AWS 白皮書

SageMaker Studio 管理最佳實務

SageMaker Studio 管理最佳實務: AWS 白皮書

Table of Contents

SageMaker Studio 管理最佳實務

發佈日期：2023 年 4 月 25 日 (文件修訂)

摘要

Amazon SageMaker AI Studio 提供單一的 Web 型視覺化界面，您可以在其中執行所有機器學習 (ML) 開發步驟，進而改善資料科學團隊的生產力。 SageMaker AI Studio 可讓您完全存取、控制和查看建構、訓練和評估模型所需的每個步驟。

在本白皮書中，我們會討論主題的最佳實務，包括操作模型、網域管理、身分管理、許可管理、網路管理、記錄、監控和自訂。此處討論的最佳實務適用於企業 SageMaker AI Studio 部署，包括多租戶部署。本文件適用於 ML 平台管理員、ML 工程師和 ML 架構師。

您是 Well-Architected 嗎？

AWS Well-Architected 架構可協助您了解在雲端建置系統時所做決策的優缺點。架構的六個支柱可讓您了解架構最佳實務，以設計和操作可靠、安全、高效、經濟實惠且永續的系統。使用 AWS Well-Architected Tool免費提供的 AWS 管理主控台，您可以透過回答每個支柱的一組問題，根據這些最佳實務來檢閱工作負載。

在 Machine Learning Lens 中，我們著重於如何在 中設計、部署和架構機器學習工作負載 AWS 雲端。此鏡頭會新增至 Well-Architected Framework 中所述的最佳實務。

簡介

當您將 SageMaker AI Studio 管理為 ML 平台時，您需要最佳實務指導，以做出明智的決策，協助您隨著工作負載的成長擴展 ML 平台。如需佈建、操作化和擴展 ML 平台，請考慮下列事項：

- 選擇正確的操作模型，並組織您的 ML 環境以符合您的業務目標。
- 選擇如何設定使用者身分的 SageMaker AI Studio 網域身分驗證，並考慮網域層級限制。
- 決定如何將使用者的身分和授權聯合到 ML 平台，以進行精細的存取控制和稽核。
- 請考慮為 ML 角色的各種角色設定許可和護欄。
- 考量 ML 工作負載的敏感度、使用者數量、執行個體類型、應用程式和啟動的任務，規劃您的虛擬私有雲端 (VPC) 網路拓撲。
- 使用加密來分類和保護您的靜態資料和傳輸中資料。

- 考慮如何記錄和監控各種應用程式程式設計介面 (APIs) 和使用者活動，以確保合規性。
- 使用您自己的映像和生命週期組態指令碼自訂 SageMaker AI Studio 筆記本體驗。

操作模型

操作模型是一種架構，可將人員、程序和技術結合在一起，以協助組織以可擴展、一致、有效率的方式提供商業價值。ML 操作模型為整個組織的團隊提供標準產品開發程序。根據大小、複雜性和業務驅動因素，實作操作模型有三種模型：

- 集中式資料科學團隊：在此模型中，所有資料科學活動都會集中在單一團隊或組織中。這與卓越中心 (COE) 模型類似，其中所有業務單位都會前往此團隊進行資料科學專案。
- 分散式資料科學團隊：在此模型中，資料科學活動會分散到不同的業務職能或部門，或根據不同的產品線。
- 聯合資料科學團隊 — 在此模型中，程式碼儲存庫、持續整合和持續交付 (CI/CD) 管道等共用服務函數由集中式團隊管理，而每個業務單位或產品層級函數則由分散團隊管理。這類似於中樞和發言模型，每個業務單位都有自己的資料科學團隊；但是，這些業務單位團隊會與集中式團隊協調其活動。

在決定為生產使用案例啟動第一個 Studio 網域之前，請考慮您的操作模型和 AWS 最佳實務來組織您的環境。如需詳細資訊，請參閱使用多個帳戶組織您的 AWS 環境。

