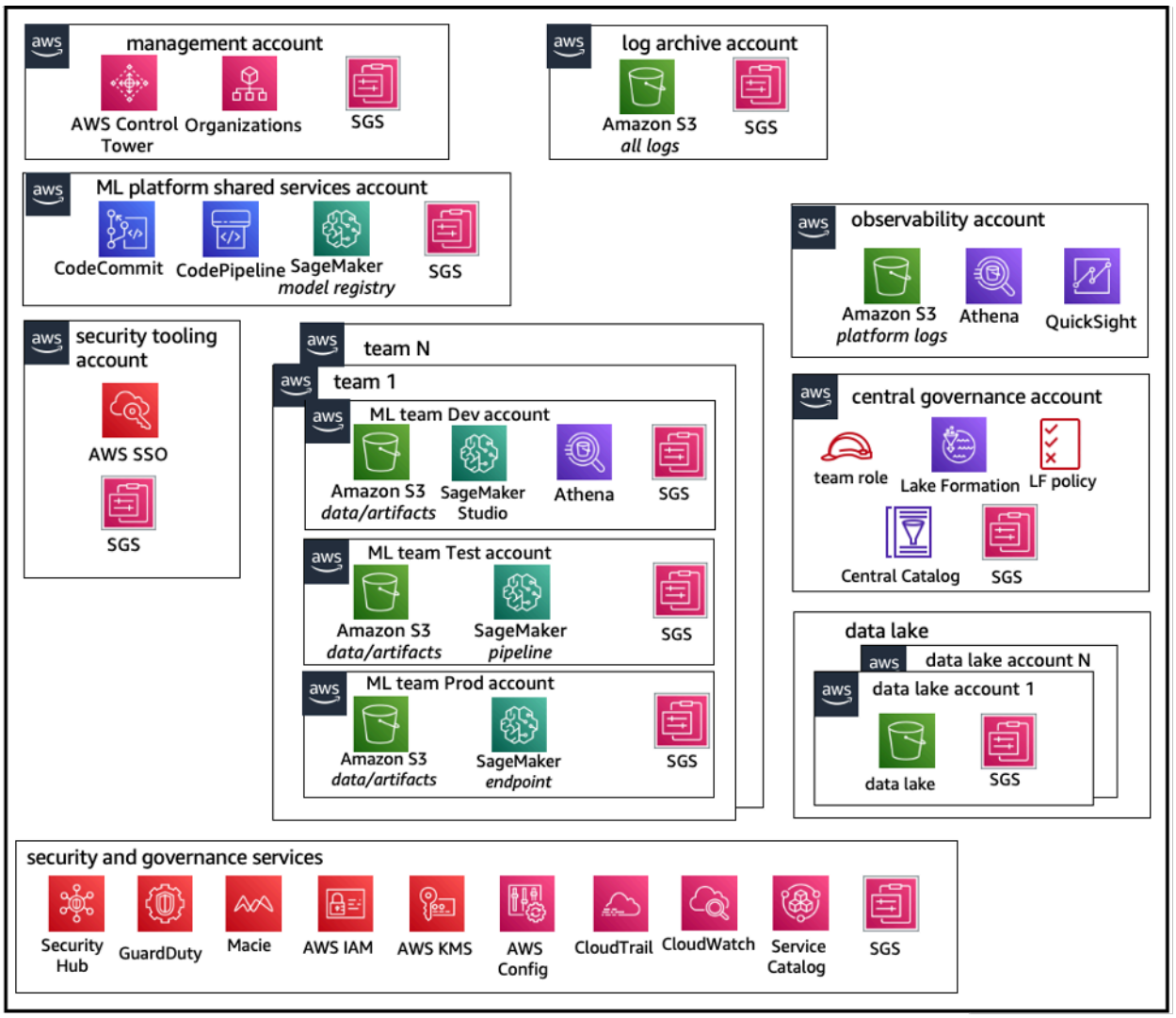
在此模型中，每个机器学习团队均独自负责预置、管理并治理机器学习账户和资源。亚马逊建议机器学习团队使用支持可观测性和数据治理的集中式模型，以简化数据治理和审计管理流程。



分布式运营模式账户结构

联合模型账户结构

该模型与集中式模型类似；但是，主要区别在于，每个数据science/ML team gets their own set of development/test/production工作负载都允许对其机器学习资源进行强大的物理隔离，并且还使每个团队能够在不影响其他团队的情况下独立扩展。



联合运营模式账户结构

机器学习平台多租户架构

多租户是一种软件架构，其中的单个软件实例可以为多个不同的用户组提供服务。租户是一组用户，共享对软件实例的特定访问权限。例如，您在开发多个机器学习产品时，可以将具有相似访问权限要求的产品团队都视为租户或团队。

虽然可以在一个 SageMaker AI Studio 实例（例如 AI Dom SageMaker ain）中部署多个团队，但当您将多个团队引入单个 A SageMaker I Studio 域时，请权衡这些优势和爆炸半径、成本归因和账户等级限制等权衡。以下章节详细说明了这些利弊和最佳实践。

如果您需要绝对的资源隔离，可以考虑为不同账户中的每个租户实现 SageMaker AI Studio 域。根据您的隔离要求，您可以在单个账户和区域内将多个业务线 (LOBs) 作为多个域名实施。使用共享空间在同一团队的成员之间进行近乎实时的协作/LOB. 对于多个域，您仍将使用身份访问管理 (IAM) 策略和权限来确保资源隔离。

SageMaker 从域创建的 AI 资源会自动使用域 Amazon 资源名称 (ARN) 和用户配置文件或空间ARN进行标记，以便于资源隔离。有关示例策略，请参阅域资源隔离文档。在这里，您可以看到有关何时使用多账户或多域策略的详细参考以及文档中的功能比较，还可以查看用于回填存储库中现有域名标签的示例脚本。GitHub

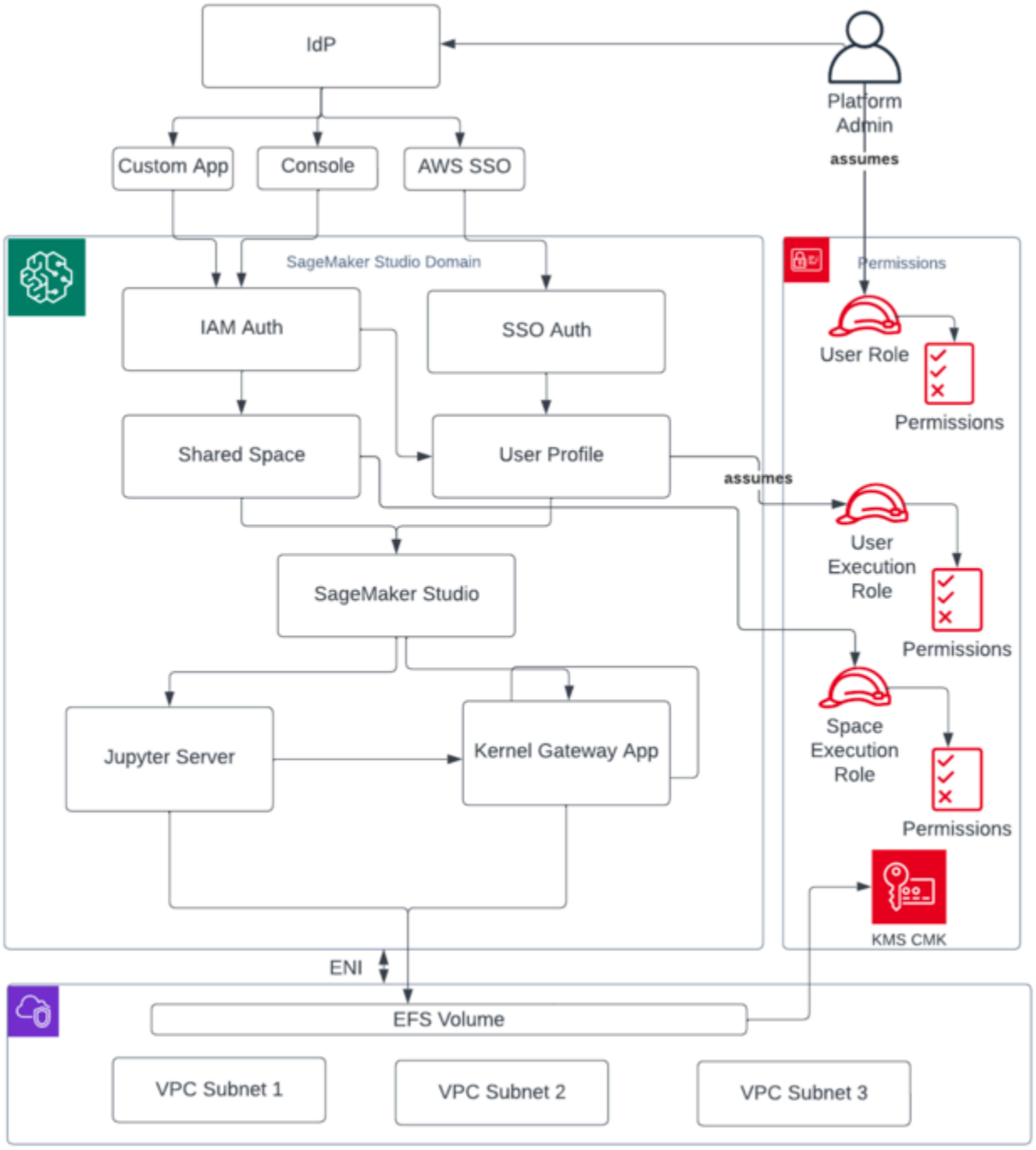
最后，您可以使用将 SageMaker AI Studio 资源自助部署到多个账户AWS Service Catalog。有关更多信息，请参阅管理多个 AWS 账户 和中的 AWS Service Catalog 商品 AWS 区域。

域管理

亚马逊 SageMaker AI 域名包括：

- 关联的亚马逊 Elastic File Syst em（亚马逊EFS）卷
- 授权用户列表
- 各种安全、应用程序、策略和亚马逊虚拟私有云 (AmazonVPC) 配置

下图提供了构成 SageMaker AIStudio域的各种组件的高级视图：

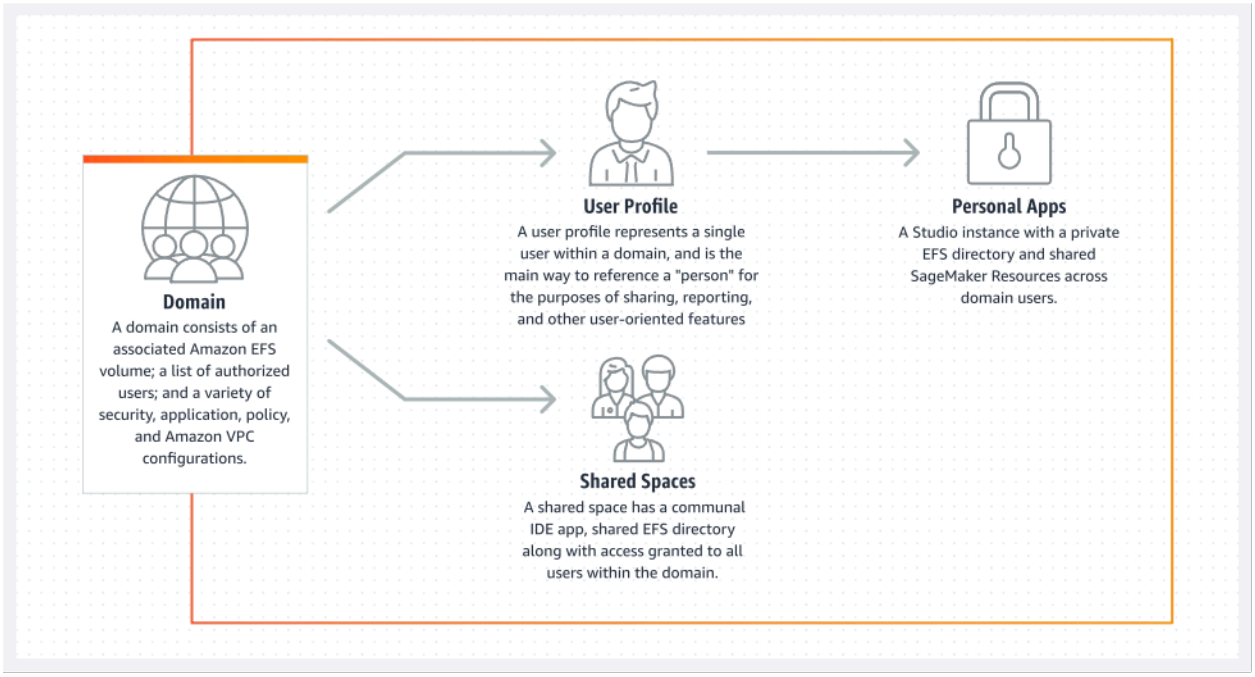


构成 SageMaker AI Studio 域的各种组件的高级视图

多域和共享空间

Amazon SageMaker AI 现在支持在单个账户中创建多个 AWS 区域 A SageMaker I 域。每个域可以有自己的域设置（例如身份验证模式）和网络设置（例如VPC和子网）。用户配置文件无法跨域共享。如果用户加入了按域划分的多个团队，则每个域都要为其创建用户配置文件。如需了解为现有域回填标签的方法，请参阅多域概述。

