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 Studio 管理最佳实践

发布日期：2023 年 4 月 25 日 (文档修订)

摘要

Amazon SageMaker Studio 提供基于 Web 的单一可视化界面，以便您执行所有机器学习 (ML) 开发步骤，进而提高数据科学团队的工作效率。SageMaker Studio 赋予您访问、控制并查看构建、训练与评估模型的每个必要步骤的充分权限。

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级 SageMaker 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，AWSWell-Architected Framework 有助于您了解所作决策的利弊。本框架的六个支柱有助于您了解设计并运行可靠、安全、高效、实惠且可持续发展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，亚马逊重点介绍了如何在 AWS Cloud 中设计、部署和并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

当您将 SageMaker Studio 作为机器学习平台进行管理时，需要根据最佳实践指南做出明智决策，从而随着工作负载的增加而扩展平台。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。
- 选择为用户身份信息设置 SageMaker Studio 域身份验证的方法，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。
- 您可以根据机器学习工作负载的敏感程度、用户数、实例类型、应用程序和已启动作业来规划虚拟私有云 (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。

- 考虑记录并监控各种应用程序编程接口 (API) 和用户活动的方法，以确保合规性。
- 使用您的映像和生命周期配置脚本，自定义 SageMaker 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) 模式，即所有业务部门加入该团队，共同完成数据科学项目。
- 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。

请先考虑适用于组织环境的运营模式和 AWS 最佳实践，再决定启动适用于生产用例的首个 Studio 域。有关更多信息，请参阅使用多个账户组织 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

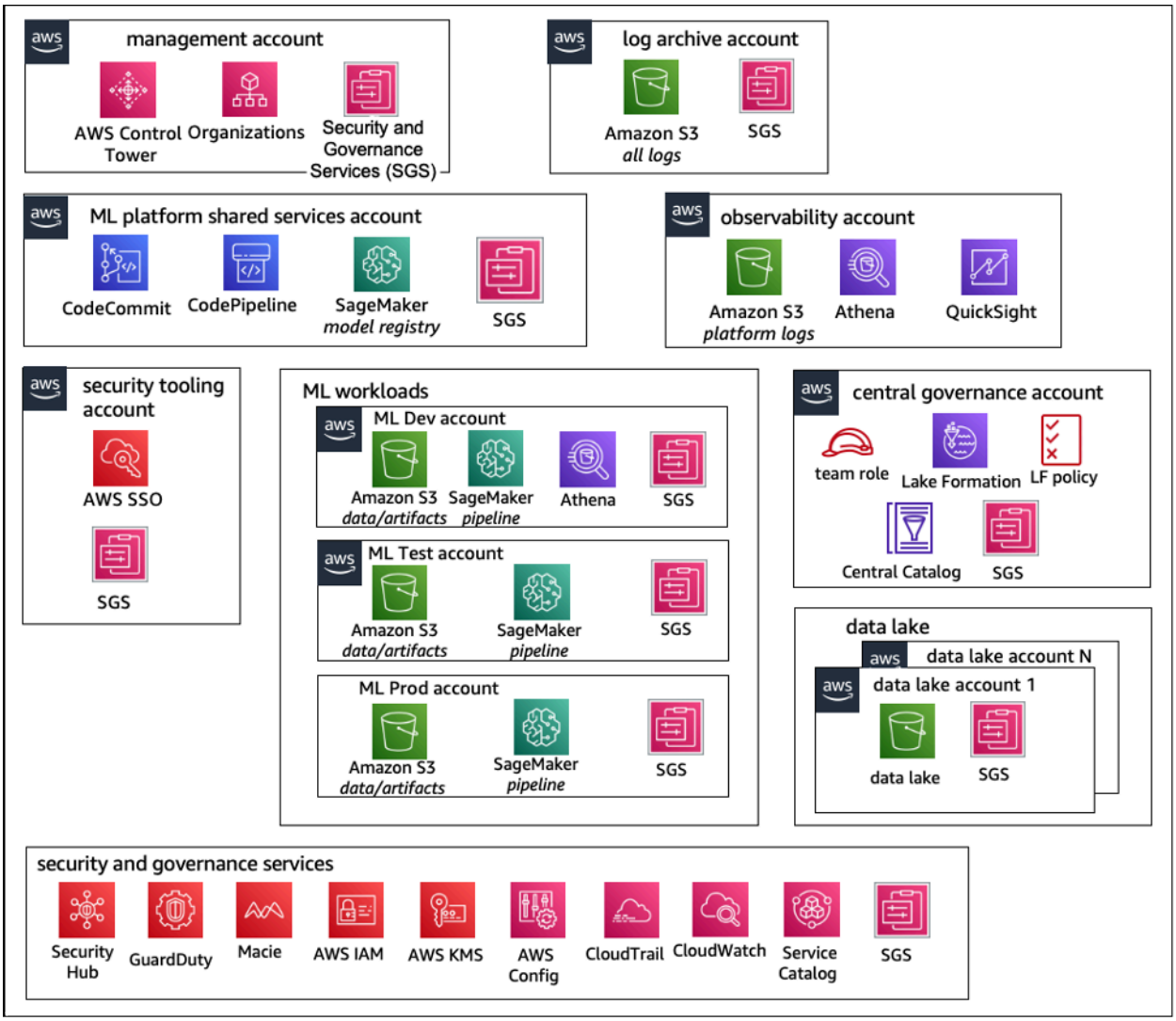
本节简要介绍了一种运营模式账户结构，以便您根据组织的运营要求初步应用并进行修改。无论您选择哪种运营模式，亚马逊都建议您实施以下常见的最佳实践：

- 使用 AWS Control Tower 设置、管理并监管账户。
- 使用身份提供者 (IdP) 和设有委派管理员 Securitiy Tooling 账户的 AWS IAM Identity Center，集中管理您的身份，并确保安全访问工作负载。
- 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。
- 根据组织和工作负载的要求，为每个账户嵌入具有适当预防性和检测性防护机制的安全和治理服务 (SGS)，确保其安全性和合规性。

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：

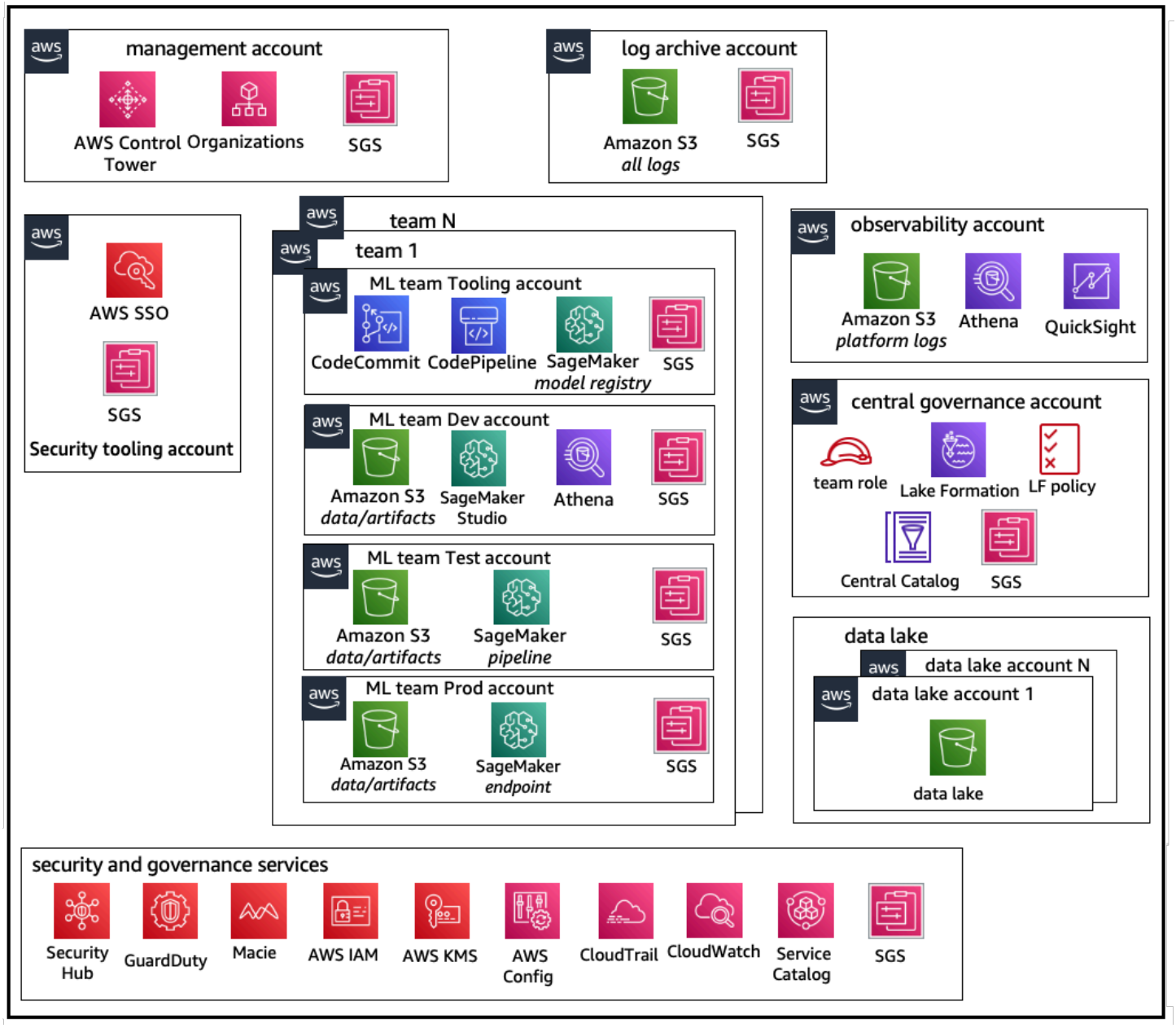
- 共享服务工具账户，可满足数据科学团队的所有机器学习操作 (MLOps) 要求。
- 跨数据科学团队共享账户，可开发、测试并生产机器学习工作负载。
- 监管策略，可确保独立运行各数据科学团队的工作负载。
- 常见的最佳实践。