下一節提供針對每個操作模型組織帳戶結構的指引。

建議的帳戶結構

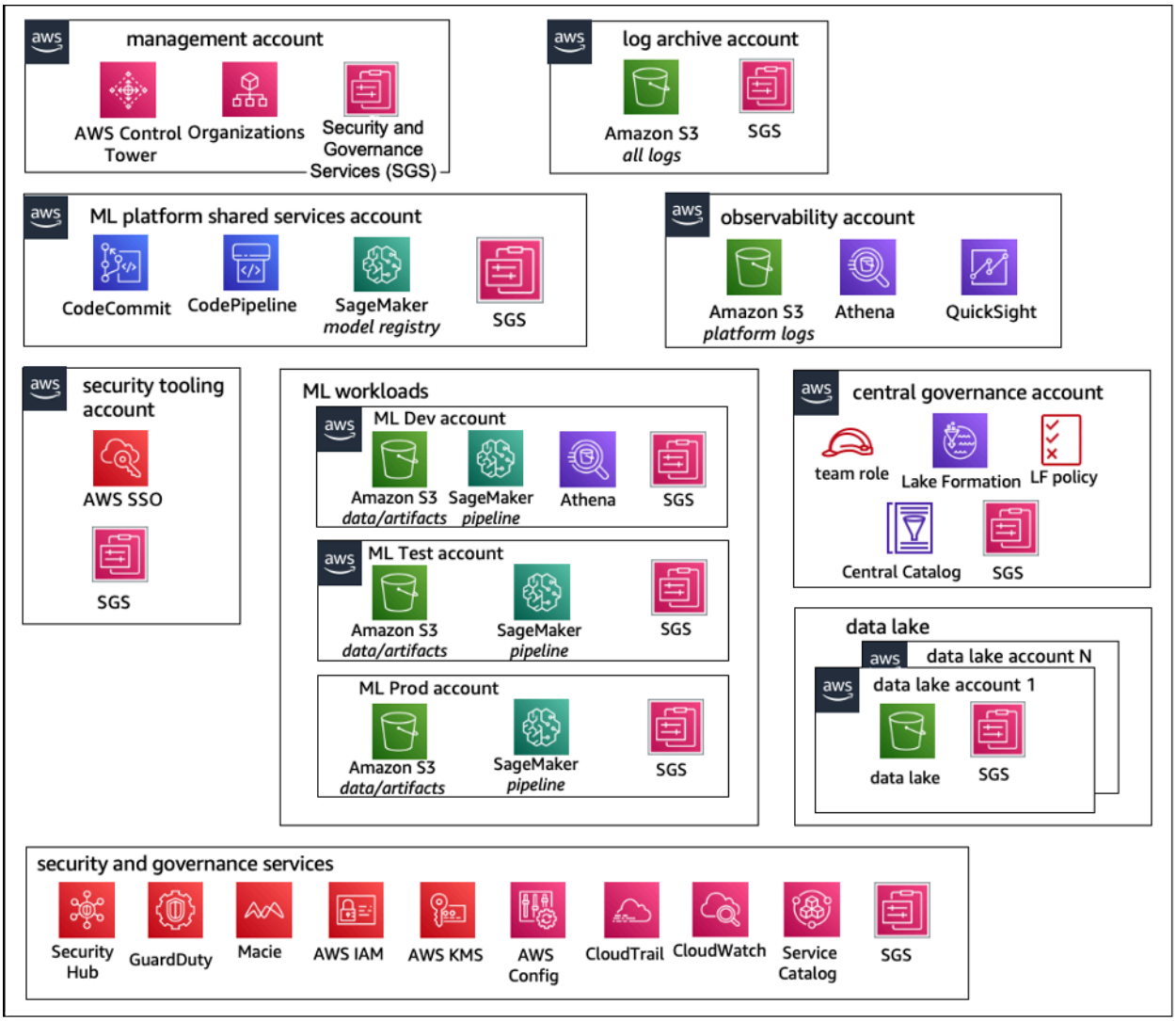
在本節中，我們簡要介紹了操作模型帳戶結構，您可以根據組織的操作需求，從 開始修改。無論您選擇何種操作模型，我們建議實作下列常見最佳實務：

- 使用 AWS Control Tower 設定、管理和控管您的帳戶。
- 使用 Identity Provider (IdP) 和 AWS IAM Identity Center 與委派管理員 Securitiy Tooling 帳戶集中您的身分，並啟用工作負載的安全存取。
- 跨開發、測試和生產工作負載，以帳戶層級隔離執行 ML 工作負載。
- 將 ML 工作負載日誌串流到日誌封存帳戶，然後在可觀測性帳戶中篩選和套用日誌分析。
- 執行集中式控管帳戶，以佈建、控制和稽核資料存取。
- 根據您的組織和工作負載需求，在每個帳戶中嵌入具有適當預防性和偵測性防護機制的安全和管理服務 (SGS)，以確保安全性和合規性。