在IAM身份验证模式下设置的每个域都可以利用共享空间在用户之间进行近乎实时的协作。通过共享空间，用户可以访问共享的 Amazon EFS 目录和用户界面的共享JupyterServer应用程序，并且可以近乎实时地共同编辑。管理员可使用共享空间创建的资源自动标记功能跟踪项目成本。共享 JupyterServer 用户界面还会筛选实验和模型注册表项等资源，以便仅显示与共享机器学习工作相关的项目。下图概述了每个域中的私有应用程序和共享空间。



单个域中的私有应用程序和共享空间概述

在您的域中设置共享空间

共享空间通常是为特定的机器学习工作或项目创建的，其中单个域的成员需要近乎实时地访问相同的底层文件存储和IDE. 用户可以近实时地访问、读取、编辑并共享其笔记本，能够以最快的速度开始与同行迭代。

如需创建共享空间，必须先指定空间默认执行角色，负责管理使用该空间的任何用户的权限。在编写时，域内所有用户均可访问自己域内的所有共享空间。有关向现有域添加共享空间的最新文档，请参阅创建共享空间。

为IAM联盟设置您的域名

在为 SageMaker AI Studio 域设置 AWS Identity and Access Management (IAM) IAM 联合之前，您需要在 IdP 中设置联合用户角色（例如平台管理员），如身份管理部分所述。

有关使用该IAM选项设置 SageMaker AI Studio 的详细说明，请参阅使用IAM身份中心加入亚马逊 SageMaker 域名。

为单点登录 (SSO) 联合设置您的域

要使用单点登录 (SSO) 联合，您需要 AWS IAM Identity Center 在需要运行 SageMaker AI Studio 的同一区域的AWS Organizations管理账户中启用。域设置步骤与IAM联合身份验证步骤类似，不同之处在于您在 “身份验证” 部分中选择 AWS IAM Identity Center(iDc)。

有关详细说明，请参阅使用IAM身份中心登录 Amazon SageMaker 域名。

SageMaker AI Studio 用户资料

用户配置文件代表域中的单个用户，也是为了使用共享、报告和其他面向用户的功能而引用“人员”的主要方式。该实体是在用户加入 toSageMaker AI Studio 时创建的。如果管理员通过电子邮件邀请用户或使用 IdC 导入文件，则用户配置文件会自动创建。用户个人资料是个人用户设置的主要持有者，它引用了用户的私有 Amazon Elastic File Syst em (AmazonEFS) 主目录。我们建议为 SageMaker AI Studio 应用程序的每个物理用户创建用户配置文件。每个用户在 Amazon 上都有自己的专用目录 EFS，并且不能在同一个账户中跨域共享用户资料。

每个共享 SageMaker AI Studio 域的用户个人资料都会获得用于运行笔记本的专用计算资源（例如 SageMaker AI A mazon Elastic Compute Cloud (AmazonEC2) 实例）。分配给一号用户和二号用户的计算实例是完全隔离的。同样，分配给 AWS 账户的计算资源与其他账户中用户的计算资源也是完全分隔的。每位用户可以在隔离的 Docker 容器中最多运行四款应用程序（应用），也可以在相同的实例类型上运行映像。

Jupyter 服务器应用程序

当您通过访问预签名URL或使用 iD AWS IAM C 登录为用户启动 SageMaker Amazon AI Studio 笔记本时，Jupyter Server 应用程序将在人工智能服务托管实例中 SageMaker 启动。VPC每位用户都能在私有应用程序中获得专用的 Jupyter 服务器应用程序。默认情况下，适用于 SageMaker AI Studio 笔记本的 Jupyter Server 应用程序在专用`ml.t3.medium`实例（预留为系统实例类型）上运行。客户无需支付实例计算费用。

Jupyter 内核网关应用程序

Kernel Gateway 应用程序可以通过API或 SageMaker AI Studio 接口创建，并在选定的实例类型上运行。此应用程序可以使用预先配置了流行数据科学的内置 SageMaker AI Studio 映像以及深度学习包（例如 Apache MXNet 和）来运行。TensorFlowPyTorch

用户可以在同一 Studio 中启动和运行多个 Jupyter 笔记本内核、终端会话和交互式控制台。SageMaker image/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

您需要使用不同的实例类型，才能创建其他应用程序。一个用户配置文件只能运行一个实例，类型不限。例如，用户可以在同一个实例上运行使用 SageMaker AI Studio 内置数据科学图像的简单笔记本电脑，也可以使用内置 TensorFlow 图像运行另一台笔记本电脑。用户需要付费运行实例。为了避免在用户未主动运行 SageMaker AI Studio 时产生成本，用户需要关闭实例。有关更多信息，请参阅关闭和更新 Studio 应用程序。

每次从 SageMaker AI Studio 界面关闭并重新打开 Kernel Gateway 应用程序时，该应用程序都会在新实例上启动。这意味着重启同一应用程序时需要重新安装软件包。同样，如果用户更改笔记本上的实例类型，则已安装的软件包和会话变量都会丢失。但是，您可以使用诸如自带图像和生命周期脚本之类的功能，将用户自己的软件包引入 SageMaker AI Studio，并通过实例切换和新实例启动来保留它们。

Amazon Elastic File System 卷

创建域时，将创建一个 Amazon Elastic File System (AmazonEFS) 卷，供该域内的所有用户使用。每个用户个人资料都会在 Amazon EFS 卷中收到一个私有主目录，用于存储用户的笔记本、 GitHub 存储库和数据文件。域中的每个空间都会在 Amazon EFS 卷中接收一个私有目录，多个用户配置文件可以访问该目录。对文件夹的访问由用户通过文件系统权限进行隔离。 SageMaker AI Studio 为每个用户配置文件或空间创建一个全局唯一的用户 ID，并将其作为便携式操作系统接口 (POSIX) 应用user/group ID for the user's home directory on EFS, which prevents other users/spaces于访问其数据。

备份和恢复

现有EFS卷无法连接到新的 SageMaker AI 域。在生产环境中，请确保已将 Amazon EFS 卷备份（备份到另一个EFS卷或亚马逊简单存储服务 (Amazon S3)）。如果EFS卷被意外删除，管理员必须拆除并重新创建 SageMaker AI Studio 域。流程如下：

通过、、和DescribeSpaceAPI呼叫备份用户配置文件、空间和关联EFS用户 IDs (UIDs) 的列表。ListUserProfiles DescribeUserProfile List Spaces

1. 创建新的 A SageMaker I Studio 域名。
2. 创建用户配置文件和空间。
3. 对于每个用户个人资料，请复制 EFS /Amazon S3 上备份的文件。
4. （可选）删除旧 A SageMaker I Studio 域中的所有应用程序和用户配置文件。

有关详细说明，请参阅附录部分 SageMaker AI Studio 域备份和恢复。

> **Note**
>
> 还可以使用 `LifecycleConfigurations`，在用户每次启动应用程序时，在 S3 之间来回备份数据。

亚马逊EBS交易量

每个 AI Studio Notebook 实例上还附有一个亚马逊 Elastic Bloc k Store（SageMaker AmazonEBS）存储卷。它将作为运行在实例上的容器或映像的根卷。虽然亚马逊EFS存储空间是永久性的，但附加到容器的亚马逊EBS卷是临时的。如果客户删除应用程序，本地存储在 Amazon EBS 卷上的数据将无法保存。

保护对预签名的访问权限 URL

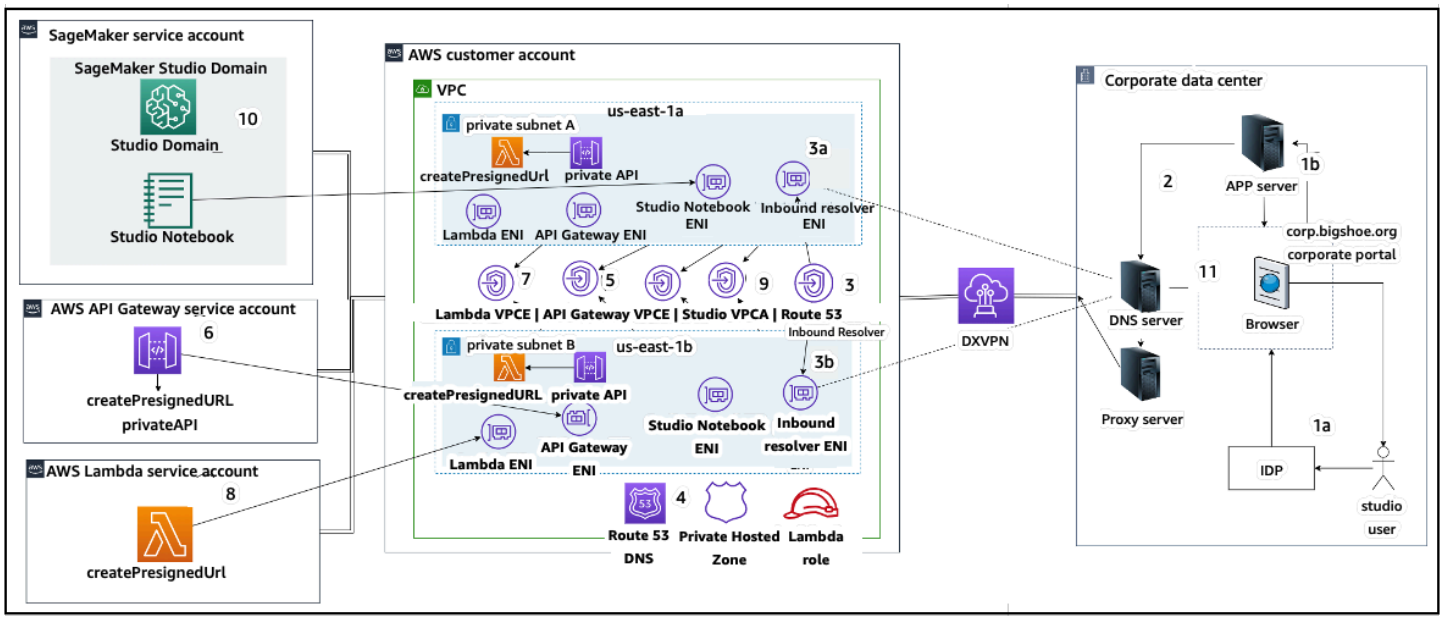
当 A SageMaker I Studio 用户打开笔记本链接时， SageMaker AI Studio 会验证联合用户授权访问的IAM策略，并生成并解析该用户的预签名URL策略。由于 SageMaker AI 控制台在 Internet 域上运行，因此在浏览器会话中可以看到生成的预签名URL。这可能会导致发生数据失窃，以及在未采取适当措施的情况下被人获取客户数据。

Studio 支持几种针对预签名URL数据盗窃实施访问控制的方法：

- 使用IAM策略条件验证客户端 IP `aws:sourceIp`
- 使用IAM条件VPC验证客户端 `aws:sourceVpc`
- 使用IAM策略条件验证客户端VPC端点 `aws:sourceVpce`

从 SageMaker AI 控制台访问 AI Studio 笔记本电脑时，唯一可用的选项是将客户端 IP 验证与IAM策略条件一起使用`aws:sourceIp`。 SageMaker 您也可以使用 Zscaler 等浏览器流量路由产品，确保员工访问互联网的规模与合规性。这些流量路由产品会生成专属源 IP，IP 范围不受企业客户控制。因此，这些企业客户无法使用 `aws:sourceIp` 条件。

要使用IAM策略条件进行客户端VPC端点验证`aws:sourceVpce`，URL需要在部署 AI Studio 的同一客户VPC中创建预签名，并且URL需要通过客户上的 SageMaker AI Studio VPC 端点解决预签名问题。SageMaker VPC使用DNS转发规则（URL在 Zscaler 和企业版中DNS），然后使用 A mazon Route 53 入站解析器进入客户VPC终端节点，如以下架构所示，可以解析企业网络用户在访问期间的预签名：



通过企业网络访问使用VPC端点预先URL签名的 Studio

有关设置上述架构的 step-by-step指南，请参阅预先签名的安全 Amazon A SageMaker I Studio URLs 第 1 部分：基础基础设施。