集中式运营模式账户结构

分布式模型账户结构

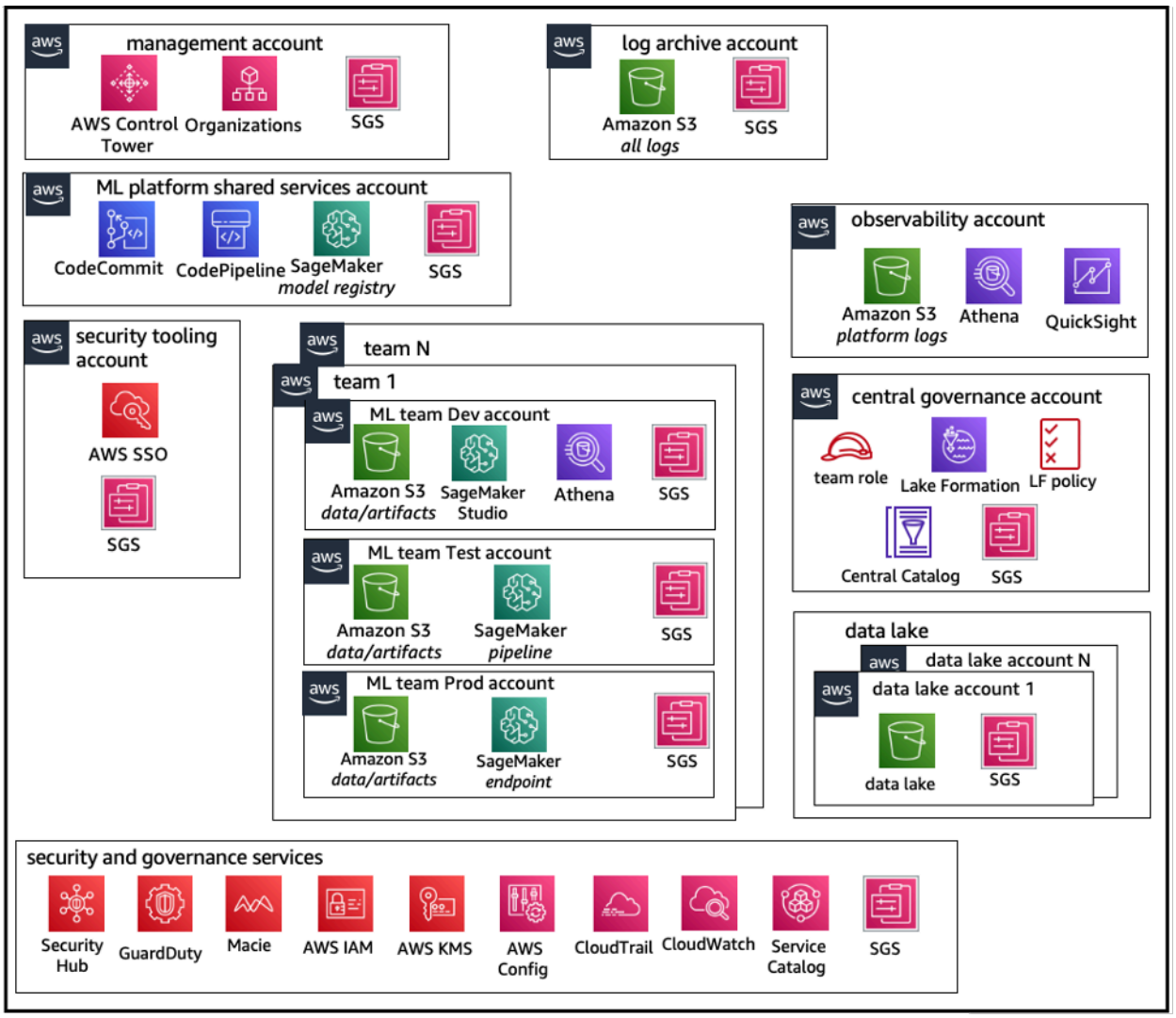
在此模型中，每个机器学习团队均独自负责预置、管理并治理机器学习账户和资源。亚马逊建议机器学习团队使用支持可观测性和数据治理的集中式模型，以简化数据治理和审计管理流程。



分布式运营模式账户结构

联合模型账户结构

此模型与集中式模型类似，关键区别在于，每个数据科学/机器学习团队都有一组独有的开发/测试/生产工作负载账户，能够有效地对机器学习资源进行物理隔离，还能让各团队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。

虽然单个 SageMaker Studio 实例（如 SageMaker 域）中也许能部署多个团队，但在多个团队代入单个 SageMaker Studio 域时，请权衡这些优势与爆炸半径、成本归属和账户级别限制等利弊。以下章节详细说明了这些利弊和最佳实践。

如需彻底隔离资源，可以考虑为不同账户中的每个租户都实施 SageMaker Studio 域。根据隔离要求，可实施多条业务线 (LOB)，作为单个账户和区域中的多个域。使用共享空间，在同一团队/业务线中的成员之间开展近实时的协作。您仍能使用 Identity Access Management (IAM) 策略和权限，确保实现多域资源隔离。

域中创建的 SageMaker 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。此文档介绍了关于多账户或多域策略使用时的详细信息和功能比较信息，并说明了为 GitHub 存储库上的现有域回填标记的脚本示例。

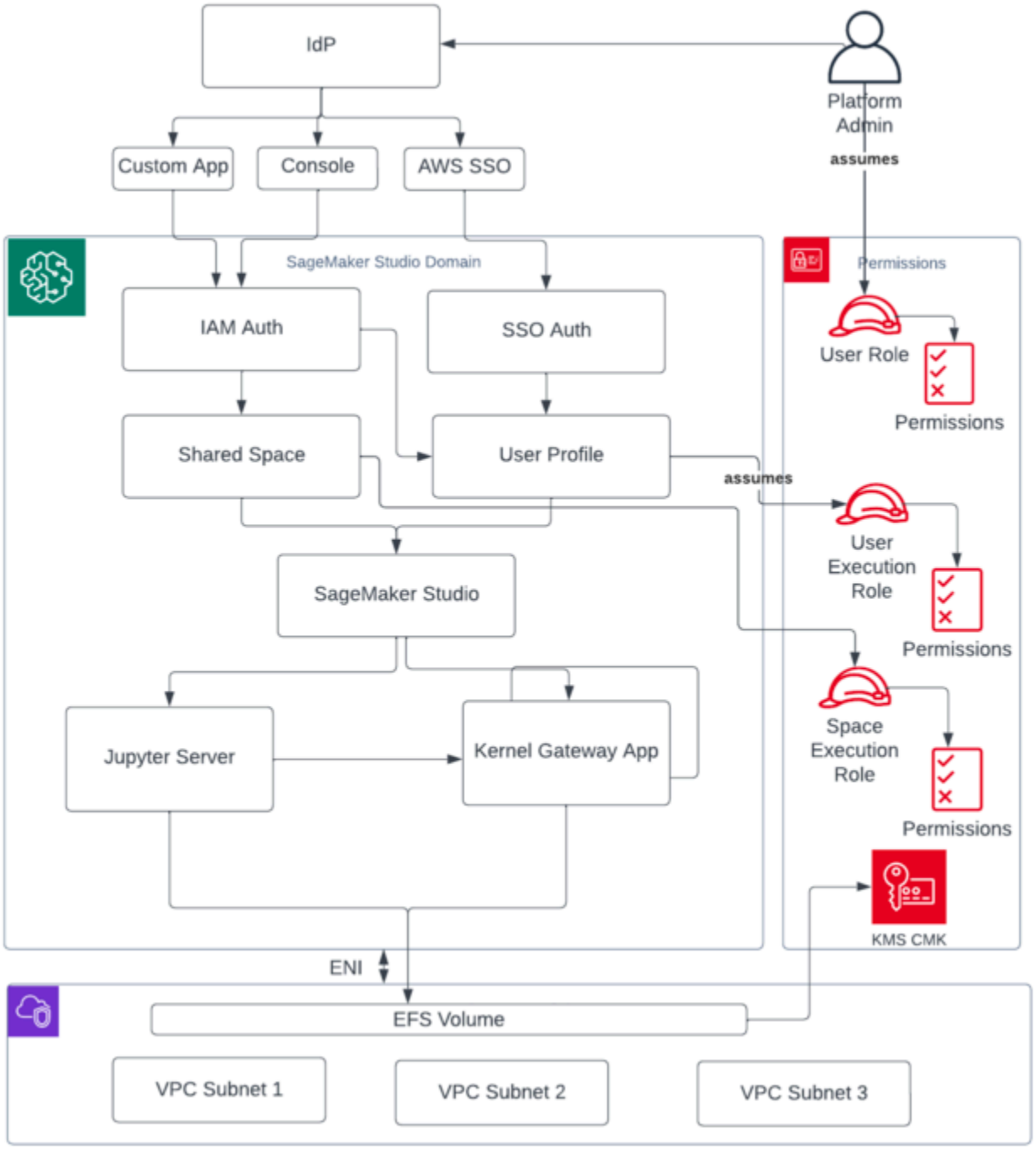
最后可用 AWS Service Catalog 将 SageMaker Studio 资源的自助部署到多个账户中。有关更多信息，请参阅在多个 AWS 账户 和 AWS 区域 中管理 AWS Service Catalog 产品。

域管理

Amazon SageMaker 域包括：

- 关联的 Amazon Elastic File System (Amazon EFS) 卷
- 授权用户列表
- 各种安全性、应用程序、策略和 Amazon Virtual Private Cloud (Amazon VPC) 配置

下图大致介绍了组成 SageMaker Studio 域的各种组件：

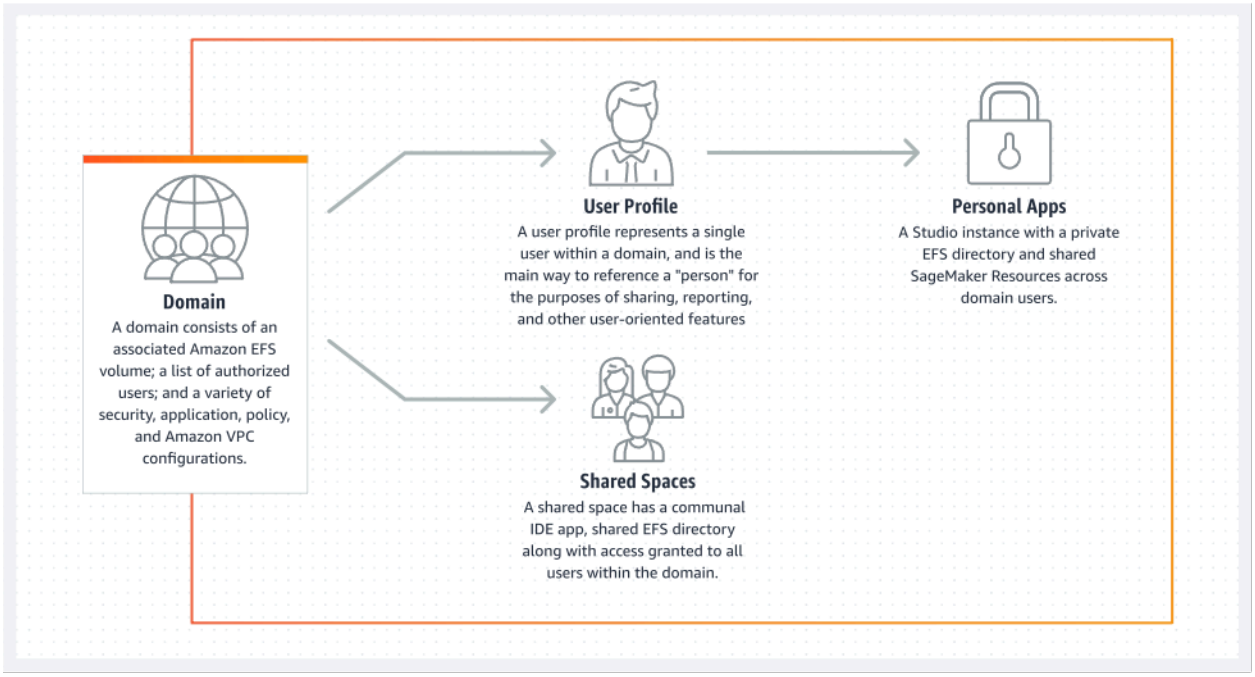


组成 SageMaker Studio 域的各种组件的总览

多域和共享空间

Amazon SageMaker 现在支持在单个AWS 区域账户中创建多个 SageMaker 域名。每个域都有专属的域设置（如身份验证模式）和联网设置（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所有域都可使用共享空间，实现用户间近实时的协作。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任务或项目创建的，其中单个域中的成员需要近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。