集中化模型帳戶結構

在此模型中，ML 平台團隊負責提供：

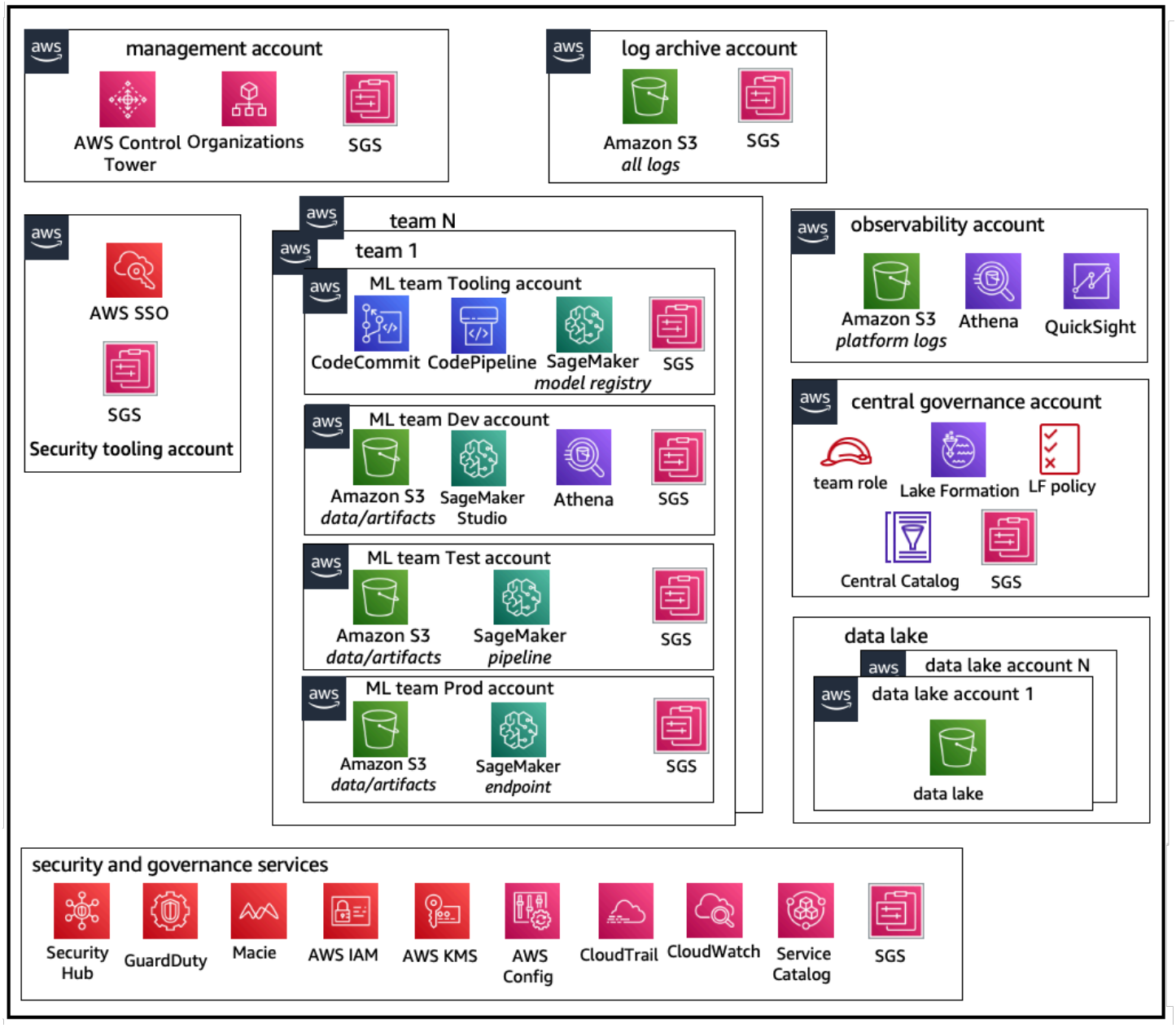
- 共用服務工具帳戶，可解決跨資料科學團隊的Machine Learning操作 (MLOps) 需求。
- 跨資料科學團隊共用的 ML 工作負載開發、測試和生產帳戶。
- 確保每個資料科學團隊工作負載獨立執行的控管政策。
- 常見最佳實務。