SageMaker AI 域配额和限制

- SageMaker AI Studio 域名SSO联合仅在配置 AWS 身份中心的 AWS 组织的成员账户中支持该区域。

- 使用 Ident AWS ity Center 设置的域目前不支持共享空间。
- VPC并且在创建域后无法更改子网配置。但是，您可以使用不同的VPC子网配置创建新域。
- 创建域后，无法在IAM和SSO模式之间切换域访问权限。您可以使用不同的身份验证模式创建新域。
- 每位用户使用每种实例类型时，最多只能启动四个内核网关应用程序。
- 每个用户只能启动每个实例类型的一个实例。
- 域内消耗的资源也会受限，如按实例类型启动的实例数量，以及可创建的用户配置文件数量。有关服务限制的完整列表，请参阅服务配额页面。
- 客户可以提交企业支持案例并说明商业理由，以便根据账户级防护机制放宽默认资源限制，例如增加域数量或用户配置文件数量。
- 每个账户的并发应用程序数量的硬限制为 2500 个。该硬限制决定了域和用户配置文件的数量限制。例如，账户可以有单个域，域中包含 1000 个用户配置文件，也可以有 20 个域，每个域中包含 50 个用户配置文件。

身份管理

本节讨论公司目录中的员工用户如何联合进入 AI Studio AWS 账户 并访问 SageMaker AI Studio。首先，亚马逊会简要说明用户、组和角色的映射方法，以及用户联合身份验证的工作原理。

用户、组和角色

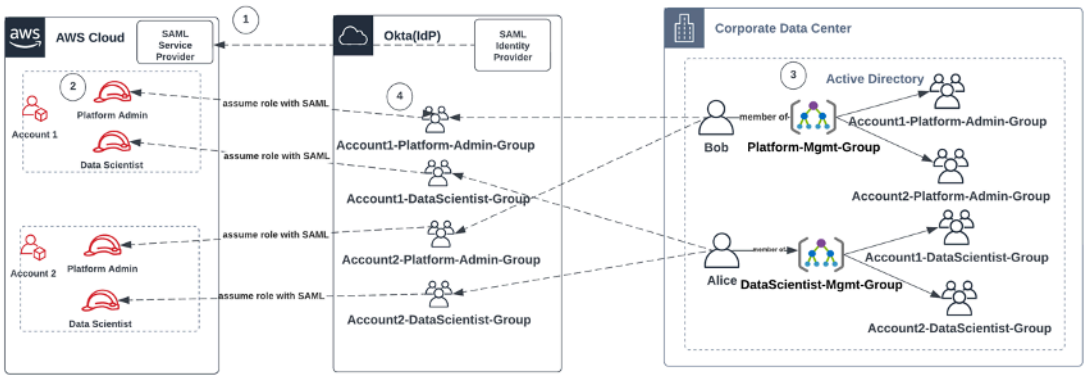
在中 AWS，使用用户、群组和角色管理资源权限。客户可以通过或在公司目录（例如Active Directory (AD)）中管理其用户和群组，该目录通过外部 IdP（例如 Okta）启用，允许他们对云端和本地运行的各种应用程序进行用户身份验证。IAM

如 AWS 安全支柱身份管理部分所述，在中央 IdP 中管理用户身份是一种最佳实践，因为这有助于轻松地与后端人力资源流程集成，并有助于管理员工用户的访问权限。

IdPs 例如 Okta 允许最终用户使用SSO安全断言标记语言 () 对一个或多个角色进行身份验证AWS 账户并获得对特定角色的访问权限。SAMLIdP 管理员可以将角色从 IdP 下载到 AWS 账户 IdP 中，然后将这些角色分配给用户。登录时 AWS，最终用户会看到一个 AWS 屏幕，其中显示了一个或多个分配给他们的 AWS 角色列表 AWS 账户。用户可选择登录时要代入的角色，该角色定义用户在身份验证会话期间可享有的权限。

针对您想要提供访问权限的特定账户和角色组合，IdP 必须一一建立对应组。这些组可视为 AWS 角色特定组。角色特定组内所有成员用户都将获得一项权限：可访问特定 AWS 账户中的特定角色。但是，这种单一的授权流程无法通过分派用户到特定 AWS 角色组来扩展用户访问权限的管理范围。为了简化管理，我们还建议您为组织中需要不同权限集的所有不同用户集创建多个群组 AWS 。

为了说明中央 IdP 设置，可考虑一家采用 AD 设置的企业，其用户和组均能同步到 IdP 目录。在中 AWS，这些 AD 组映射到IAM角色。工作流的主要步骤如下：



加入 AD 用户、AD 组和IAM角色的工作流程

1. 在中 AWS，为每个人设置 AWS 账户 与 IdP 的SAML集成。
2. 在中 AWS，在每个角色中设置角色 AWS 账户 并同步到 IdP。
3. 在企业 AD 系统中：
   a. 为每个账户角色创建一个 AD 组并同步到 IdP（例如，`Account1-Platform-Admin-Group`（又名 AWS 角色组））。
   b. 在每个角色级别创建管理组（例如`Platform-Mgmt-Group`），并将 AWS 角色组分配为成员。
   c. 将用户分配到该管理组以允许访问 AWS 账户 角色。
4. 在 IdP 中，将 AWS 角色组（例如`Account1-Platform-Admin-Group`）映射到 AWS 账户 角色（例如 Account1 中的平台管理员）。
5. 当数据科学家 Alice 登录 Idp 时，他们会看到一个 AWS 联邦应用程序用户界面，其中有两个选项可供选择：“账户 1 数据科学家” 和 “账户 2 数据科学家”。
6. Alice 选择 “账户 1 数据科学家” 选项，他们将连接到 AWS 账户 1（SageMaker AI 控制台）中的授权应用程序。

有关设置SAML账户联盟的详细说明，请参阅 Okta 的 “如何为 AWS 账户联合配置 SAML 2.0”。

用户联合身份验证

SageMaker AI Studio 的身份验证可以使用IAM或 i IAM DC 完成。如果用户是通过管理的IAM，则他们可以选择IAM模式。如果企业使用外部 IdP，则可以通过或 IdC 进行IAM联合。IAM请注意，现有 SageMaker AI Studio 域的身份验证模式无法更新，因此在创建生产 SageMaker AI Studio 域之前做出决定至关重要。

如果 SageMaker AI Studio 设置为IAM模式， SageMaker AI Studio 用户将通过预签名访问应用程序 URL，当用户通过浏览器访问时，该应用程序会自动登录到 SageMaker AI Studio 应用程序。

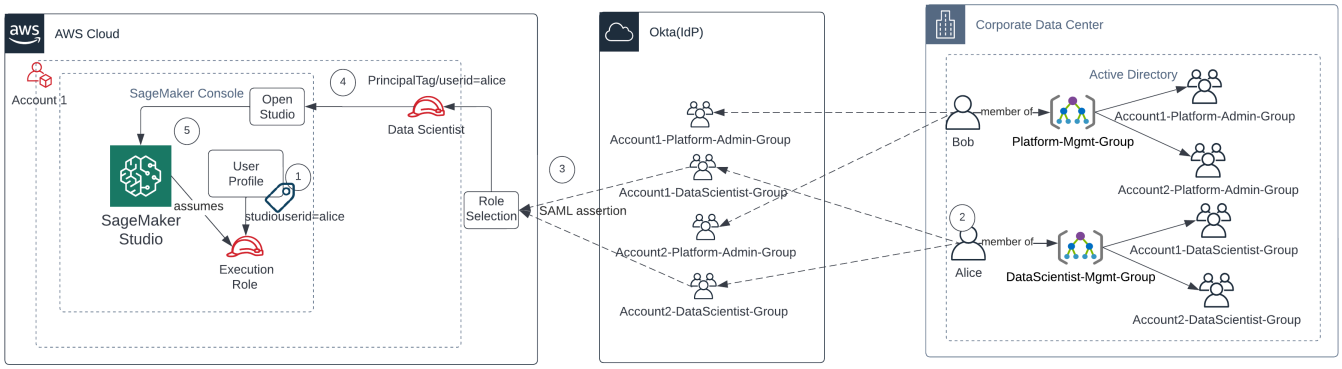
IAM 用户

对于IAM用户，管理员为每个用户创建 SageMaker AI Studio 用户配置文件，并将用户配置文件与允许用户在 Studio 中执行必要操作的IAM角色关联起来。要限制 AWS 用户仅访问其 SageMaker AI Studio 用户个人资料，管理员应为 SageMaker AI Studio 用户配置文件添加标签，并向该用户附加一个IAM策略，允许他们仅在标签值与 AWS 用户名相同时才允许他们进行访问。此策略语句与以下内容类似：

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

AWS IAM或账户联合

AWS 账户 联合方法使客户能够从其 Id SAML P（例如 Okta）联合到 SageMaker AI 控制台。要限制用户仅访问其用户个人资料，管理员应标记 SageMaker AI Studio 用户个人资料，添加 `PrincipalTags` IdP，然后将其设置为传递标签。下图描述了如何授权联合用户（数据科学家 Alice）访问自己的 SageMaker AI Studio 用户个人资料。



在IAM联合模式下访问 SageMaker AI Studio

1. Alice SageMaker AI Studio 用户配置文件标有他们的用户 ID，并与执行角色相关联。
2. Alice 向 IdP (Okta) 进行身份验证。

3. IdP 对 Alice 进行了身份验证，并发布了 Alice 所属的两个角色（账户 1 和账户 2 的数据科学家）的 SAML断言。Alice 选择数据科学家账户 1 角色。
4. Alice 以数据科学家的角色登录账户 1 SageMaker AI 控制台。Alice 在 Studio 应用程序实例列表中打开对应的应用程序实例。
5. 代入角色会话中的 Alice 主体标签已根据所选的 SageMaker AI Studio 应用程序实例用户配置文件标签进行验证。如果配置文件标签有效，则以执行角色启动 SageMaker AI Studio 应用程序实例。

如果您想在用户入职过程中自动创建 SageMaker AI Execution 角色和策略，以下是实现这一目标的一种方法：

1. 为每个账户和 Studio 域级别设置 AD 组，例如 `SageMaker AI-Account1-Group`。
2. 当你需要让用户加入 SageMaker AI Studio 时，将 ai-account1-group 添加到 SageMaker 用户的群组成员资格中。

设置一个自动化流程，监听`SageMaker AI-Account1-Group`成员资格事件，并使用 AWS APIs该流程根据其广告组成员资格创建角色、策略、标签和 SageMaker AI Studio 用户个人资料。附加角色到用户配置文件。有关策略示例，请参阅阻止 SageMaker AI Studio 用户访问其他用户个人资料。

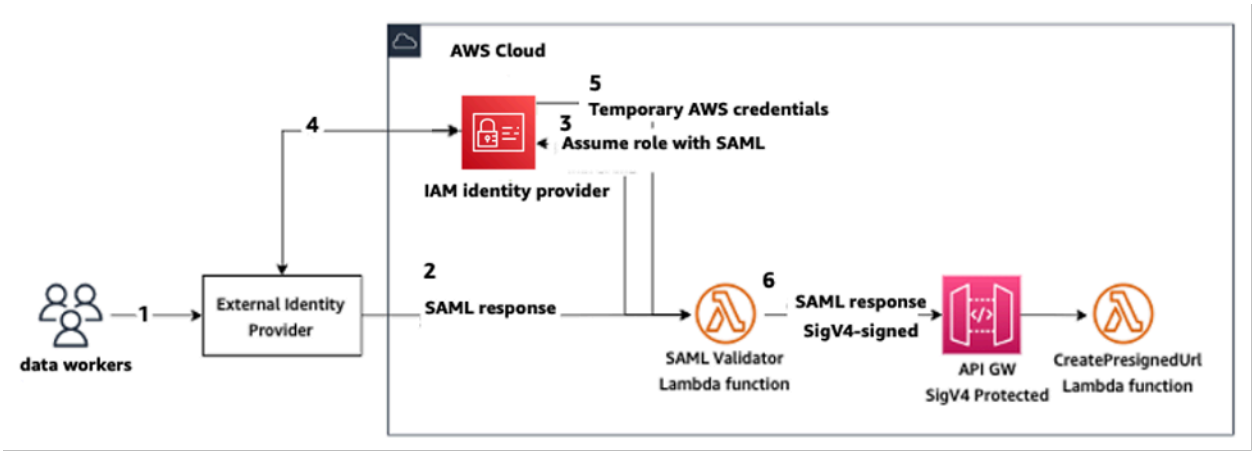
SAML使用身份验证 AWS Lambda

在IAM模式下，用户还可以使用SAML断言在 SageMaker AI Studio 中进行身份验证。在此架构中，客户拥有现有 IdP，他们可以在其中创建SAML应用程序供用户访问 Studio（而不是 AWS 身份联合应用程序）。客户的 IdP 已添加到。IAM AWS Lambda 函数使用IAM和帮助验证SAML断言STS，然后直接调用API网关或 Lambda 函数来创建预签名域。URL