为进行 IAM 联合身份验证而设置域

在为 SageMaker Studio 域设置 AWS Identity and Access Management (IAM) 联合身份之前，您需要在 IdP 中设置 IAM 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 IAM 选项设置 SageMaker Studio 的详细说明，请参阅使用 IAM 身份中心加入亚马逊 SageMaker 域名。

为进行单点登录 (SSO) 联合身份验证而设置域

要使用单点登录 (SSO) 联合，您需要在需要运行 SageMaker Studio 的同一区域的AWS Organizations管理账户AWS IAM Identity Center中启用单点登录 (SSO)。域设置步骤与 IAM 联合身份验证步骤相似，但不包括在身份验证部分选择 AWS IAM Identity Center (IdC) 的情况。

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

SageMaker 工作室用户个人资料

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。该实体是在用户加载 toSageMaker 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。用户配置文件是个人用户设置的主要保存方式，引用了用户的 Amazon Elastic File System (Amazon EFS) 私有主目录。我们建议为 SageMaker Studio 应用程序的每位实际用户创建用户配置文件。每位用户在 Amazon EFS 上都有自己的专用目录，但不能在同一个账户中跨域共享用户配置文件。

每个共享 SageMaker 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 亚马逊弹性计算云 (Amazon EC2) 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 AWS IAM iDC 登录为用户启动 A mazon SageMaker Studio 笔记本时，Jupyter 服务器应用程序将在服务 SageMaker 托管 VPC 实例中启动。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 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 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 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 SageMaker Studio Image/Kernel Gateway 应用程序中启动和运行多个 Jupyter 笔记本内核、终端会话和交互式控制台。还可以在同一个物理实例上运行最多四个内核网关应用程序或映像，由实例对应的容器/映像一一隔离。

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 Studio 内置数据科学映像的简单笔记本电脑，也可以使用内置 TensorFlow 映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 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 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镜像和生命周期脚本之类的功能，将用户自己的软件包带到 SageMaker Studio，并在实例切换和新实例启动时保留它们。

Amazon Elastic File System 卷

创建域后，将创建单一的 Amazon Elastic File System (Amazon EFS) 卷，以供域中所有用户使用。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 卷中收到一个私有目录，可通过多个用户配置文件进行访问。对文件夹的访问由用户通过文件系统权限进行隔离。 SageMaker 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 EFS 上用户主目录的便携式操作系统接口 (POSIX) 用户/组 ID 应用，从而防止其他用户/空间访问其数据。

备份和恢复

现有 EFS 卷无法连接到新 SageMaker 域。请确保已将生产设置中的 Amazon EFS 卷进行备份 [备份到另一个 EFS 卷或 Amazon Simple Storage Service (Amazon S3)]。如果 EFS 卷被意外删除，管理员必须拆除并重新创建 SageMaker Studio 域。流程如下：

使用 `ListUserProfiles`、`DescribeUserProfile`、`List Spaces` 和 `DescribeSpace` API 调用，备份用户配置文件、空间和关联的 EFS 用户 ID (UID) 列表。

1. 创建一个新的 SageMaker Studio 域名。
2. 创建用户配置文件和空间。
3. 从 EFS/Amazon S3 上的备份文件中复制每个用户配置文件。
4. （可选）删除旧 SageMaker Studio 域中的所有应用程序和用户个人资料。

有关详细说明，请参阅附录部分 SageMaker 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。

Amazon EBS 卷

每个 Studio Notebook 实例上还附有一个亚马逊 Elastic Bloc k SageMaker Store (Amazon EBS) 存储卷。它将作为运行在实例上的容器或映像的根卷。Amazon EFS 存储虽然是永久的，但与容器挂载的 Amazon EBS 卷却是临时的。如果客户删除应用程序，则存储在 Amazon EBS 卷上的本地数据也会丢失。

保护对预签名 URL 的访问权限

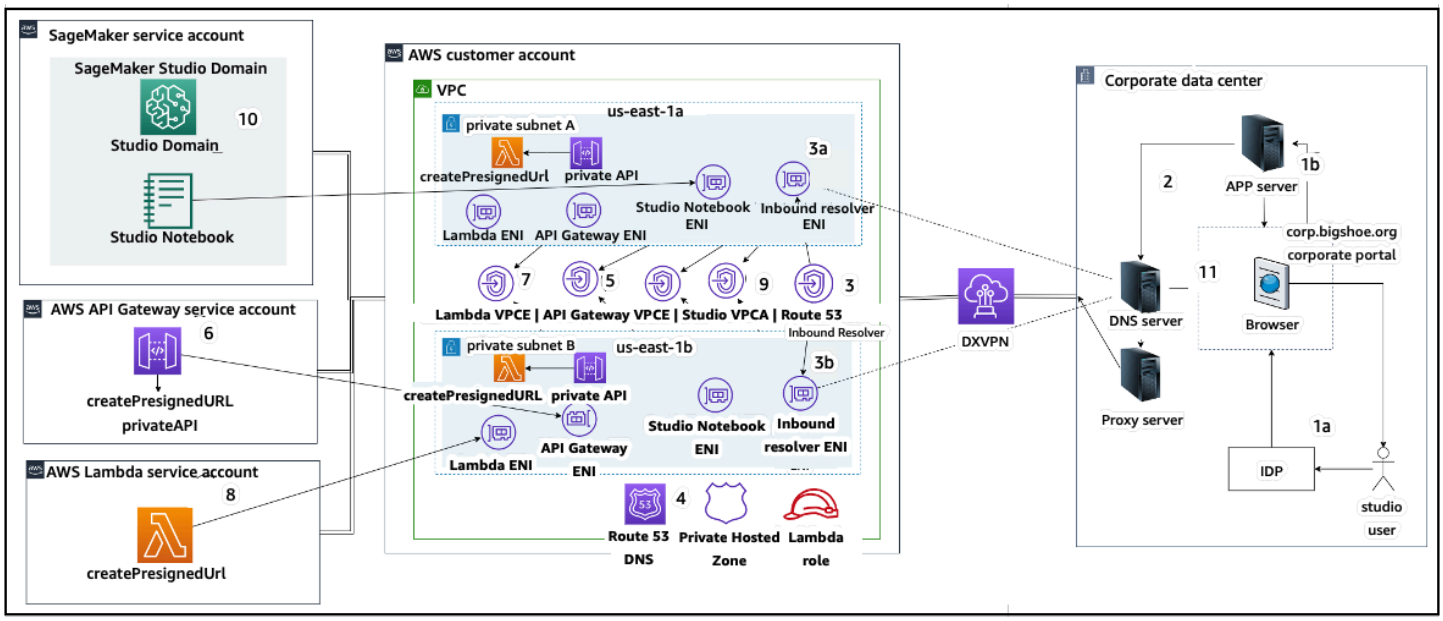
当 SageMaker Studio 用户打开笔记本链接时， SageMakerStudio 会验证联合用户的 IAM 策略以授权访问权限，并生成和解析该用户的预签名 URL。由于 SageMaker 控制台在 Internet 域上运行，因此此生成的预签名 URL 在浏览器会话中可见。这可能会导致发生数据失窃，以及在未采取适当措施的情况下被人获取客户数据。

针对预签名 URL 数据失窃问题，Studio 支持以下几种访问控制措施：

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

当您从 SageMaker控制台访问 SageMaker Studio 笔记本时，唯一可用的选项是将客户端 IP 验证与 IAM 策略条件结合使用`aws:sourceIp`。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 策略条件使用客户端 VPC 终端节点验证`aws:sourceVpce`，需要在部署 Studio 的同一客户 VPC 中创建预签名 URL，并且需要通过客户 VPC 上的 SageMaker Studio SageMaker VPC 终端节点解析预签名 URL。企业网络用户在访问时可以使用 DNS 转发规则（在 Zscaler 和企业 DNS 中）来解析预签名 URL，然后使用 Amazon Route 53 入站解析器解析客户 VPC 端点，其架构如下：



使用 VPC 端点在企业网络上访问 Studio 预签名 URL

有关设置上述架构的 step-by-step 指南，请参阅安全 Amazon SageMaker Studio 预签名 URL 第 1 部分：基础基础架构。

SageMaker 域名配额和限制

- SageMaker 仅在配置AWS身份中心的AWS组织的成员账户的区域中支持 Studio 域 SSO 联合。
- 目前，使用 AWS Identity Center 设置的域不支持共享空间。

- 创建域后无法更改 VPC 和子网配置。但可以使用不同的 VPC 和子网配置创建新域。
- 创建域后无法在 IAM 和 SSO 模式间切换域访问权限。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 Studio AWS 账户 并访问 SageMaker 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

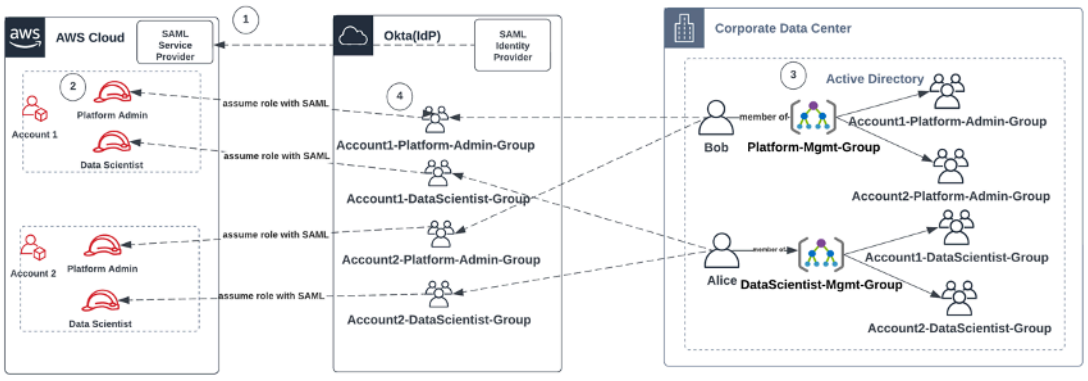
在中 AWS，使用用户、组和角色管理资源权限。客户可以使用 IAM 或者在由外部 IdP（如 Okta）启用的企业目录中 [如 Active Directory (AD)] 管理其用户与组。客户可使用该目录，对运行于云端和本地的各种应用程序进行用户身份验证。

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 允许最终用户使用带有安全断言标记语言 (SAML) 的 SSO 对一个或多个角色进行身份验证AWS 账户 并访问特定角色。IdP 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：



添加 AD 用户、AD 组和 IAM 角色的工作流

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 控制台）中的授权应用程序。

有关设置 SAML 账户联合的详细说明，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 Studio 的身份验证可以使用 IAM 或 IAM IdC 完成。由 IAM 管理的用户可选用 IAM 模式。使用外部 IdP 的企业可使用 IAM 或 IAM IdC 进行联合身份验证。请注意，现有 SageMaker Studio 域的身份验证模式无法更新，因此在创建正式版 SageMaker Studio 域之前做出决定至关重要。