该解决方案的优势在于，Lambda 函数可以自定义访问 SageMaker AI Studio 的逻辑。例如：

- 如果没有用户配置文件，则自动创建该文件。
- 通过解析SAML属性，为 SageMaker AI Studio 执行角色附加或移除角色或策略文档。
- 通过添加生命周期配置 (LCC) 和添加标签来自定义用户配置文件。

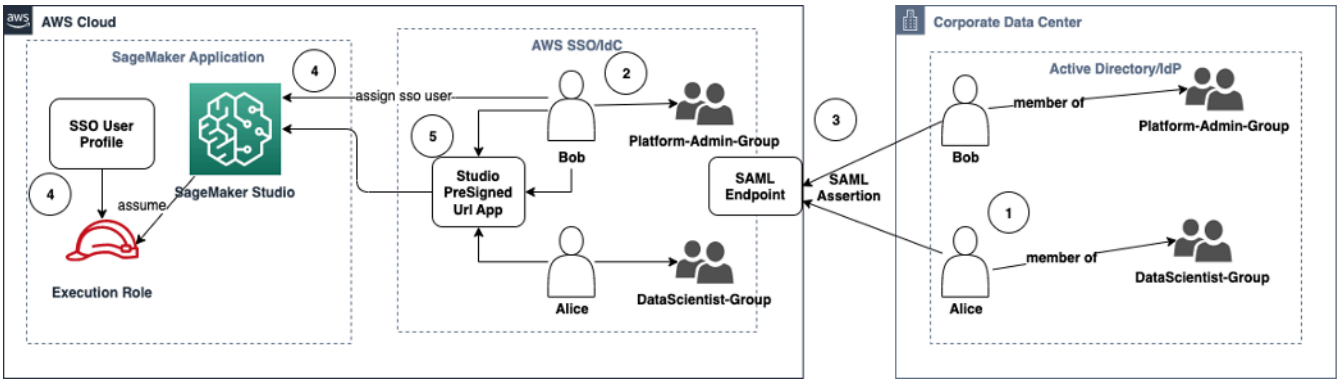
总而言之，此解决方案将 SageMaker AI Studio 公开为 SAML2 .0 应用程序，具有用于身份验证和授权的自定义逻辑。有关实现的详细信息，请参阅附录部分 SageMaker Studio 使用SAML断言进行访问。



使用自定义SAML应用程序访问 SageMaker AI Studio

AWSIAMIDC 联合会

iDC 联合方法使客户能够从其 Id SAML P（例如 Okta）直接联合到 SageMaker AI Studio 应用程序。下图描述了如何授权联合用户访问自己的 SageMaker AI Studio 实例。



在 iD IAM C 模式下访问 SageMaker AI Studio

1. 企业 AD 中的用户属于 AD 组，如平台管理员组和数据科学家组。
2. 来自身份提供商 (IdP) 的 AD 用户和 AD 组将同步到 Ident AWS IAM ity Center，分别作为单点登录用户和群组进行分配。
3. IdP 向 IdC 终端节点发布SAML断言。 AWS SAML
4. 在 SageMaker AI Studio 中，iDc 用户被分配给 SageMaker Studio 应用程序。此任务可以使用 iDC 群组完成， SageMaker AI Studio 将应用于每个 iDC 用户级别。创建此任务后， SageMaker AI Studio 会创建 IdC 用户配置文件并附加域执行角色。

5. 用户使用 iDC 作为云应用程序URL托管的安全预签名访问 SageMaker AI Studio 应用程序。SageMaker AI Studio 承担附加到其 iDC 用户个人资料的执行角色。

域身份验证指南

选择域身份验证模式时，需要考虑以下几点：

1. 如果您希望用户不直接访问 AWS Management Console 和查看 SageMaker AI Studio 用户界面，请使用 i AWS IAM DC 的单点登录模式。
2. 如果您希望用户不在IAM模式下直接访问 AWS Management Console 和查看 SageMaker AI Studio 用户界面，则可以在后端使用 Lambda 函数URL为用户配置文件生成预签名，然后将其重定向到 AI Studio 用户界面。 SageMaker
3. 在 IdC 模式下，每位用户映射一个用户配置文件。
4. 在 IdC 模式下，为所有用户配置文件自动分配默认执行角色。如果您希望为用户分配不同的执行角色，则需要使用更新用户配置文件UpdateUserProfileAPI。
5. 如果您想在IAM模式下（使用生成的预签名URL）将 SageMaker AI Studio UI 访问权限限制为VPC端点，而不必穿越互联网，则可以使用自定义解析器。DNS请参阅 Secure Amazon SageMaker AI Studio 预签名第 1 URLs 部分：基础基础设施博客文章。

权限管理

本节讨论设置用于配置和运营 SageMaker AI Studio 域的常用IAM角色、策略和防护栏的最佳实践。

IAM 角色和策略

作为最佳实践，您可能需要首先确定机器学习生命周期中涉及的相关人员和应用程序（即委托人），以及需要向他们授予哪些 AWS 权限。由于 A SageMaker I 是一项托管服务，因此您还需要考虑服务主体，即可以代表用户API拨打电话的 AWS 服务。下图说明了您可能要创建的不同IAM角色，这些角色对应于组织中的不同角色。

SageMaker AI IAM 角色

详细描述了这些角色，并提供了一些IAMpermissions他们需要的具体示例。

- 机器学习管理员用户角色 — 该负责人通过创建工作室域和用户个人资料 (`sagemaker:CreateDomain`,`sagemaker:CreateUserProfile`)、为用户创建 AWS Key Management Service (AWS KMS) 密钥、为数据科学家创建 S3 存储桶以及创建 Amazon 存储ECR库来存放容器来为数据科学家配置环境。他们还可以为用户设置默认配置和生命周期脚本，构建自定义映像并将其附加到 SageMaker AI Studio 域，并提供 Service Catalog 产品，例如定制项目、Amazon EMR 模板。

  例如，由于该委托人不会运行训练作业，因此他们不需要权限即可启动 SageMaker AI 训练或处理作业。如果他们使用基础架构作为代码模板（例如 CloudFormation 或 Terraform）来配置域和用户，则配置服务将扮演这个角色来代表管理员创建资源。此角色可能使用对 SageMaker AI 具有只读访问权限 AWS Management Console。

  此用户角色还需要某些EC2权限才能在私有环境中启动域VPC、加密EFS卷的KMS权限以及为 Studio 创建服务关联角色的权限 (`iam:CreateServiceLinkedRole`)。此类精细权限将在后文说明。

- 数据科学家用户角色 — 该主体是登录 SageMaker AI Studio、浏览数据、创建处理和训练作业和管道等的用户。用户需要的主要权限是启动 SageMaker AI Studio 的权限，其余策略可以由 SageMaker AI 执行服务角色管理。
- SageMaker AI 执行服务角色 — 由于 SageMaker AI 是一项托管服务，因此它代表用户启动作业。由于很多客户都选用单个执行角色来运行训练作业、处理作业或模型托管作业，所以就获准权限而言，此角色往往是最常用的角色。虽然这是一种简单的入门方法，但由于客户在旅程中成熟，他们通常会将笔记本执行角色分成不同的角色来执行不同的API操作，尤其是在已部署的环境中运行这些作业时。

  创建角色后，您可以将角色与 SageMaker AI Studio 域相关联。但是，由于客户可能需要灵活地将不同的角色与域中的不同用户配置文件相关联（例如，根据他们的工作职能），因此您也可以将单独的IAM角色与每个用户配置文件相关联。亚马逊建议您将单个物理用户映射到对应的用户配置文件。如果您在创建用户配置文件时没有将角色附加到用户配置文件，则默认行为是将 SageMaker AIStudio 域执行角色与用户配置文件关联起来。

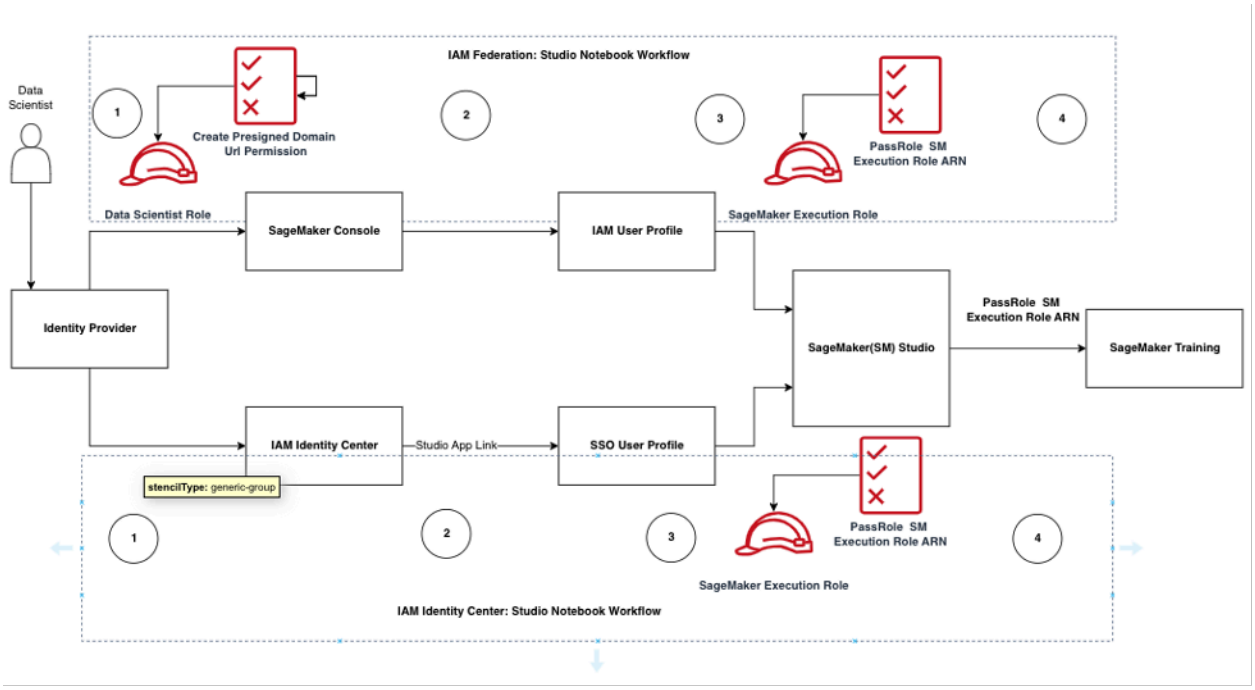
  如果多个数据科学家和机器学习工程师共同处理一个项目，并且需要共享权限模型来访问资源，我们建议您创建一个团队级的 SageMaker AI 服务执行角色，以便在团队成员之间共享IAM权限。在需要锁定每个用户级别的权限的情况下，您可以创建单独的用户级 SageMaker AI 服务执行角色；但是，您需要注意自己的服务限制。

SageMaker AI Studio 笔记本授权工作流程

本节讨论 SageMaker AI Studio 笔记本授权如何适用于数据科学家需要执行的各种活动，以便直接从 SageMaker AI Studio Notebook 上构建和训练模型。A SageMaker I 域支持两种授权模式：

- IAM 联合身份验证
- IAM身份中心

下文将介绍每种模式的数据科学家授权工作流。



适用于 Studio 用户的身份验证和授权工作流

IAM联合： SageMaker Studio 笔记本工作流程

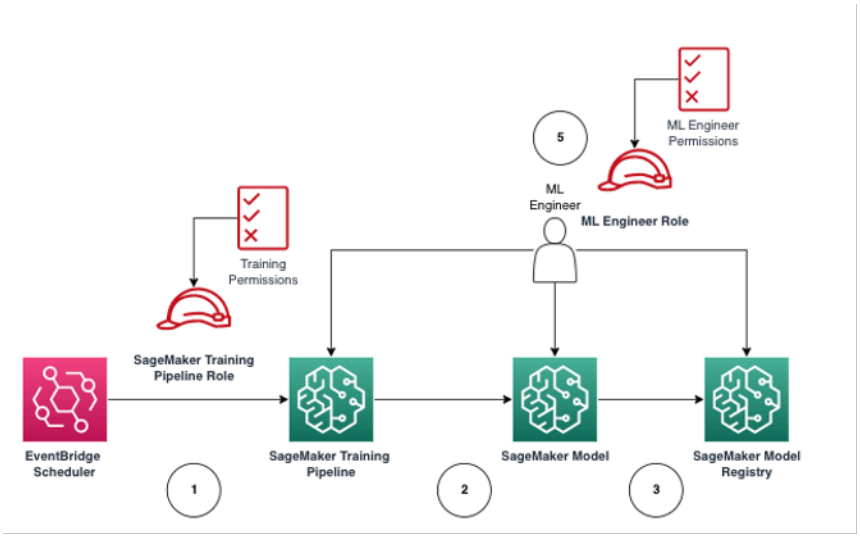
1. 数据科学家向其企业身份提供者进行身份验证，并在 SageMaker AI 控制台中担任数据科学家用户角色（用户联合角色）。此联合角色拥有对 SageMaker AI 执行角色的`iam:PassRole`API权限，可以将角色亚马逊资源名称 (ARN) 传递给 SageMaker Studio。
2. 数据科学家从其 Studio IAM 用户个人资料中选择与 SageMaker AI 执行角色关联的 Open Studio 链接
3. 假设用户配置文件具有 SageMaker 执行角色权限， SageMaker Studio IDE 服务即会启动。此角色拥有对 SageMaker AI 执行角色的`iam:PassRole`API权限，可以将该角色ARN传递给 SageMaker AI 训练服务。
4. 当数据科学家在远程计算节点中启动训练作业时， SageMaker AI 执行角色ARN将传递给 SageMaker AI 训练服务。这将使用它创建一个新的角色会话ARN并运行训练作业。如果您需要进一步缩小训练作业的权限范围，则可以创建特定于训练的角色并在调用 training ARN 时传递该角色API。

IAM身份中心： SageMaker AI Studio 笔记本工作流程

1. 数据科学家向其企业身份提供商进行身份验证，然后单击 Ident AWS IAM ity Center。数据科学家会看到 Identity Center 用户门户。
2. 数据科学家点击根据其 iDC 用户个人资料创建的 SageMaker AI Studio 应用程序链接，该链接与 SageMaker AI 执行角色相关联。
3. 假设用户配置文件具有 SageMaker AI 执行角色权限，则启动 A SageMaker I Studio IDE 服务。此角色拥有对 SageMaker AI 执行角色的`iam:PassRole`API权限，可以将该角色ARN传递给 SageMaker AI 训练服务。
4. 当数据科学家在远程计算节点中启动训练作业时， SageMaker AI 执行角色ARN将传递给 SageMaker AI 训练服务。执行角色使用ARN它创建新的角色会话ARN，并运行训练作业。如果您需要进一步缩小训练作业的权限范围，则可以创建特定于训练的角色并在调用培训ARN时传递该角色。API

部署环境： SageMaker AI 训练工作流程

在系统测试和生产等已部署环境中，作业通过自动调度程序和事件触发器运行， SageMaker AI Studio Notebook 对这些环境的访问受到限制。本节讨论IAM角色如何在已部署环境中与 SageMaker AI 训练管道配合使用。



SageMaker 托管生产环境中的 AI 训练工作流程

1. Amazon EventBridge 计划程序会触发 SageMaker AI 训练管道作业。
2. A SageMaker I 训练管道作业担任 SageMaker AI 训练管道角色来训练模型。
3. 经过训练的 SageMaker AI 模型已注册到 SageMaker AI 模型注册表中。

4. 机器学习工程师扮演机器学习工程师用户角色来管理训练管道和 SageMaker AI 模型。

数据权限

SageMaker AI Studio 用户访问任何数据源的能力受与其 SageMaker AI IAM 执行角色关联的权限的约束。所附的策略可以授权他们读取、写入或删除某些 Amazon S3 存储桶或前缀，以及连接到 Ama RDS zon 数据库。

访问 AWS Lake Formation 数据

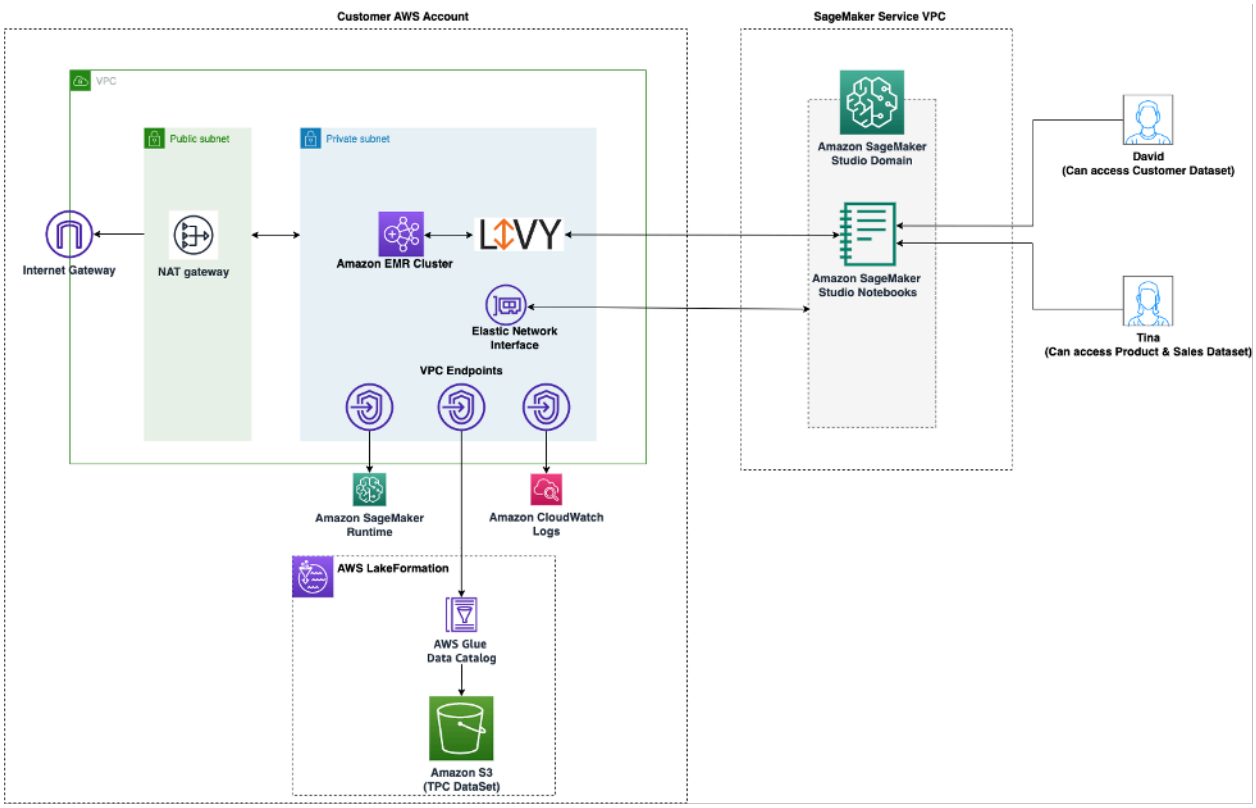
已经有不少企业开始使用受 AWS Lake Formation 监管的数据湖，为用户提供精细数据访问权限。以此类受管数据为例，管理员可为部分用户屏蔽敏感列，同时可查询同一基础表。

要使用 SageMaker AI Studio 的 Lake Formation，管理员可以 SageMaker 将 AI IAM 执行角色注册为`DataLakePrincipals`。有关更多信息，请参阅 Lake Formation 权限参考。获得授权后，有三种主要方法可以从 SageMaker AI Studio 访问和写入受管理的数据：

1. 在 A SageMaker I Studio 笔记本上，用户可以利用 Amazon A thena 等查询引擎或基于 boto3 构建的库将数据直接提取到笔记本中。f AWSSDKor Pandas（以前称为 awswrangler）是一个受欢迎的图书馆。以下代码示例说明了无缝操作的方法：

```
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. 使用 SageMaker AI Studio 与亚马逊的原生连接EMR来大规模读取和写入数据。通过使用 Apache Livy 和 Amazon EMR 运行时角色， SageMaker AI Studio 建立了原生连接，允许您将 SageMaker 人工智能执行IAM角色（或其他授权角色）传递给亚马逊EMR集群进行数据访问和处理。有关 up-to-date说明，请参阅从 Studio 连接到亚马逊EMR集群。



用于访问 SageMaker Studio 中由 Lake Formation 管理的数据的架构

3. 使用 SageMaker AI Studio 与交AWS Glue 互式会话的原生连接来大规模读取和写入数据。SageMaker AI Studio 笔记本具有内置内核，允许用户以交互方式在上运行命令。AWS Glue用户可以大规模使用 Python、Spark 或 Ray 后端，从受管数据来源中无缝读取并写入大量数据。内核允许用户传递其 SageMaker 执行角色或其他授权IAM角色。有关更多信息，请参阅使用 AWS Glue 交互式会话准备数据。

通用防护机制

本节讨论使用策略、资源策略、VPC端点策略和服务控制IAM策略（ ）对机器学习资源进行监管的最常用防护措施。SCPs

限制笔记本访问特定实例

此服务控制策略可限制数据科学家在创建 Studio 笔记本时可访问的实例类型。请注意，任何用户都需要允许的 “系统” 实例来创建托管 SageMaker AI Studio 的默认 Jupyter Server 应用程序。

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

限制不合规的 SageMaker AI Studio 域

对于 SageMaker AI Studio 域，可以使用以下服务控制策略来强制访问客户资源的流量，这样他们就不会通过公共互联网而是通过客户的互联网进行访问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"
```

```
                }
            }
        }
    ]
}
```

限制启动未经授权的 SageMaker AI 镜像

以下策略可防止用户在其域内启动未经授权的 SageMaker AI 镜像：f

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

仅通过 SageMaker AI VPC 端点启动笔记本电脑

除了 SageMaker AI 控制平面的VPC端点之外，A SageMaker I 还支持用户连接到 AI Studio 笔记本电脑或 SageMaker A SageMaker I 笔记本实例的VPC端点。如果您已经为 SageMaker AI Studio/Notebook 实例设置了VPC终端节点，则以下IAM条件键仅允许通过 SageMaker AI Studio VPC 终端节点或 AI 终端节点连接到 SageMaker AI Studio 笔记本电脑。 SageMaker API

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

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

将 SageMaker AI Studio 笔记本电脑的访问权限限制在有限的 IP 范围内

公司通常会将 SageMaker AI Studio 的访问权限限制在某些允许的企业 IP 范围内。以下带有`SourceIP`条件键的IAM策略可以对此进行限制。

```
{
    "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"
                    ]
                }
```

```
            }
        }
    ]
}
```

阻止 SageMaker AI Studio 用户访问其他用户个人资料

作为管理员，在创建用户配置文件时，请确保该配置文件标有 SageMaker AI Studio 用户名和标签密钥`studiouserid`。还应使用 `studiouserid` 键（可随意命名此标签，不限于 `studiouserid`）标记主体（用户或其附加的角色）。

接下来，将以下策略附加到用户在启动 SageMaker AI Studio 时将扮演的角色。

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

执行标记操作

数据科学家需要使用 SageMaker AI Studio 笔记本电脑来探索数据、构建和训练模型。对笔记本应用标签这一行为有助于监控使用情况并控制成本，同时保障所有权和可审核性。

对于 SageMaker AI Studio 应用程序，请确保已标记用户个人资料。这些标签会自动从用户配置文件传播到应用程序。要强制使用标签创建用户个人资料（通过CLI和支持SDK），请考虑将此策略添加到管理员角色：

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

对于训练作业和处理作业等资源，可采用以下策略强制要求进行标记：

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

SageMaker AI Studio 中的根访问权限

在 SageMaker AI Studio 中，笔记本在 Docker 容器中运行，默认情况下，该容器没有主机实例的根访问权限。同样，除了默认的 run-as user 之外，容器内的所有其他用户 ID 范围都将在主机实例本身IDs上重新映射为非特权用户。这样就只有笔记本容器本身会产生权限升级威胁。

创建自定义映像时，可能要为用户提供非根权限以加强控制；例如，不以根用户身份运行不良流程或安装公开软件包。在此情况下，您可以在 Dockerfile 中创建以非根用户身份运行的映像。无论您以 root 用户还是非 root 用户身份创建用户，都需要确保自定义应用程序UID/GID of the user is identical to the UID/GID中的，这会为 SageMaker AI 创建使用自定义映像运行应用程序的配置。AppImageConfig如果您是为了以下非根用户构建的 Dockerfile：

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

该AppImageConfig文件需要提及同样的内容UID，并在其GID中KernelGatewayConfig：

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

对于自定义图像，Studio 图像的可接受UID/GID值为 0/0 和 1000/100。有关构建自定义映像和 AppImageConfig 关联设置的示例，请参阅此 Github 存储库。

为避免用户篡改此权限，请勿向 SageMaker AI Studio 笔记本电脑用户授予CreateAppImageConfigUpdateAppImageConfig、或DeleteAppImageConfig权限。