如果 SageMaker Studio 设置为 IAM 模式， SageMaker Studio 用户将通过预签名 URL 访问应用程序，该网址在用户通过浏览器访问时会自动登录 SageMaker Studio 应用程序。

IAM 用户

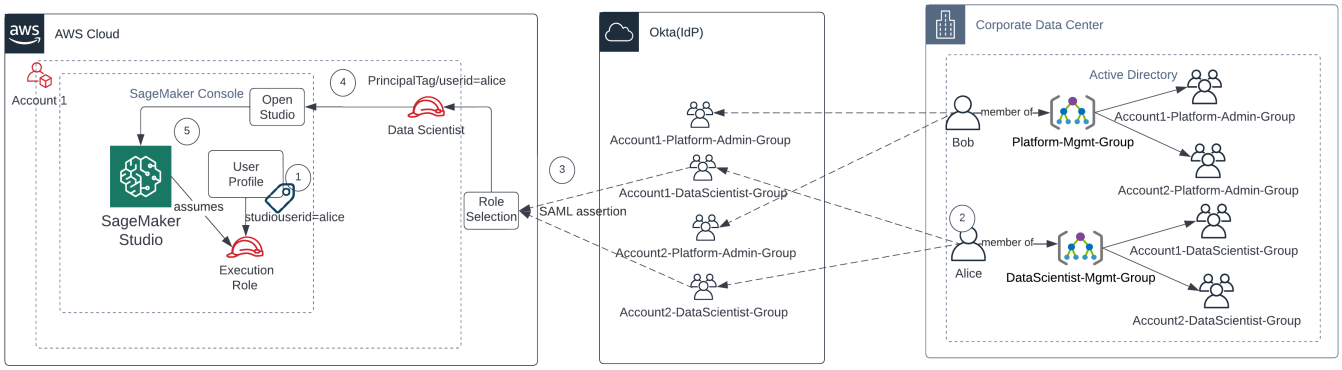
对于 IAM 用户，管理员为每个用户创建 SageMaker Studio 用户配置文件，并将用户配置文件与允许用户在 Studio 中执行必要操作的 IAM 角色关联起来。要限制 AWS 用户仅访问其 SageMaker Studio 用户个人资料，管理员应为 SageMaker Studio 用户配置文件添加标签，并向该用户附加一个 IAM 策略，允许他们仅在标签值与 AWS 用户名相同时才允许他们进行访问。此策略语句与以下内容类似：

```
{
    "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 SAML IdP（例如 Okta）联合进入 SageMaker 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 Studio 用户个人资料，添加 `PrincipalTags` IdP，然后将其设置为传递标记。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 Studio 用户个人资料。



在 IAM 联合模式下访问 SageMaker Studio

1. Alice SageMaker Studio 用户配置文件标有其用户 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 进行身份验证，并发布了具有两个角色（数据科学家账户 1 和数据科学家账户 2）的 SAML 断言，而 Alice 就是其中一个角色。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 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 主体标签将根据所选的 SageMaker 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 Studio 应用程序实例。

如果你想在用户入职过程中自动创建 SageMaker 执行角色和策略，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-Account1-Group`。
2. 当您需要让用户加入 St SageMaker udio 时，将-Account1-Group 添加到用户的群组成员资格中。SageMaker

设置监听`SageMaker-Account1-Group`成员资格事件的自动化流程，并使用 AWS API 根据其广告组成员资格创建角色、策略、标签和 SageMaker 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 Studio 用户访问其他用户配置文件。

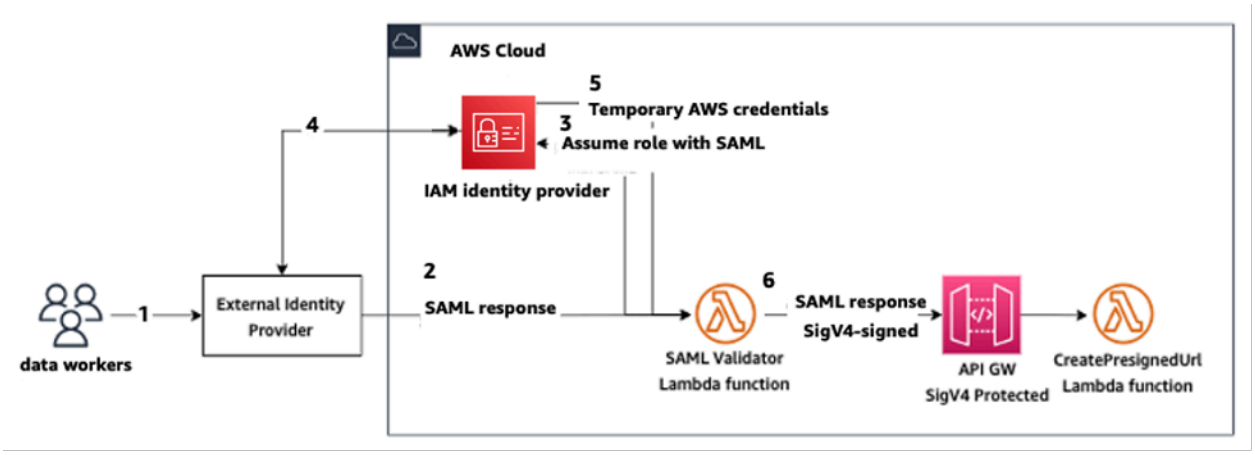
使用 SAML 身份验证 AWS Lambda

在 IAM 模式下，用户也可以使用 SAML 断言在 SageMaker 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）。已添加客户 IdP 到 IAM。 AWS Lambda 函数使用 IAM 和 STS 来帮助验证 SAML 断言，然后直接调用 API 网关或 Lambda 函数来创建预签名的域网址。

此解决方案的优势在于，Lambda 函数可以自定义访问 Studio 的逻辑。 SageMaker 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，向 SageMaker Studio 执行角色附加或删除角色或策略文档。
- 添加生命周期配置 (LCC) 和标签，以自定义用户配置文件。

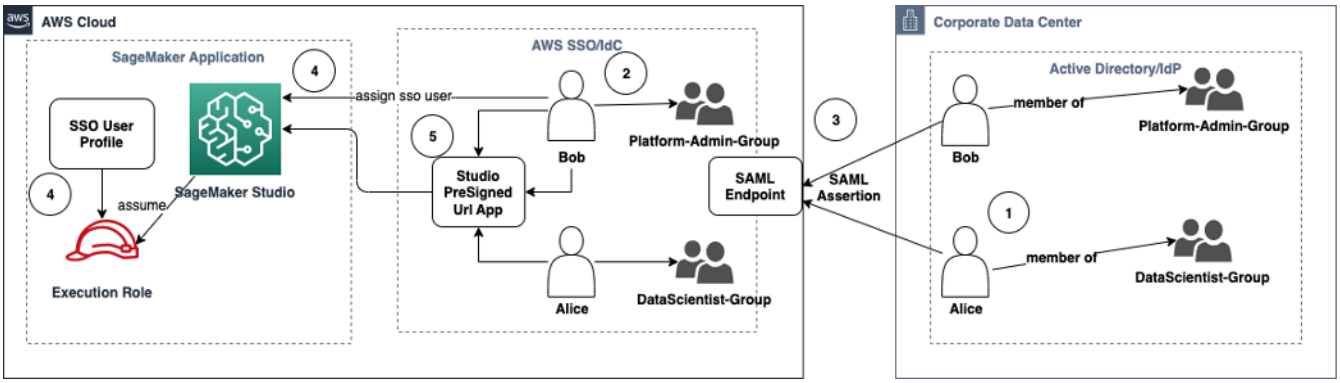
总而言之，此解决方案将 SageMaker Studio 作为一个 SAML2.0 应用程序公开，具有用于身份验证和授权的自定义逻辑。有关实现的详细信息，请参阅附录部分使用 SAML 断言访问 SageMaker Studio。



使用自定义 SAML 应用程序访问 SageMaker Studio

AWS IAM IdC 联合身份验证

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



在 IAM IdC 模式下访问 SageMaker 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 AWS IAM Identity Center，并分别作为单点登录用户和群组进行分配。
3. IdP 向 IdC SAML 端点发布 SAML 断言。 AWS
4. 在 SageMaker Studio 中，将 iDC 用户分配给 SageMaker Studio 应用程序。此任务可以使用 iDC 群组完成， SageMaker Studio 将应用于每个 iDC 用户级别。创建此任务后， SageMaker 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 作为云应用程序托管的安全预签名 URL 访问 SageMaker Studio 应用程序。SageMaker 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 Studio 用户界面，请在 AWS IAM iDc 中使用单点登录模式。
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 Studio 用户界面，则可以在后端使用 Lambda 函数为用户个人资料生成预签名 URL，然后将其重定向到 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 UpdateUserProfileAPI 更新用户配置文件。
5. 如果您想限制在 IAM 模式下（使用生成的预签名 URL）对 VPC 终端节点的 SageMaker Studio 界面访问权限，而不必穿越互联网，则可以使用自定义 DNS 解析器。请参阅 Secure Amazon SageMaker Studio 预签名 URL 第 1 部分：基础基础设施博客文章。

权限管理

本节讨论用于设置常用 IAM 角色、策略和防护机制的最佳实践，以实现 SageMaker Studio 域的预置和运行。

IAM 角色和策略

对于最佳实践，您需要先确定相关人员和应用程序（机器学习生命周期的参与主体），并确定需要授予的 AWS 权限。SageMaker 是托管服务，因此您还需要考虑服务主体，即能够代表用户调用 API 的 AWS 服务。下图说明了您希望创建的不同 IAM 角色，分别对应组织中的不同角色。

SageMaker IAM 角色

下文将详细介绍这些角色，并举例说明其所需的 IAM 特定权限。

- 机器学习管理员用户角色 — 该主体通过创建 Studio 域和用户配置文件（`sagemaker:CreateDomain`，`sagemaker:CreateUserProfile`）、用户适用的 AWS Key Management Service (AWS KMS) 密钥、数据科学家适用的 S3 存储桶和用于存放容器的 Amazon ECR 存储库，为数据科学家预置环境。此类角色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至 SageMaker Studio 域，以及提供 Service Catalog 产品，如自定义项目、Amazon EMR 模板。

  例如，因为主体不会运行训练作业，所以角色也无需启动 SageMaker 训练或处理作业的权限。如果角色使用基础设施即代码模板（如 CloudFormation 或 Terraform）预置域和用户，则预置服务会代入此角色，以管理员身份创建资源。此角色可通过 AWS Management Console 获取 SageMaker 只读访问权限。

  此用户角色还需要可在私有 VPC 中启动域的 EC2 特定权限，以加密 EFS 卷的 KMS 权限，以及为 Studio (`iam:CreateServiceLinkedRole`) 创建服务相关角色的权限。此类精细权限将在后文说明。

- 数据科学家用户角色 — 该主体用户可登录 SageMaker Studio、浏览数据、创建处理和训练作业与管道等。此用户主要需要启动 SageMaker Studio 的权限，其余策略可由 SageMaker 执行服务角色管理。
- SageMaker 执行服务角色 — SageMaker 托管服务可以代表用户启动作业。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。虽说这是一种能让客户在实践中熟练起来的轻松入门方法，但客户通常会将笔记本执行角色拆分为单独角色，用于执行不同的 API 操作，尤其是在已部署环境中运行此类作业时。

  您可以将创建好的角色关联到 SageMaker Studio 域。由于客户需要将不同角色灵活地关联到域中的不同用户配置文件（例如根据工作职能进行关联），您也可以将单独的 IAM 角色关联到每个用户配置文件。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。如果在创建角色时未将其挂载到用户配置文件，则默认将 SageMakerStudio 域执行角色关联到用户配置文件。

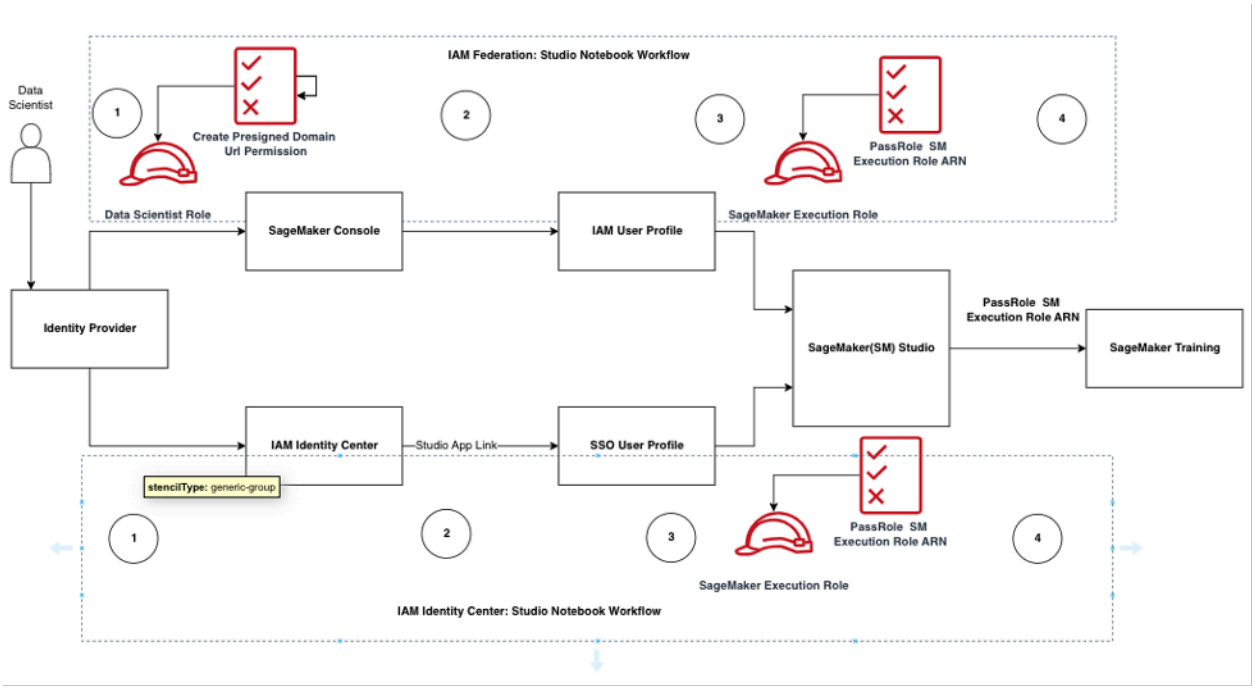
  如果由多个数据科学家和机器学习工程师共同处理一个项目，且需要使用共享权限模型来访问资源，则建议您创建团队级 SageMaker 服务执行角色，以便团队成员共享 IAM 权限。如需锁定各用户级别权限，您可以单独创建用户级 SageMaker 服务执行角色，但要注意个人服务限制。

SageMaker Studio 笔记本授权工作流

本节讨论如何将 SageMaker Studio 笔记本授权工作流应用于数据科学家希望执行的各项活动，从而在 SageMaker Studio 笔记本中直接构建并训练模型。SageMaker 域支持两种授权模式：

- IAM 联合身份验证
- IAM Identity Center 验证

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 联合身份验证：SageMaker Studio 笔记本工作流

1. 数据科学家对其企业身份提供者进行身份验证，并在 SageMaker 控制台中代入数据科学家的用户角色（用户联合身份验证角色）。此联合身份验证角色具有 SageMaker 执行角色的 `iam:PassRole` API 权限，可将角色的 Amazon 资源名称 (ARN) 传递给 SageMaker Studio。
2. 数据科学家从 Studio IAM 用户配置文件中选择了与 SageMaker 执行角色相关联的 Open Studio 链接。
3. SageMaker Studio IDE 服务已启动，具有用户配置文件的 SageMaker 执行角色权限。此角色具有 SageMaker 执行角色的 `iam:PassRole` API 权限，可将角色 ARN 传递给 SageMaker 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 执行角色 ARN 会传递至 SageMaker 训练服务。这样就能使用此 ARN 创建新角色会话，并运行训练作业。如需进一步缩小训练作业的权限范围，您可以创建训练特定角色，并在调用训练 API 时传递该角色 ARN。

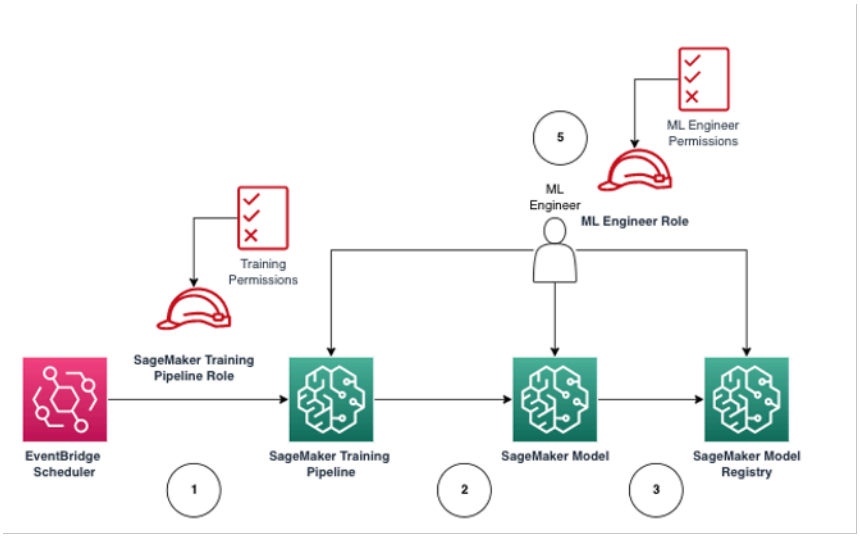
IAM Identity Center：SageMaker Studio 笔记本工作流

1. 数据科学家对其企业身份提供者进行身份验证，然后单击 AWS IAM Identity Center。数据科学家会看到 Identity Center 用户门户。

2. 数据科学家单击其 IdC 用户配置文件中创建的与 SageMaker 执行角色相关联的 SageMaker Studio 应用程序链接。
3. SageMaker Studio IDE 服务已启动，具有用户配置文件的 SageMaker 执行角色权限。此角色具有 SageMaker 执行角色的 `iam:PassRole` API 权限，可将角色 ARN 传递给 SageMaker 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 执行角色 ARN 会传递至 SageMaker 训练服务。执行角色 ARN 使用此 ARN 创建新角色会话，并运行训练作业。如需进一步缩小训练作业的权限范围，可以创建训练特定角色并在调用训练 API 时传递该角色 ARN。

已部署环境：SageMaker 训练工作流

系统测试和生产等已部署环境使用自动调度器和事件触发器运行作业，并使用 SageMaker Studio 笔记本限制人员访问这些环境。本节讨论 IAM 角色如何在已部署环境中与 SageMaker 训练管道搭配使用。



托管生产环境中的 SageMaker 训练工作流

1. Amazon EventBridge 调度器触发 SageMaker 训练管道作业。
2. SageMaker 训练管道作业代入 SageMaker 训练管道角色，以训练模型。
3. 经过训练的 SageMaker 模型注册到 SageMaker 模型注册表中。
4. 机器学习工程师代入对应的用户角色，管理训练管道和 SageMaker 模型。

数据权限

SageMaker Studio 用户的数据来源访问权限受制于 SageMaker IAM 执行角色的关联权限。附加策略授予用户权限，可读取、写入或删除特定的 Amazon S3 存储桶或前缀，并连接至 Amazon RDS 数据库。

访问 AWS Lake Formation 数据

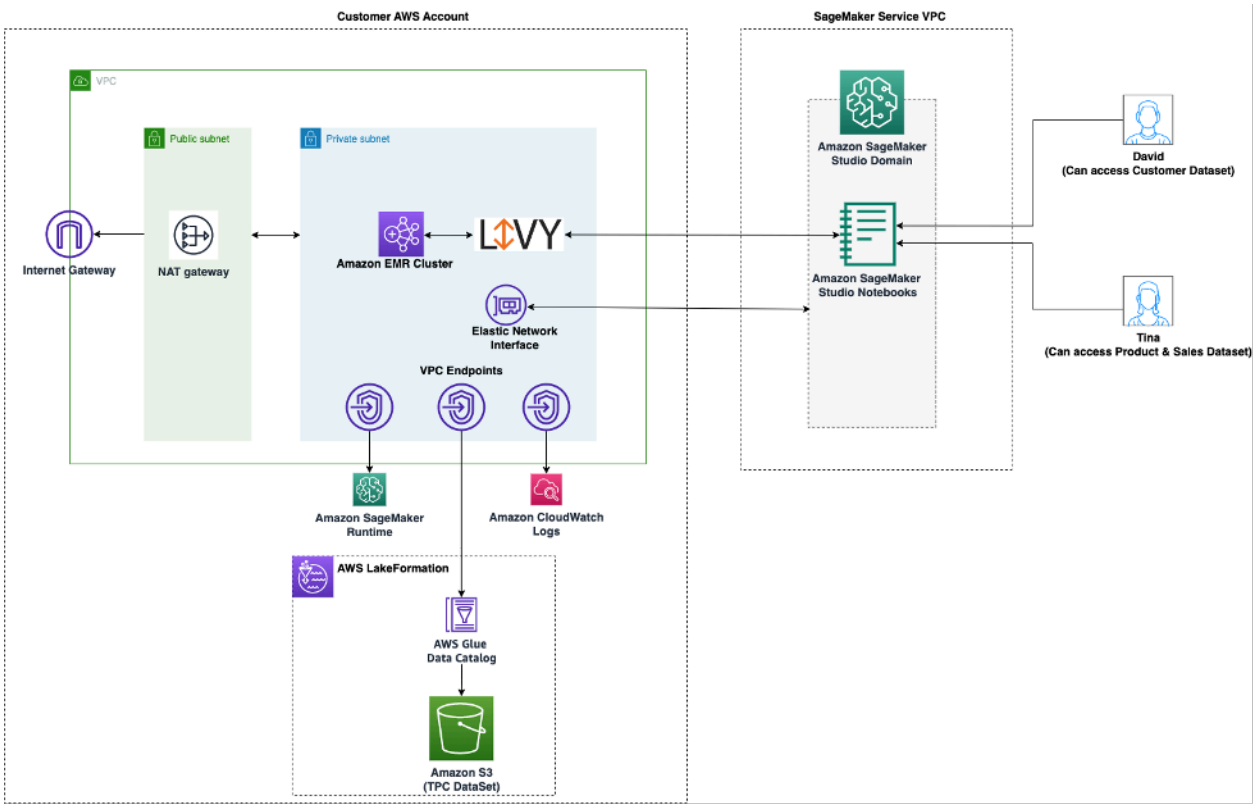
已经有不少企业开始使用受 AWS Lake Formation 监管的数据湖，为用户提供精细数据访问权限。以此类受管数据为例，管理员可为部分用户屏蔽敏感列，同时可查询同一基础表。

如需使用 SageMaker Studio 中的 Lake Formation，管理员可以将 SageMaker IAM 执行角色注册为 `DataLakePrincipals`。有关更多信息，请参阅 Lake Formation 权限参考。用户获得授权后，主要可采用三种方法访问并写入 SageMaker Studio 的受管数据：

1. 用户可使用 SageMaker Studio 笔记本中的查询引擎（如 Amazon Athena）或基于 Boto3 构建的库，直接将数据传输到笔记本。适用于 Pandas 的 AWS SDK（旧称为 awswrangler）就是一种常用的库。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 Studio 与 Amazon EMR 的本地连接，大规模读取并写入数据。SageMaker Studio 使用 Apache Livy 和 Amazon EMR 运行时系统角色构建本地连接，您可以将自己的 SageMaker 执行 IAM 角色（或其他授权角色）传递到 Amazon EMR 集群，用于数据访问和处理。有关最新说明，请参阅从 Studio 连接到 Amazon EMR 集群。



从 SageMaker Studio 访问 Lake Formation 管理数据的架构

3. 使用 SageMaker Studio 与 AWS Glue 交互式会话的本地连接，大规模读取并写入数据。SageMaker Studio 笔记本的内置内核允许用户在 AWS Glue 上交互运行命令。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。这些内核让用户能够传递其 SageMaker 执行角色或其他 IAM 授权角色。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

本节讨论在使用 IAM 策略、资源策略、VPC 端点策略和服务控制策略 (SCP) 治理机器学习资源时最常用的防护机制。

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，所有用户都要使用“系统”实例，创建托管 SageMaker Studio 的默认 Jupyter 服务器应用程序。