网络管理

要设置 SageMaker AI Studio 域，您需要指定VPC网络、子网和安全组。指定VPC和子网时，请确保在分配时IPs考虑以下各节中讨论的使用量和预期增长。

VPC网络规划

与 SageMaker AI Studio 域关联的客户VPC子网必须使用相应的无类域间路由 (CIDR) 范围创建，具体取决于以下因素：

- 用户数。
- 每位用户的应用程序数量。
- 每位用户的唯一实例类型数量。
- 每位用户的训练实例平均数。
- 预期增长百分比。

SageMaker AI 和参与的 AWS 服务将弹性网络接口 (ENI) 注入到客户VPC子网中，用于以下用例：

- Amazon 为 ENI A SageMaker I 域EFS注入一个EFS挂载目标（每个子网/附加到 AI 域的可用区一个 IP）。 SageMaker
- SageMaker AI Studio 会ENI为用户个人资料或共享空间使用的每个唯一实例注入一个。例如：
  - 如果用户配置文件运行一个默认 Jupyter 服务器应用程序（“系统”实例）、一个数据科学应用程序和一个 Base Python 应用程序（均运行在 `ml.t3.medium` 实例上），则 Studio 会注入两个 IP 地址。
  - 如果用户配置文件运行默认 Jupyter 服务器应用程序（一个 “系统” 实例）、一个 Tensorflow GPU 应用程序（在一个`ml.g4dn.xlarge`实例上）和一个数据管理器应用程序（在一个`ml.m5.4xlarge`实例上），Studio 会注入三个 IP 地址。
- ENI为跨域VPC子网/可用区域的每个VPC端点注入一个（ SageMaker AI VPC 端点IPs为四个；参与IPs的服务VPC端点（例如 S3、ECR和。 ） CloudWatch
- 如果以相同的VPC配置启动 SageMaker AI 训练和处理作业，则每个作业需要每个实例两个 IP 地址。

> **Note**
>
> VPC SageMaker AI Studio 的设置（例如子网和VPC仅限流量）不会自动传递到从 AI Studio 创建的训练/处理作业。 SageMaker 用户在调用 Create* APIs Job 时需要根据需要VPC设置设置和网络隔离。有关更多信息，请参阅在互联网免费模式下运行训练和推理容器。

场景：数据科学家在两种不同的实例类型上运行实验

在这种情况下，假设 A SageMaker I 域设置为VPC仅限流量模式。设置了VPC终端节点，例如 SageMaker AI API、A SageMaker I 运行时、Amazon S3 和亚马逊ECR。

数据科学家正在 Studio 笔记本上运行实验，选用两种不同的实例类型（如 `ml.t3.medium` 和 `ml.m5.large`）并分别启动两个应用程序。

假设数据科学家还同时在`ml.m5.4xlarge`实例上运行具有相同VPC配置的训练作业。

在这种情况下， SageMaker AI Studio 服务将按ENIs如下方式注入：

表 1 — VPC 针对实验场景向客户ENIs注入

值。	目标	ENI注射	注意	级别
EFS挂载目标	VPC子网	三	三个 AZs / subnets	域
VPC 端点	VPC子网	30	三个 AZs /子网，每个子网 10 VPCE	域
Jupyter 服务器	VPC 子网	One	每个实例对应一个 IP	用户
KernelGateway 应用程序	VPC 子网	二	每种实例类型对应一个 IP	用户
训练	VPC 子网	二	IPs每个训练实例两个	用户

值。	目标	ENI注射	注意	级别
			如果使用，EFA则IPs每个训练实例五个	

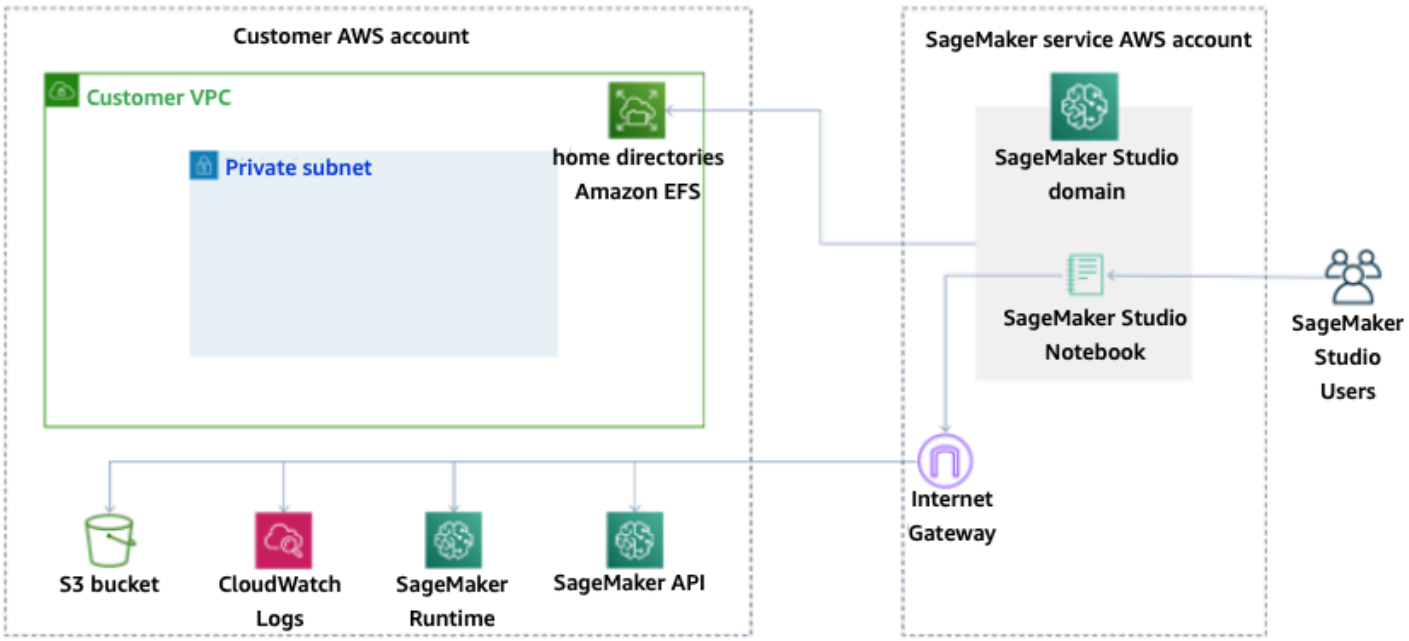
在此场景中，客户总共使用了 38 IPs 个，VPC其中 33 IPs 个在域级别的用户之间共享，5 IPs 个在用户级别消费。如果您在该域中有 100 个用户配置文件相似的用户同时执行这些活动，那么除了域级别 IP 消耗（IPs每个子网 11 个）之外，您还将在用户级别消耗 5 x 100 = 500IPs，总共消耗 5 IPs 11 个。在这种情况下，您需要创建CIDR带有 /22 的子网，该VPC子网将分配 1024 个 IP 地址，并有增长空间。

VPC网络选项

A SageMaker I Studio 域支持使用以下选项之一配置VPC网络：

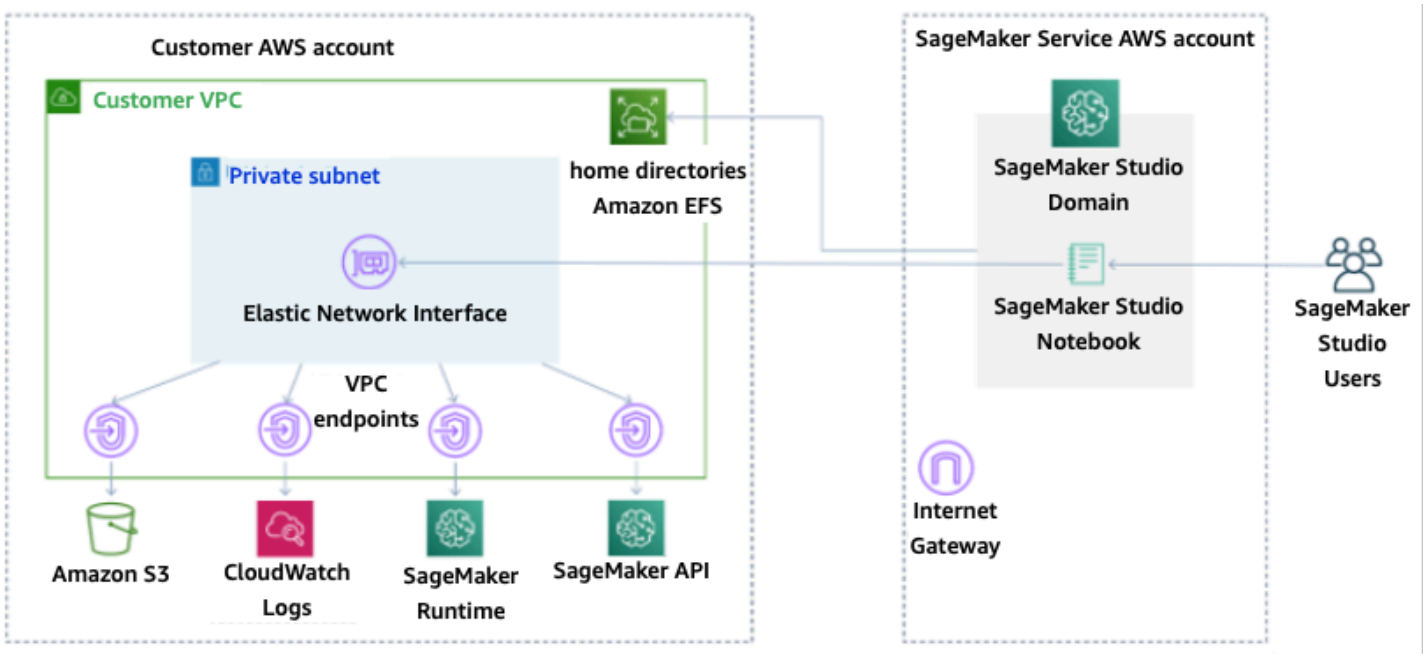
- 仅限公共互联网
- 仅限 VPC

仅限公共互联网选项允许 SageMaker AI API 服务通过中配置的互联网网关使用公共互联网VPC，由 SageMaker AI 服务账户管理，如下图所示：



默认模式：通过 SageMaker AI 服务账号访问互联网

VPC唯一的选项禁用由 SageMaker AI 服务帐户VPC管理的互联网路由，并允许客户将流量配置为通过VPC端点路由，如下图所示：



VPC仅限模式：无法通过 SageMaker AI 服务帐户访问互联网

对于在VPC仅限模式下设置的域，请为每个用户配置文件设置一个安全组，以确保底层实例完全隔离。 AWS 账户中的每个域名都可以有自己的VPC配置和互联网模式。有关设置VPC网络配置的更多详细信息，请参阅外部资源中的 Connect SageMaker AI Studio 笔记本电脑。VPC

限制

- 创建 A SageMaker I Studio 域后，您无法将新子网关联到该域。
- VPC网络类型（仅限公共互联网或VPC仅限公共互联网）无法更改。

数据保护

应该先建立影响安全性的基础实践，再搭建机器学习工作负载的架构。例如，数据分类可根据敏感级别划分，而加密手段能够阻止未经授权的访问者，从而保护数据。这些方法有助于避免误操作或履行监管义务等，意义重大。

SageMaker AI Studio 提供了多种保护静态和传输中数据的功能。但是，如责任AWS 共担模型中所述，客户有责任保持对 AWS 全球基础设施上托管的内容的控制。本节介绍了客户利用这些功能保护数据安全的方法。

保护静态数据

为了保护您的 SageMaker AI Studio 笔记本以及模型构建数据和模型工件， SageMaker AI 会对笔记本以及训练和批处理转换作业的输出进行加密。 SageMaker 默认情况下，AI 使用适用于 A mazon S3 的 AWS 托管密钥对这些密钥进行加密。此适用于 Amazon S3 的 AWS 托管密钥无法共享以供跨账户访问。对于跨账户访问，请在创建 SageMaker AI 资源时指定您的客户管理的密钥，以便可以共享该密钥以进行跨账户访问。

使用 SageMaker AI Studio，可以将数据存储在以下位置：

- S3 存储桶 — 启用可共享笔记本后， SageMaker AI Studio 会在 S3 存储桶中共享笔记本快照和元数据。
- EFSv olume — SageMaker AI Studio 将一个EFS卷连接到您的域中，用于存储笔记本和数据文件。即使删除域名后，此EFS卷仍会保留。
- EBSvol EBS ume — 连接到运行笔记本电脑的实例。实例运行期间，此卷持续存在。

静态加密 AWS KMS

- 您可以传递AWS KMS 密钥来加密附加到笔记本、训练、调整、批量转换作业和端点的EBS卷。
- 如果您未指定KMS密钥， SageMaker AI 会使用系统管理KMS的密钥对操作系统 (OS) 卷和 ML 数据卷进行加密。
- 出于合规原因需要使用KMS密钥加密的敏感数据应存储在 ML 存储卷或 Amazon S3 中，这两者都可以使用您指定的KMS密钥进行加密。

保护传输中的数据

SageMaker AI Studio 确保机器学习模型工件和其他系统工件在传输过程中和静态时都经过加密。对 SageMaker AI API 和控制台的请求是通过安全 (SSL) 连接发出的。部分网络内（服务平台内部）传输中数据未加密。其中包括：

- 服务控制面板和训练作业实例（不是客户数据）之间的命令和控制通信。
- 分布式处理和训练作业（网络内）中节点之间的通信。

您也可以对训练集群中节点之间的通信进行加密。启用容器间流量加密可能会延长训练时间，尤其是在使用分布式深度学习算法的情况下。

默认情况下，Amazon SageMaker AI 在亚马逊运行训练作业VPC，以帮助保护您的数据安全。您可以通过配置私有来增加另一个安全级别来保护您的训练容器和数据VPC。此外，您可以将 SageMaker AI Studio 域配置为VPC仅在模式下运行，并将VPC终端节点设置为通过私有网络路由流量，而不会通过 Internet 流出流量。

数据保护防护机制

加密静态的 SageMaker AI 托管卷

在托管用于在线推理的 SageMaker AI 终端节点期间，使用以下策略强制加密：

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

```
 }
```

加密模型监控期间使用的 S3 存储桶

模型监控会捕获发送到您的 SageMaker AI 终端节点的数据并将其存储在 S3 存储桶中。设置 Data Capture Config 时需要加密 S3 存储桶。目前对此尚无补偿控制措施。

除了捕获端点输出外，模型监控服务还会对照预先指定的基线，检查是否出现偏差。输出流量和用于监控偏差的中间存储卷均需加密。

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

加密 A SageMaker I Studio 域存储卷

对挂载至 Studio 域的存储卷执行加密操作。此策略要求用户提供对附加到 st CMK udio 域的存储卷进行加密。

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

```
            "Action": [
                "sagemaker:CreateDomain"
            ],
            "Resource": "*",
            "Condition": {
                "Null": {
                    "sagemaker:VolumeKmsKey": "false"
                }
            }
        }
    ]
}
```

加密 S3 中存储的用于共享笔记本的数据

以下策略用于加密存储在存储桶中用于在 SageMaker AI Studio 域中的用户之间共享笔记本的所有数据：

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

限制

- 创建域后，您将无法使用自定义 AWS KMS 密钥更新附加的EFS卷存储。

- 一旦创建了训练/处理任务或终端节点配置，就无法使用KMS密钥更新它们。

日志记录和监控

为了帮助您调试编译作业、处理作业、训练作业、终端节点、转换作业、笔记本实例和笔记本实例生命周期配置，算法容器、模型容器或笔记本实例生命周期配置发送到 stdout 或 stderr 的任何内容也会发送到 Amazon Logs。 CloudWatch 您可以使用 Amazon 监控 A SageMaker I Studio CloudWatch，亚马逊会收集原始数据并将其处理为可读的近乎实时的指标。这些统计数据会保存 15 个月，方便您访问历史信息并更好地了解 Web 应用程序或服务的运行情况。

使用登录 CloudWatch

数据科学流程本质上具有实验性和迭代性，必须记录笔记本使用情况、训练/处理作业运行时间、训练指标和端点服务指标（如调用延迟）等活动。默认情况下， SageMaker AI 会将指标发布到 CloudWatch 日志，并且可以使用使用 AWS KMS客户管理的密钥对这些日志进行加密。

您也可以使用VPC终端节点向其发送日志， CloudWatch 而无需使用公共互联网。还可以设置特定阈值监视警报，在达到对应阈值时发送通知或采取行动。有关更多信息，请参阅 A mazon CloudWatch 用户指南。

SageMaker AI 在下方为 Studio 创建了一个日志组`/aws/sagemaker/studio`。此日志组下的每个用户配置文件和应用程序都有专属日志流，而生命周期配置脚本也有专属日志流。例如，“studio-user”用户配置文件具有 Jupyter 服务器应用程序和附加的生命周期脚本，而数据科学内核网关应用程序具有以下日志流：

```
/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
```

为了 SageMaker 让 AI 代表你向其 CloudWatch 发送日志，Training/Processing/Transform任务的调用者APIs需要以下权限：

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

要使用自定义 AWS KMS 密钥加密这些日志，您首先需要修改密钥策略以允许 CloudWatch 服务加密和解密密钥。创建日志加密 AWS KMS 密钥后，请修改密钥策略以包括以下内容：

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

```
    ]
}
```

请注意，对于要加密的 CloudWatch 日志，您可以随时使用ArnEquals并提供特定的 Amazon 资源名称 (ARN)。为简单起见，亚马逊为您演示如何使用此密钥加密账户中的所有日志。此外，训练、处理和模型端点还会发布有关实例CPU和内存利用率、托管调用延迟等的指标。您可以进一步配置AmazonSNS，使其在超过特定阈值时将事件通知管理员。训练和处理的使用者APIs需要具有以下权限：

```
{
     "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"
```

```
        }
    ]
}
```

使用审计 AWS CloudTrail

为了改善您的合规状况，请对您的所有内容APIs进行审计 AWS CloudTrail。默认情况下，所有 SageMaker AI 都会APIs被记录下来AWS CloudTrail。您不需要任何其他IAM权限即可启用 CloudTrail。

除InvokeEndpoint和InvokeEndpointAsync之外，所有 SageMaker AI 操作都由操作记录 CloudTrail 并记录在操作中。例如，对CreateTrainingJobCreateEndpoint、和CreateNotebookInstance操作的调用会在 CloudTrail 日志文件中生成条目。

每个 CloudTrail 事件条目都包含有关谁生成请求的信息。身份信息有助于您确定以下内容：

- 请求是使用根用户凭证还是 AWS IAM 用户凭证发出的。
- 请求是使用角色还是联合用户的临时安全凭证发出的。
- 请求是否由其他 AWS 服务发出。有关示例事件，请参阅附带 CloudTrail文档的 “记录 SageMaker AI API 通话”。

默认情况下，将用户配置文件的 Studio 执行角色名称 CloudTrail 记录为每个事件的标识符。此方法适用于每位用户都有专属执行角色的情况。如果多个用户共享同一个执行角色，则可以使用该sourceIdentity配置将 Studio 用户配置文件名称传播到 CloudTrail。要启用该sourceIdentity功能，请参阅监控来自 Amazon A SageMaker I Studio 的用户资源访问权限。在共享空间中，所有操作都将该空间ARN视为来源，您无法通过审计sourceIdentity。

成本归属

SageMaker AI Studio 内置功能可帮助管理员跟踪其个人域、共享空间和用户的支出。

自动标记

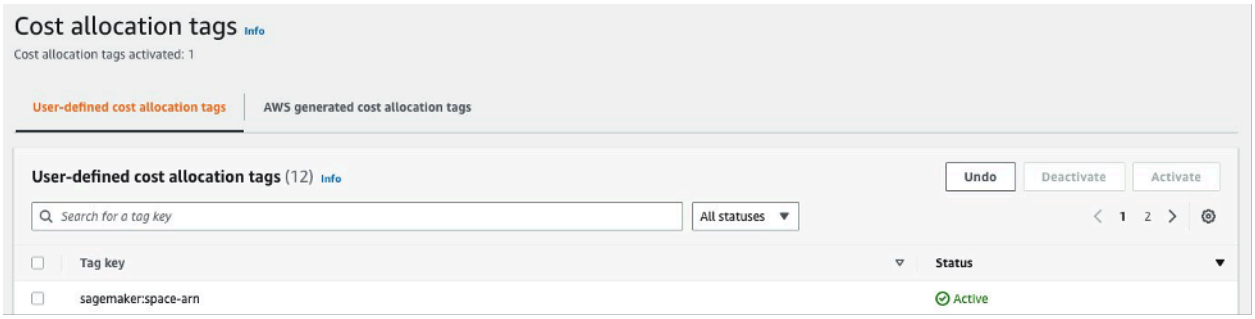
SageMaker AI Studio 现在会自动为训练作业、处理作业和内核应用程序等新 SageMaker 资源添加相应的标签`sagemaker:domain-arn`。在更精细的层面上， SageMaker AI 还使用`sagemaker:user-profile-arn`或标记资源，`sagemaker:space-arn`以指定资源的主要创建者。

SageMaker AI 域EFS卷使用以域值命名的密钥`ManagedByAmazonSageMakerResource`进行标记ARN。这些卷没有精细标签，无法了解每个用户级别上的空间使用情况。不过，管理员可以将EFS卷连接到EC2实例以进行定制监控。

成本监控

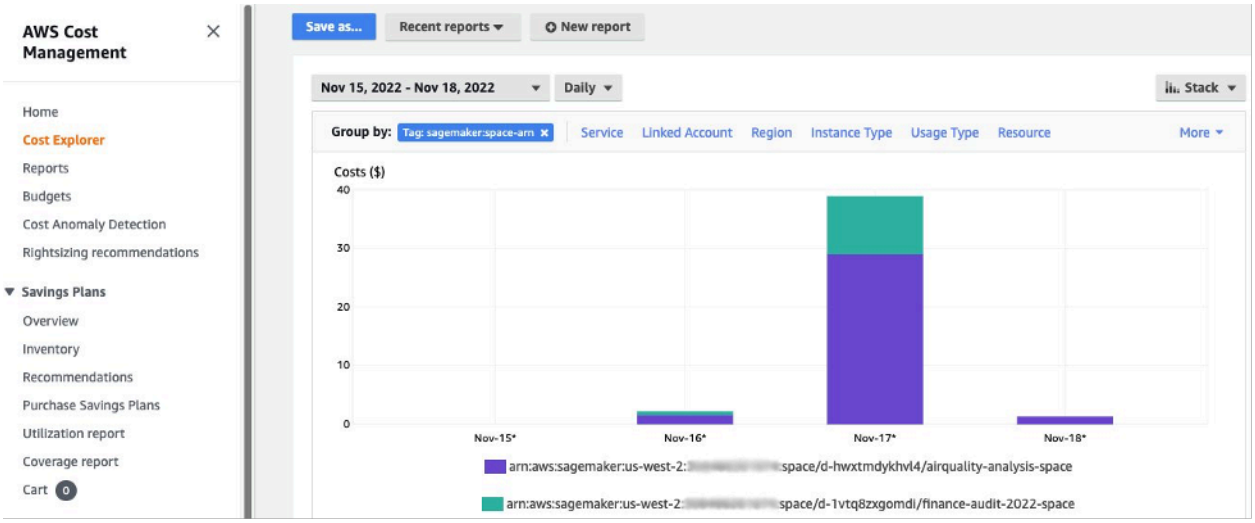
自动标签使管理员能够通过和等 out-of-the-box解决方案以及基于AWS 成本AWS Cost Explorer和使用情况报告中的数据构建的自定义解决方案（CURs）来跟踪AWS Budgets、报告和监控您的机器学习支出。

必须先在 AWS Billing 控制台的成本分配标签一节中激活已附加的标签，才能利用其分析成本。标签最多可能需要 24 小时才能显示在成本分配标签面板中，因此您需要先创建 A SageMaker I 资源，然后才能启用它们。



在 Cost Explorer 上ARN启用空间作为成本分配标签

启用成本分配标签后， AWS 将开始跟踪您标记的资源，24-48 小时后，这些标签将在成本资源管理器中显示为可选筛选器。



按照示例域的共享空间分组的成本

成本控制

当第一个 SageMaker AI Studio 用户加入后， SageMaker AI 会为该域创建一个EFS音量。此EFS卷会产生存储成本，因为笔记本和数据文件存储在用户的主目录中。当用户启动 Studio 笔记本，就会对支持笔记本运行的计算实例开始计费。有关详细的成本明细，请参阅 SageMaker Amazon AI 定价。

管理员可以使用通用护栏部分中提到的IAM策略，通过指定用户可以启动的实例列表来控制计算成本。此外，我们建议客户使用 SageMaker AI Studio 自动关闭扩展程序，通过自动关闭闲置的应用程序来节省成本。此服务器扩展程序会定期轮询每个用户配置文件中正在运行的应用程序，并根据管理员设置的超时时间关闭空闲的应用程序。