```
{
    "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 Studio 域

对于 SageMaker Studio 域，可使用以下服务控制策略来运行访问客户资源的流量，使其不必流经公共互联网，而是流经客户 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"
```

```
                }
            }
        }
    ]
}
```

限制对未经授权的 SageMaker 映像的启动权限

以下策略可防止用户在自有域内启动未经授权的 SageMaker 映像：

```
{
    "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 VPC 端点启动笔记本

除了适用于 SageMaker 控制面板的 VPC 端点外，SageMaker 还支持适用于用户的 VPC 端点，以连接 SageMaker Studio 笔记本或 SageMaker 笔记本实例。如果您已为 SageMaker Studio/笔记本实例设置 VPC 端点，则以下使用 SageMaker Studio VPC 端点或 SageMaker API 端点生成的 IAM 条件密钥只能连接 SageMaker Studio 笔记本。

```
{
    "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 Studio 笔记本访问权限控制在有限 IP 范围内

公司通常会将 SageMaker Studio 访问权限控制在经过允许的特定企业 IP 范围内。可使用以下具有 SourceIP 条件密钥的 IAM 策略进行限制。

```
{
    "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 Studio 用户访问其他用户配置文件

当您以管理员身份创建用户配置文件时，请确保使用带有 `studiouserid` 标签键的 SageMaker Studio 用户名标记文件。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）。

然后将以下策略附加到用户在启动 SageMaker Studio 时要代入的角色。

```
{
    "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 Studio 笔记本探索数据、构建并训练模型。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。

请确保已标记 SageMaker Studio 应用程序的用户配置文件。这些标签会自动从用户配置文件传播到应用程序。如需使用标签创建用户配置文件（由 CLI 和 SDK 提供支持），可考虑为管理员角色添加此策略：

```
{
    "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 Studio 中的根访问权限

SageMaker Studio 笔记本在 Docker 容器中运行，默认对该主机实例不享有根访问权限。同样，容器内的所有用户 ID 范围（默认运行身份用户除外）都将重新映射为主机实例上的非特权用户 ID。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。无论您是以根用户还是非根用户的身份创建用户，都要保证用户的 UID/GID 与 AppImageConfig 中自定义应用程序的 UID/GID 一致，为 SageMaker 创建可使用自定义映像运行应用程序的配置。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
```

则 AppImageConfig 文件需要在 KernelGatewayConfig 中提到相同的 UID 和 GID：

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

自定义映像和 Studio 映像的 UID/GID 可接受值分别为 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。

如需阻止用户进行篡改，请勿向 SageMaker Studio 笔记本用户授予 CreateAppImageConfig、UpdateAppImageConfig 或 DeleteAppImageConfig 权限。

网络管理

要设置 SageMaker Studio 域，您需要指定 VPC 网络、子网和安全组。在指定 VPC 和子网时，请确保先衡量以下章节讨论的使用量和预期增长再分配 IP。

VPC 网络规划

与 SageMaker 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 参与的AWS服务将弹性网络接口 (ENI) 注入客户 VPC 子网，用于以下用例：

- Amazon EFS 为该域的 EFS 挂载目标注入一个 ENI（每个子网/连接到该 SageMaker 域的可用区域各一个 IP）。 SageMaker
- SageMaker Studio 会为用户配置文件或共享空间使用的每个唯一实例注入一个 ENI。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一个默认 Jupyter 服务器应用程序（“系统”实例）、一个 Tensorflow GPU 应用程序（运行在 `ml.g4dn.xlarge` 实例上）和一个 Data Wrangler 应用程序（运行在 `ml.m5.4xlarge` 实例上），则 Studio 会注入三个 IP 地址。
- 为跨域 VPC 子网/可用区的每个 VPC 终端节点注入一个 ENI（VPC 终端节点有四个 IP；参与的服务 SageMaker VPC 终端节点（例如 S3、ECR 和. ）约六个 IP CloudWatch
- 如果使用相同的 VPC 配置启动 SageMaker 训练和处理任务，则每个任务需要每个实例两个 IP 地址。

> **Note**
>
> SageMaker Studio 的 VPC 设置（例如子网和仅限 VPC 的流量）不会自动传递到从 Studio 创建的训练/处理作业。 SageMaker 用户在调用 Create*Job API 时，可按需设定 VPC 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

在此场景中，假设 SageMaker 域设置为仅限 VPC 的流量模式。设置了 VPC 终端节点，例如 SageMaker API、 SageMaker 运行时、Amazon S3 和 Amazon 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 Studio 服务将按如下方式注入 ENI：

表 1 — 为实验场景注入客户 VPC 的弹性网络接口

实体	目标	已注入弹性网络接口	备注	级别
EFS 挂载目标	VPC 子网	三	三个可用区/子网	域
VPC 端点	VPC 子网	30	三个可用区/子网，各有 10 个 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	每个训练实例对应两个 IP	用户

实体	目标	已注入弹性网络接口	备注	级别
			如果使用 EFA，则每个训练实例有五个 IP	

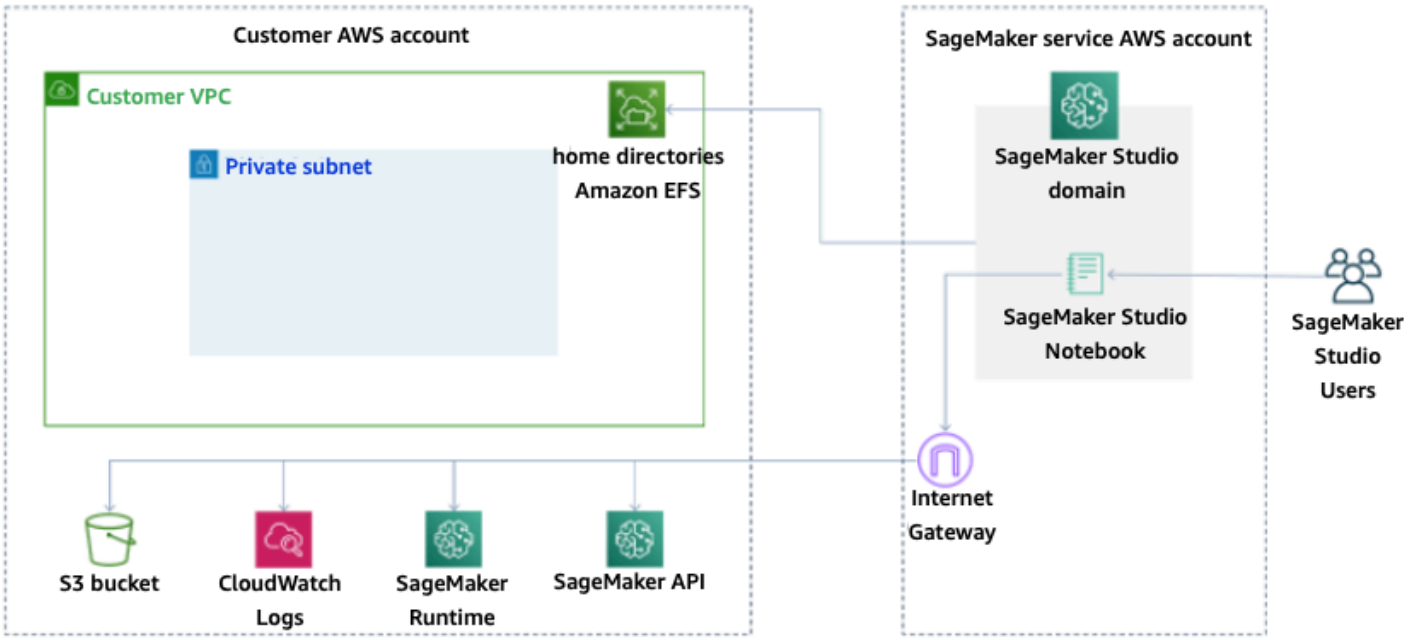
在此场景中，客户 VPC 共使用 38 个 IP，其中 33 个 IP 由域级别用户共享，其余 5 个 IP 由用户级别使用。如果域中有 100 位具有相似用户配置文件的用户同时执行这些活动，则除了在域级别上使用的 IP 数量（每个子网对应 11 个 IP）外，您还将在用户级别上使用 5 x 100 = 500 个 IP，即总共使用 511 个 IP。在此场景中，您需要创建 /22 范围的 VPC 子网 CIDR，用于分配 1024 个 IP 地址，并预留增长空间。

VPC 网络选项

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

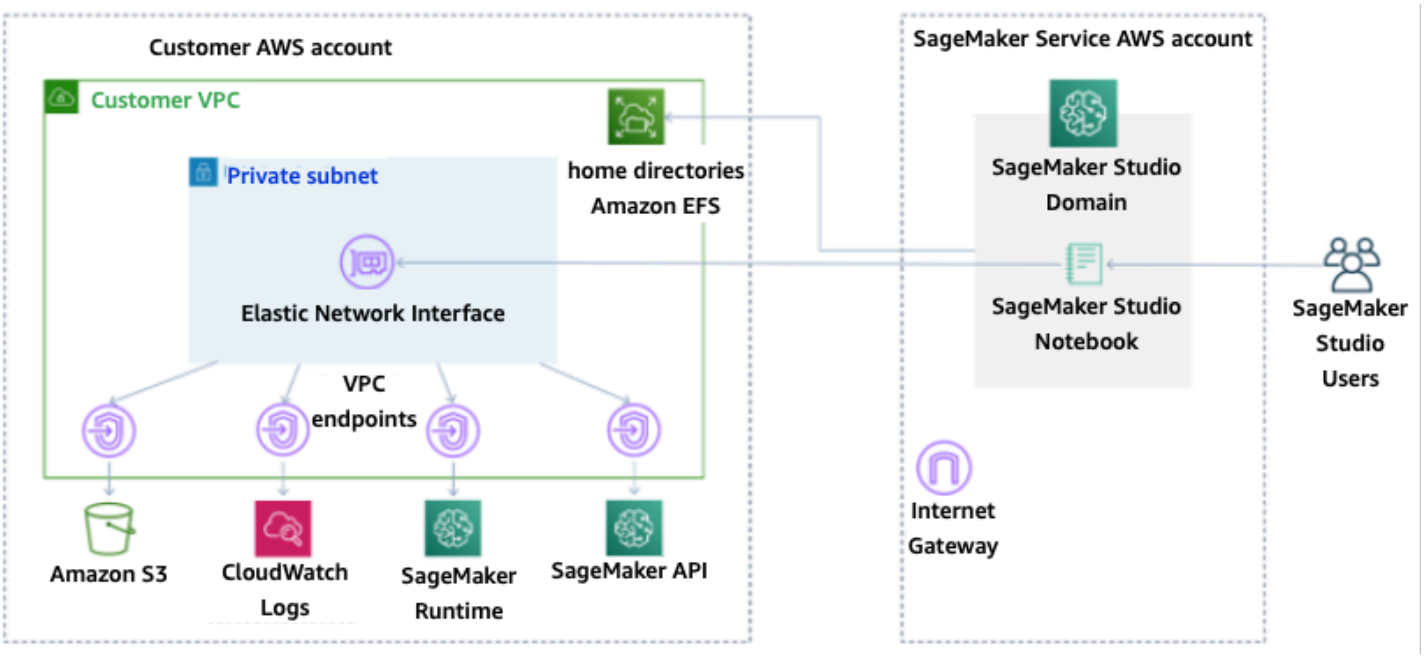
- 仅限公共互联网
- 仅限 VPC

仅限公共互联网选项允许 SageMaker API 服务通过 VPC 中配置的互联网网关使用公共互联网，该网关由 SageMaker服务账号管理，如下图所示：



默认模式：通过 SageMaker 服务账号访问互联网

仅限 VPC 选项禁用来自 SageMaker 服务账户管理的 VPC 的互联网路由，并允许客户将流量配置为通过 VPC 终端节点路由，如下图所示：



仅限 VPC 模式：无法通过 SageMaker 服务账号访问互联网

对于在仅限 VPC 模式下设置的域，请为每个用户配置文件设置对应的安全组，确保完全隔离底层实例。AWS 账户中的每个域都能设置独有的 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，请参阅将 VP C 中的 SageMaker Studio 笔记本连接到外部资源。

限制

- 创建 SageMaker Studio 域后，您无法将新子网与该域关联起来。
- 无法更改 VPC 网络类型（仅限公共互联网或仅限 VPC）。

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 Studio 提供了多种保护静态数据和传输中数据的功能。但正如 AWS 责任共担模式所述，客户有责任对托管在 AWS 全球基础设施上的内容加以管控。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

为保护您的 SageMaker Studio 笔记本以及模型构建数据和模型构件，SageMaker 对笔记本以及训练与批量转换作业的结果都进行了加密。SageMaker 默认使用适用于 Amazon S3 的 AWS 托管密钥进行加密。对于跨账户访问，此 Amazon S3 的 AWS 托管密钥无法共享。若要进行跨账户访问，则可在创建 SageMaker 资源时，指定您的客户托管密钥以共享密钥。

SageMaker Studio 可将数据存储到以下位置：

- S3 存储桶 — 启用可共享笔记本后，SageMaker Studio 会在 S3 存储桶中共享笔记本快照和元数据。
- EFS 卷 — SageMaker Studio 会在您的域中挂载 EFS 卷，用于存储笔记本和数据文件。删除域后，EFS 卷仍存在。
- EBS 卷 — 挂载 EBS 到笔记本运行实例。实例运行期间，此卷持续存在。

使用 AWS KMS 进行静态加密

- 您可以传递 AWS KMS 密钥，对挂载到笔记本、训练、优化、批量转换作业和端点的 EBS 卷进行加密。
- 若不指定 KMS 密钥，SageMaker 则会使用系统托管的 KMS 密钥，对操作系统 (OS) 卷和机器学习数据卷进行加密。
- 出于合规性原因，需要使用 KMS 密钥加密的敏感数据应该存储在机器学习存储卷或 Amazon S3 中，这两个位置均可使用您指定的 KMS 密钥进行加密。

保护传输中的数据

SageMaker Studio 确保机器学习模型构件和其他系统构件在传输过程中和静止状态下均加密。可通过安全 (SSL) 连接对 SageMaker API 和控制台发出请求。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。

Amazon SageMaker 默认在 Amazon VPC 中运行训练作业，为您的数据安全保驾护航。您可以再添一道安全保障，通过配置私有 VPC 来保护训练容器和数据。还可以配置 SageMaker Studio 域，使其仅在 VPC 模式下运行，并设置 VPC 端点，使其通过私有网络路由流量，不会通过互联网输出流量。

数据保护防护机制

加密 SageMaker 静态托管卷

在托管 SageMaker 端点期间，使用以下策略对在线推理执行加密：