如需为域中所有用户设置此扩展程序，您可以使用自定义一节中描述的生命周期配置。也可以使用扩展检查程序，确保域中所有用户都安装了此扩展程序。

自定义

生命周期配置

生命周期配置是由 SageMaker AI Studio 生命周期事件（例如启动新 SageMaker 的 AI Studio 笔记本电脑）启动的 shell 脚本。您可以使用这些 shell 脚本自动对 SageMaker AI Studio 环境进行自定义，例如安装自定义包、用于自动关闭非活动笔记本应用程序的 Jupyter 扩展以及设置 Git 配置。有关如何构建生命周期配置的详细说明，请参阅此博客：使用生命周期配置自定义 Amazon SageMaker AI Studio。

SageMaker AI Studio 笔记本电脑的自定义镜像

Studio 笔记本附带一组预先构建的镜像，其中包括 Amazon A SageMaker I Python SDK 和最新版本的IPython运行时或内核。借助此功能，您可以将自己的自定义图像带到 Amazon A SageMaker I 笔记本上。之后，所有通过身份验证进入域的用户均可使用这些映像。

开发人员和数据科学家可能要对以下几种用例使用自定义映像：

- 访问常用 ML 框架的特定或最新版本 TensorFlow，例如、MXNet PyTorch、或其他。
- 将本地开发的自定义代码或算法带到 SageMaker AI Studio 笔记本中，用于快速迭代和模型训练。
- 通过访问数据湖或本地数据存储APIs。管理员需要为映像加入相应的驱动程序。
- 访问除了 R、Julia 或其他）之外的后端运行时IPython（也称为内核）。您也可以使用所述方法安装自定义内核。

有关如何构建自定义镜像的详细说明，请参阅创建自定义 A SageMaker I 镜像。

JupyterLab 扩展

有了 SageMaker AI Studio JuypterLab 3 Notebook，你可以充分利用不断增长的开源 JupyterLab扩展社区。本节重点介绍一些自然适合 SageMaker AI 开发者工作流程的扩展，但我们鼓励您浏览可用的扩展程序，甚至创建自己的扩展。

JupyterLab 3 现在大大简化了打包和安装扩展的过程。您可以使用 Bash 脚本安装上述扩展程序。例如，在 SageMaker AI Studio 中，从 Studio 启动器打开系统终端并运行以下命令。还可以使用生命周期配置自动安装这些扩展程序，使其在 Studio 重启期间也能继续生效。您可以为域中所有用户或在个人用户级别上配置此扩展程序。

例如，如需为 Amazon S3 文件浏览器安装扩展程序，请在系统终端中运行以下命令并刷新浏览器：

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

有关扩展管理的更多信息，包括如何编写适用于 JupyterLab 笔记本版本 1 和 3 的生命周期配置以实现向后兼容，请参阅安装 JupyterLab 和 Jupyter Server 扩展。

Git 存储库

SageMaker AI Studio 预装了 Jupyter Git 扩展程序，供用户进入URL定制的 Git 存储库、将其克隆到EFS您的目录、推送更改和查看提交历史记录。管理员可配置域级别的建议 Git 存储库，将其作为最终用户的下拉选项。有关 up-to-date说明，请参阅将建议的 Git 存储库附加到 Studio。

如果是私有存储库，则扩展程序会要求用户使用 Git 标准安装程序，将其凭证输入终端。或者，用户可以将 ssh 凭据存储在其个人EFS目录中，以便于管理。

Conda 环境

SageMaker AI Studio 笔记本电脑使用亚马逊EFS作为永久存储层。数据科学家可利用永久存储建立 Conda 自定义环境，进而创建内核。这些内核由内核EFS、应用程序或 Studio 重启之间提供支持，并且在内核、应用程序或 Studio 重启之间保持不变。Studio 会自动将所有有效的环境作为KernelGateway内核获取。

虽然数据科学家能够轻松创建 Conda 环境，但内核仍要等待约一分钟才会填充到内核选择器上。如需创建环境，请在系统终端中运行以下命令：

```
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
```

有关详细说明，请参阅在 A mazon Studio 笔记本中管理 Python 包的四种方法中的 Persist Conda 环境到 SageMaker Studio EFS 卷部分。

结论

在本白皮书中，我们回顾了运营模式、域管理、身份管理、权限管理、网络管理、日志、监控和自定义等领域的几种最佳实践，以使平台管理员能够设置和管理 SageMaker AI Studio Platform。

附录

多租户比较

表 2 — 多租户比较

多域	多账号	单个域内基于属性的访问控制 (ABAC)
资源隔离是使用标签实现的。SageMaker AI Studio 会自动使用域ARN和用户配置文件/空间标记所有资源。ARN	每个租户都有自己的账户，因此资源绝对隔离。	资源隔离是使用标签实现的。用户必须管理为创建的资源的标ABAC记。
列表APIs不能受标签限制。用户界面筛选资源是在共享空间上完成的，但是，通过 AWS CLI 或 Boto3 发出的列表API调用SDK将列出整个区域的资源。	列表APIs隔离也是可能的，因为租户位于他们的专用账户中。	列表APIs不能受标签限制。列出通过 AWS CLI 或 Boto3 API 拨打的电话SDK将列出整个地区的资源。
SageMaker 使用 Domain ARN 作为成本分配标签，可以轻松监控每个租户的 AI Studio 计算和存储成本。	SageMaker 使用专用帐户，可以轻松监控每个租户的 AI Studio 计算和存储成本。	SageMaker AI Studio 需要使用自定义标签计算每个租户的计算成本。 SageMaker 无法按域监控 AI Studio 的存储成本，因为所有租户共享相同的存储EFS容量。
服务配额是在账户级别设置的，因此单个租户仍然可以用完所有资源。	可以在账户级别为每个租户设置服务配额。	服务配额是在账户级别设置的，因此单个租户仍然可以用完所有资源。
可以通过基础设施即代码 (IaC) 或 Service Catalog 来扩展到多个租户。	扩展到多个租户涉及Organizations和出售多个帐户。	Scaling 需要为每个新租户指定租户特定的角色，并且需要用

多域	多账号	单个域内基于属性的访问控制 (ABAC)
		租户名称手动标记用户配置文件。
租户内部的用户可以通过共享空间进行协作。	租户内部的用户可以通过共享空间进行协作。	所有租户都可以访问同一个共享空间进行协作。

SageMaker AI Studio 域名备份和恢复

如果意外EFS删除，或者由于网络或身份验证更改而需要重新创建域，请按照以下说明进行操作。

选项 1：EFS使用现有备份 EC2

SageMaker 工作室域名备份

1. 列出 SageMaker Studio 中的用户个人资料和空间 (CLI, SDK)。
2. 将用户个人资料/空间映射到 on。UIDs EFS
   a. 对于 users/spaces, describe the user profile/space (CLI, SDK) 列表中的每个用户。
   b. 映射用户配置文件/空间到 `HomeEfsFileSystemUid`。
   c. 为具有不同执行角色的用户映射配置文件到 `UserSettings['ExecutionRole']`。
   d. 识别默认空间执行角色。
3. 创建新域并指定默认空间执行角色。
4. 创建用户配置文件和空间。
   - 对于用户列表中的每个用户，使用执行角色映射创建用户配置文件 (CLI, SDK)。
5. 为新的EFS和创建映射UIDs。
   a. 对于用户列表中的每个用户，请描述用户个人资料 (CLI, SDK)。
   b. 映射用户配置文件到 `HomeEfsFileSystemUid`。
6. 可以先删除所有应用程序、用户配置文件和空间，再删除域。

EFS 备份

要进行备份EFS，请按照以下说明进行操作：

1. 启动EC2实例，并将旧 SageMaker Studio 域的入站/出站安全组附加到新EC2实例（允许端口 2049 TCP 上的NFS流量）。请参阅 “外部资源” 中的 Connect SageMaker Studio 笔记本电脑。VPC
2. 将 SageMaker Studio EFS 卷挂载到新EC2实例。请参阅挂载EFS文件系统。
3. 将文件复制到EBS本地存储：`>sudo cp -rp /efs /studio-backup:`
   a. 将新的域安全组附加到实EC2例。
   b. 将新EFS卷挂载到EC2实例。
   c. 将文件复制到新EFS卷。
   d. 对于用户集合中的每位用户：
      i. 创建目录：`mkdir new_uid`。
      ii. 将文件从旧UID目录复制到新UID目录。
      iii. 更改所有文件的所有权：所有文件的 `chown <new_UID>`。

选项 2：EFS使用 S3 和生命周期配置从现有设备进行备份

1. 请参阅使用亚马逊 Linux 将您的作品迁移到亚马逊 SageMaker 笔记本实例 2。
2. 创建 S3 存储桶进行备份（例如 `>studio-backup`）。
3. 列出所有具有执行角色的用户配置文件。
4. 在当前 SageMaker Studio 域中，在域级别设置默认LCC脚本。
   - 在中LCC，将所有内容复制`/home/sagemaker-user`到 S3 中的用户配置文件前缀中（例如，`s3://studio-backup/studio-user1`）。
5. 重新启动所有默认 Jupyter Server 应用程序（LCC以便运行）。
6. 删除所有应用程序、用户配置文件和域。
7. 创建一个新的 SageMaker Studio 域名。
8. 从用户配置文件和执行角色列表新建用户配置文件。
9. LCC在域级别设置：
   - 在中LCC，将 S3 中用户配置文件前缀中的所有内容复制到 `/home/sagemaker-user`
10. 使用LCC配置 (CLI,) 为所有用户创建默认 Jupyter Server 应用程序。SDK

SageMaker 使用SAML断言访问工作室

解决方案设置步骤：

1. 在外部 IdP 中创建SAML应用程序。
2. 在中将外部 IdP 设置为身份提供者。IAM
3. 创建一个 I SAMLValidator dP 可以访问的 Lambda 函数（通过函数URL或网关）。API
4. 创建一个 GeneratePresignedUrl Lambda 函数和一个API网关来访问该函数。
5. 创建一个用户可以代入的IAM角色来调用API网关。此角色应按以下格式作为属性在SAML断言中传递：
   - 属性名称： https://aws.amazon.com/SAML/属性/角色
   - 属性值：<IdentityProviderARN> , <RoleARN>
6. 将SAML断言使用者服务 (ACS) 端点更新为SAMLValidator调用URL。

SAML验证器示例代码：

```
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
```

延伸阅读

- 在 AWS（AWS 博客）上设置安全、管理良好的机器学习环境
- 为团队和小组配置 SageMaker Amazon AI Studio，实现完全的资源隔离（AWS 博客）
- 使用 AWS SSO Okta 通用名录入亚马逊 SageMaker AI Studio（博客）AWS
- 如何为 AWS 账户联合配置 SAML 2.0（Okta 文档）
- 在 AWS上构建安全的企业机器学习平台（AWS 技术指南）
- 使用生命周期配置自定义 SageMaker Amazon AI Studio（AWS 博客）
- 将自己的自定义容器镜像带到 Amazon A SageMaker I Studio 笔记本上（AWS 博客）
- 构建自定义 SageMaker AI 项目模板-最佳实践（AWS 博客）
- 使用 Amazon A SageMaker I Pipelines 部署多账户模型（AWS 博客）
- 第 1 部分： NatWest 集团如何构建可扩展、安全和可持续的MLOps平台（AWS 博客）
- Secure Amazon SageMaker AI Studio 预签名第 1 URLs 部分：基础基础设施（博客）AWS

贡献者

本文档的贡献者包括：

- Ram Vittal，Amazon Web Services 的 ML Solutions Architect
- Sean Morgan，Amazon Web Services 的 ML Solutions Architect
- Durga Sury，Amazon Web Services 的 ML Solutions Architect

特此感谢以下人士贡献其创意、修订意见和观点：

- Alessandro Cerè，Amazon Web Services 的 AI/ML Solutions Architect
- Sumit Thakur，Amazon Web S SageMaker ervices 人工智能产品负责人
- Han Zhang，Amazon Web Services 的 Sr. Software Development Engineer
- Bhadrinath Pani，Amazon Web Services 的 Software Development Engineer

文档修订

如需获取有关本白皮书更新的通知，请订阅 RSS 源。

变更	说明	日期
已更新白皮书	已修复受损链接并进行大量编辑更改。	2023 年 4 月 25 日
初次发布	已发布白皮书。	2022 年 10 月 19 日

注意事项

客户有责任对本文档中的信息进行单独评测。本文档：（a）仅供参考，（b）代表当前的 AWS 产品和实践，如有更改，恕不另行通知，以及（c）不构成 AWS 及其附属公司、供应商或许可方的任何承诺或保证。AWS 产品或服务“按原样”提供，不附带任何明示或暗示的保证、陈述或条件。AWS 对其客户承担的责任和义务受 AWS 协议制约，本文档不是 AWS 与客户直接协议的一部分，也不构成对该协议的修改。

AWS 术语表

有关最新的 AWS 术语，请参阅《AWS 词汇表参考》中的 AWS 词汇表。

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。