```
{
   "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 SageMaker 端点的数据，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 SageMaker Studio 域存储卷

对挂载至 Studio 域的存储卷执行加密操作。本策略要求用户提供 CMK，对挂载至 Studio 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

本策略可加密存储桶中的任何数据，以便 SageMaker Studio 域中的用户共享笔记本：

```
{
    "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，您就无法使用自定义 AWS KMS 密钥更新挂载的 EFS 卷存储。
- 创建域后，您就无法使用 KMS 密钥更新训练/处理作业或端点配置。

日志记录和监控

为协助您调试编译作业、处理作业、训练作业、端点、转换作业、笔记本实例及其生命周期配置，亚马逊会将发送到 stdout 或 stderr 的算法容器、模型容器或笔记本实例生命周期配置的内容也发送到 Amazon CloudWatch Logs。您可以使用 Amazon CloudWatch 监控 SageMaker Studio，Amazon CloudWatch 会收集原始数据，并将数据处理为近实时的可读指标。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 CloudWatch 进行日志记录

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。SageMaker 默认发布指标到 CloudWatch Logs，并使用 AWS KMS 客户托管密钥加密这些日志。

您也可以在不使用公共互联网的情况下，通过 VPC 端点将日志发送到 CloudWatch。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 Amazon CloudWatch 用户指南。

SageMaker 在 `/aws/sagemaker/studio` 目录下创建了一个 Studio 日志组。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
```

训练/处理/转换作业 API 的调用者需要具备以下权限，才能让 SageMaker 以您的名义向 CloudWatch 发送日志：

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

如需使用自定义 AWS KMS 密钥加密这些日志，您首先要修改密钥策略，允许 CloudWatch 服务加密和解密密钥。创建用于日志加密的 AWS KMS 密钥后，请修改密钥策略以涵盖以下内容：

```
{
    "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，您可以随时使用 ArnEquals 并为希望加密的 CloudWatch 日志提供特定的 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。训练、处理和模型端点还会发布实例 CPU 和内存利用率、托管调用延迟等相关指标。您可以进一步配置 Amazon SNS，以便在超过特定阈值时向管理员发出事件通知。训练和处理 API 的使用者需要具备以下权限：

```
{
     "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 AWS CloudTrail 进行审计

如需提高合规性，请使用 AWS CloudTrail 对所有 API 进行审计。默认使用 AWS CloudTrail 对所有 SageMaker API 进行日志记录。启用 CloudTrail 无需其他任何 IAM 权限。

除 `InvokeEndpoint` 和 `InvokeEndpointAsync` 以外的所有 SageMaker 操作均由 CloudTrail 进行日志记录并载入操作。例如，CloudTrail 日志文件会为 `CreateTrainingJob`、`CreateEndpoint` 和 `CreateNotebookInstance` 的调用操作生成条目。

每个 CloudTrail 事件条目中都有请求生成者的相关信息。身份信息可帮助您确定以下内容：

- 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。
- 请求是否由其它 AWS 服务发出。有关示例事件，请参阅使用 CloudTrail 记录 SageMaker API 调用文档。

CloudTrail 默认将用户配置文件的 Studio 执行角色名称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。如果有多个用户共享一个执行角色，则可以使用 `sourceIdentity` 配置，将 Studio 用户配置文件名称传播到 CloudTrail。如需启用 `sourceIdentity` 功能，请参阅监控 Amazon SageMaker Studio 中的用户资源访问权限。共享空间中的所有操作都将空间 ARN 作为参考源，您无法使用 `sourceIdentity` 进行审计。

成本归属

管理员可利用 SageMaker Studio 的内置功能跟踪个人域、共享空间和用户方面的支出。

自动标记

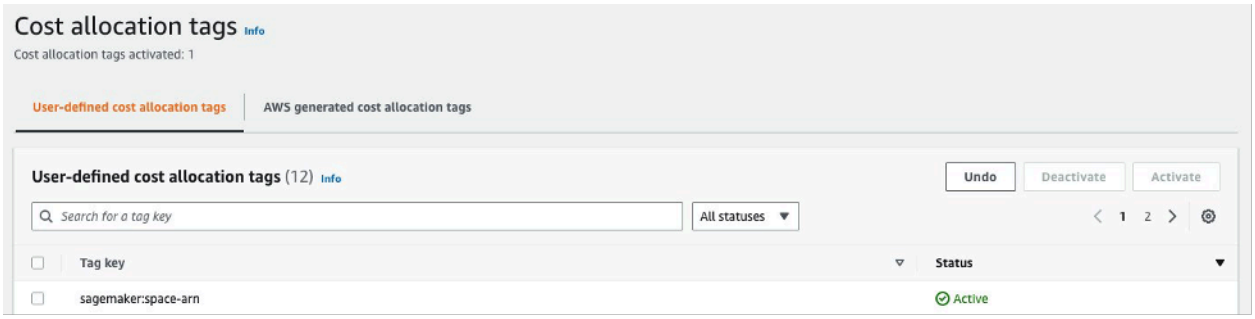
对于 SageMaker 的新资源（例如训练作业、处理作业与内核应用程序），SageMaker Studio 目前使用对应的 `sagemaker:domain-arn` 进行自动标记。对于更精细的资源，SageMaker 还会使用 `sagemaker:user-profile-arn` 或 `sagemaker:space-arn` 进行标记，以指定资源的主要创建者。

SageMaker 域中的 EFS 卷均使用具有域 ARN 值的 `ManagedByAmazonSageMakerResource` 密钥进行标记。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 卷挂载到监控专用的 EC2 实例。

成本监控

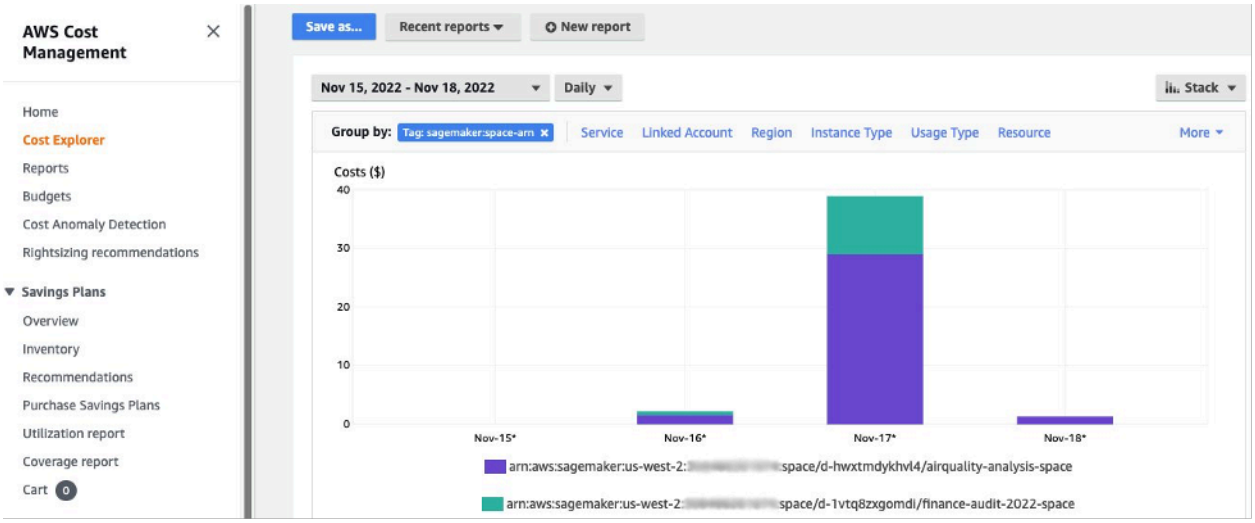
自动标签方便管理员采用创新的解决方案（例如 AWS Cost Explorer 和 AWS Budgets）和基于 AWS 成本和使用情况报告 (CUR) 的数据构建的自定义解决方案，跟踪、报告并监控机器学习支出。

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。成本分配标签面板最晚要在 24 小时后才显示标签，因此您需要先创建 SageMaker 资源再启用标签。



Cost Explorer 上作为成本分配标签启用的空间 ARN

AWS 将在您启用成本分配标签后开始跟踪标记的资源。这些标签将在 24 - 48 小时后，作为可选择的筛选条件显示在 Cost Explorer 中。



按照示例域的共享空间分组的成本

成本控制

当第一位 SageMaker Studio 用户登录时， SageMaker 会为域创建 EFS 卷。由于笔记本和数据文件均存储于用户主目录，所以 EFS 卷会产生存储成本。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。有关费用明细，请参阅 Amazon SageMaker 定价情况。

管理员使用通用防护机制一节所述的 IAM 策略指定用户可启动的实例列表，即可控制计算成本。此外，亚马逊建议客户使用 SageMaker Studio 的自动关闭扩展程序，可自动关闭空闲的应用程序以节约成本。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 Studio 生命周期事件（例如启动新的 SageMaker Studio 笔记本）启动的 Shell 脚本。您可以使用这些 Shell 脚本自动对 SageMaker Studio 环境进行自定义设置，例如安装自定义包、安装可自动关闭非活动状态的笔记本应用程序的 Jupyter 扩展程序，以及设定 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 Studio。

适用于 SageMaker Studio 笔记本的自定义映像

Studio 笔记本附带一组预构建映像，其中包括 Amazon SageMaker Python SDK 和最新版本的 IPython 运行时系统或内核。此功能可以将您的自定义映像引入 Amazon SageMaker 笔记本。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：

- 访问常用机器学习框架的特定或最新版本，例如 TensorFlow、MXNet、PyTorch 等框架。
- 引入本地开发的自定义代码或算法到 SageMaker Studio 笔记本，实现快速迭代与模型训练。
- 通过 API 访问数据湖或本地数据存储。管理员需要为映像加入相应的驱动程序。
- 访问 IPython 以外的后端运行时系统（也称为内核）（例如 R、Julia 等系统）。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映像的详细说明，请参阅创建 SageMaker 自定义映像。

JupyterLab 扩展程序

借助 SageMaker Studio JuypterLab 3 Notebook，您可以利用日益壮大的 JupyterLab 开源扩展程序的社区。本节重点介绍了一些与 SageMaker 开发者工作流程自然融合的扩展程序，但亚马逊鼓励您浏览可用的扩展程序，甚至是自己创建扩展程序。

目前，JupyterLab 3 大大简化了扩展程序的打包和安装流程。您可以使用 Bash 脚本安装上述扩展程序。例如，在 SageMaker Studio 中打开 Studio 启动器中的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
```

有关扩展程序管理的更多信息（包括如何为实现向后兼容目的编写适用于第 1 版和第 3 版 JupyterLab 笔记本的生命周期配置），请参阅安装 JupyterLab 和 Jupyter 服务器扩展程序。

Git 存储库

SageMaker Studio 预装了 Jupyter Git 扩展程序，以便用户输入 Git 存储库专属 URL、将其克隆到个人 EFS 目录、推送更改内容，并查看提交历史记录。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。有关最新说明，请参阅将建议的 Git 存储库挂载至 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。或者要求用户将 SSH 凭证存储在个人 EFS 目录上，以便管理。

Conda 环境

SageMaker Studio 笔记本使用 Amazon 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 EFS 提供支持，并且在重启内核、应用程序或 Studio 期间持续生效。Studio 会自动将一切有效环境作为 KernelGateway 内核。

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，请参阅管理 Amazon SageMaker Studio 笔记本中 Python 软件包的四种方法中提到的在 Studio EFS 卷中永久保存 Conda 环境一节。

结论

本白皮书回顾了运营模式、域管理、身份管理、权限管理、网络管理、日志记录、监控和自定义等领域的多种最佳实践，便于平台管理员设置并管理 SageMaker Studio 平台。

附录

多租户比较

表 2 — 多租户比较

多域	多账户	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 Studio 会自动使用域 ARN 和用户配置文件/空间 ARN 标记所有资源。	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 ABAC 创建的资源的标记。
列表 API 不能通过标签进行限制。资源的 UI 筛选是在共享空间上完成的，但是，通过 AWS CLI 或 Boto3 SDK 进行的 List API 调用将列出整个地区的资源。	列表 API 隔离也是可能的，因为租户位于他们的专用账户中。	列表 API 不能通过标签进行限制。列出通过 AWS CLI 或 Boto3 SDK 进行的 API 调用将列出该地区的资源。
SageMaker 使用域 ARN 作为成本分配标签，可以轻松监控每个租户的 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 Studio 计算和存储成本。	SageMaker Studio 每个租户的计算成本需要使用自定义标签进行计算。 SageMaker 由于所有租户共享相同的 EFS 卷，因此无法按域监控 Studio 存储成本。
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，并且需要用租户名称手动标记用户配置文件。

多域	多账户	单个域内基于属性的访问控制 (ABAC)
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 Studio 域名备份和恢复

如果误删 EFS 或由于联网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：使用 EC2 从现有 EFS 进行备份

SageMaker 工作室域名备份

1. 在 SageMaker Studio（CLI、SDK）中列出用户个人资料和空间。
2. 映射用户配置文件/空间到 EFS 上的 UID。
   a. 为用户/空间列表中的每位用户描述用户配置文件/空间（CLI，SDK）。
   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。
   - 使用执行角色映射，为用户列表中的每位用户创建相应的配置文件（CLI，SDK）。
5. 为新 EFS 和 UID 创建映射。
   a. 为用户列表中的每位用户描述用户配置文件（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

如需进行 EFS 备份，请按照以下说明操作：

1. 启动 EC2 实例，并将旧 SageMaker Studio 域的入站/出站安全组附加到新的 EC2 实例（允许端口 2049 上通过 TCP 进行 NFS 流量）。请参阅将 VPC 中的 SageMaker Studio 笔记本电脑连接到外部资源。
2. 将 SageMaker Studio EFS 卷挂载到新的 EC2 实例。请参阅挂载 EFS 文件系统。
3. 复制文件到 EBS 本地存储：`>sudo cp -rp /efs /studio-backup:`
   a. 为 EC2 实例附加新域安全组。
   b. 挂载 EFS 新卷到 EC2 实例。
   c. 复制文件到 EFS 新卷。
   d. 对于用户集合中的每位用户：
      i. 创建目录：`mkdir new_uid`。
      ii. 将 UID 旧目录中的文件复制到新目录。
      iii. 更改所有文件的所有权：所有文件的 `chown <new_UID>`。

选项 2：使用 Amazon S3 和生命周期配置，从现有 EFS 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 LCC 中，将 `/home/sagemaker-user` 中的所有内容复制到 S3 存储桶中的用户配置文件前缀（例如 `s3://studio-backup/studio-user1`）。
5. 重启所有默认 Jupyter 服务器应用程序（以运行 LCC）。
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. 设置域级别的 LCC：
   - 在 LCC 中，将 S3 存储桶的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，SDK）为所有用户创建默认 Jupyter 服务器应用程序。

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. 将外部 IdP 设为 IAM 中的身份提供者。
3. 创建可通过 IdP 访问的 `SAMLValidator` Lambda 函数（通过函数 URL 或 API Gateway）。
4. 创建 `GeneratePresignedUrl` Lambda 函数和 API Gateway 以访问此函数。
5. 创建用户可代入的 IAM 角色，以调用 API Gateway。此角色应作为一种属性在 SAML 断言中传递，格式如下：
   - 属性名称：https://aws.amazon.com/SAML/Attributes/Role
   - 属性值：`<IdentityProviderARN>`，`<RoleARN>`
6. 更新 SAML 断言使用者服务 (ACS) 端点到 `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 上搭建安全且管理得当的机器学习环境（AWS 博客）
- 使用完善资源隔离功能为团队和小组配置 Amazon SageMaker Studio（AWS 博客）
- 使用 AWS SSO 和 Okta 通用目录登录 Amazon SageMaker Studio（AWS 博客）
- 如何为 AWS 账户联合身份验证配置 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 Amazon SageMaker Studio（AWS 博客）
- 将您的自定义容器映像引入 Amazon SageMaker Studio 笔记本（AWS 博客）
- 构建 SageMaker 自定义项目模板—最佳实践（AWS 博客）
- 使用 Amazon SageMaker 管道部署多账户模型（AWS 博客）
- 第 1 部分：NatWest Group 如何构建具有可扩展性与可持续性的安全 MLOps 平台（AWS 博客）
- 保护 Amazon SageMaker Studio 的预签名 URL 第 1 部分：基础设施（AWS 博客）

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 Services 的 SageMaker Product Leader
- 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。