集中式操作模型帳戶結構

分散式模型帳戶結構

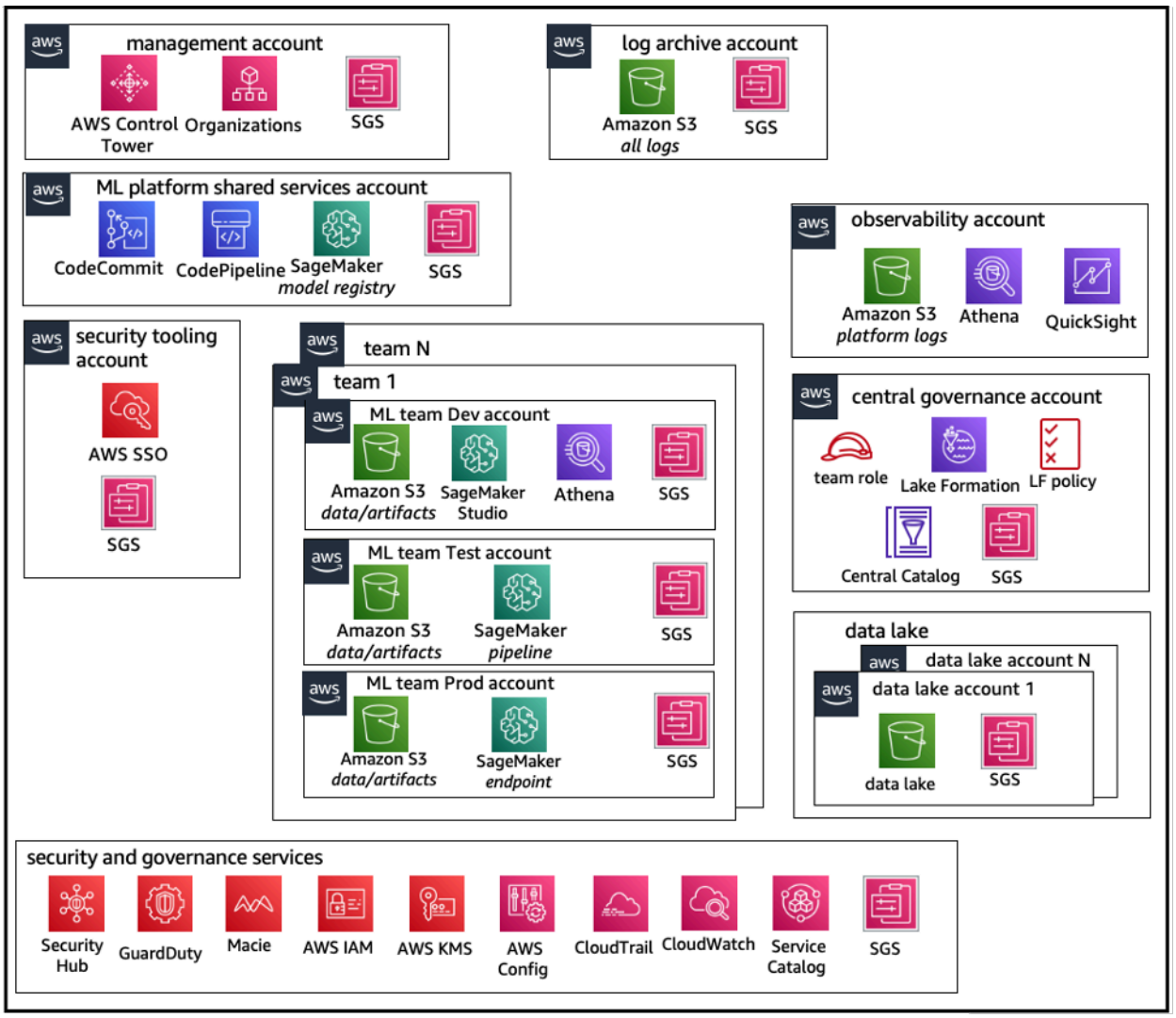
在此模型中，每個 ML 團隊都會獨立運作，以佈建、管理和管理 ML 帳戶和資源。不過，我們建議 ML 團隊使用集中式可觀測性和資料控管模型方法來簡化資料控管和稽核管理。



分散式操作模型帳戶結構

聯合模型帳戶結構

此模型與集中式模型類似；不過，關鍵差異在於每個資料science/ML team gets their own set of development/test/production工作負載帳戶，其可對其 ML 資源進行強健的實體隔離，並使每個團隊能夠獨立擴展，而不會影響其他團隊。



聯合操作模型帳戶結構

ML 平台多租戶

多租戶是一種軟體架構，單一軟體執行個體可以提供多個不同的使用者群組。租用戶是一組使用者，這些使用者共用軟體執行個體的特定權限。例如，如果您正在建置數個 ML 產品，則具有類似存取要求的每個產品團隊都可以被視為租戶或團隊。

雖然可以在 SageMaker AI Studio 執行個體（例如 SageMaker AI Domain) 中實作多個團隊，但當您將多個團隊帶入單一 SageMaker AI Studio 網域時，可以權衡這些優勢與權衡，例如爆量半徑、成本歸因和帳戶層級限制。在以下各節中進一步了解這些權衡和最佳實務。

如果您需要絕對資源隔離，請考慮為不同帳戶中的每個租戶實作 SageMaker AI Studio 網域。根據您的隔離需求，您可以在單一帳戶和區域中實作多行業務 (LOBs) 做為多個網域。使用共用空間，在同一團隊的成員之間進行近乎即時的協同合作/LOB。使用多個網域時，您仍然會使用身分存取管理 (IAM) 政策和許可來確保資源隔離。

SageMaker 從網域建立的 AI 資源會使用網域 Amazon Resource Name (ARN) 和使用者描述檔或空間自動標記，ARN以方便資源隔離。如需範例政策，請參閱網域資源隔離文件。您可以在此查看何時使用多帳戶或多網域策略的詳細參考，以及文件中的功能比較，而且您可以檢視範例指令碼，以回填GitHub 儲存庫上現有網域的標籤。

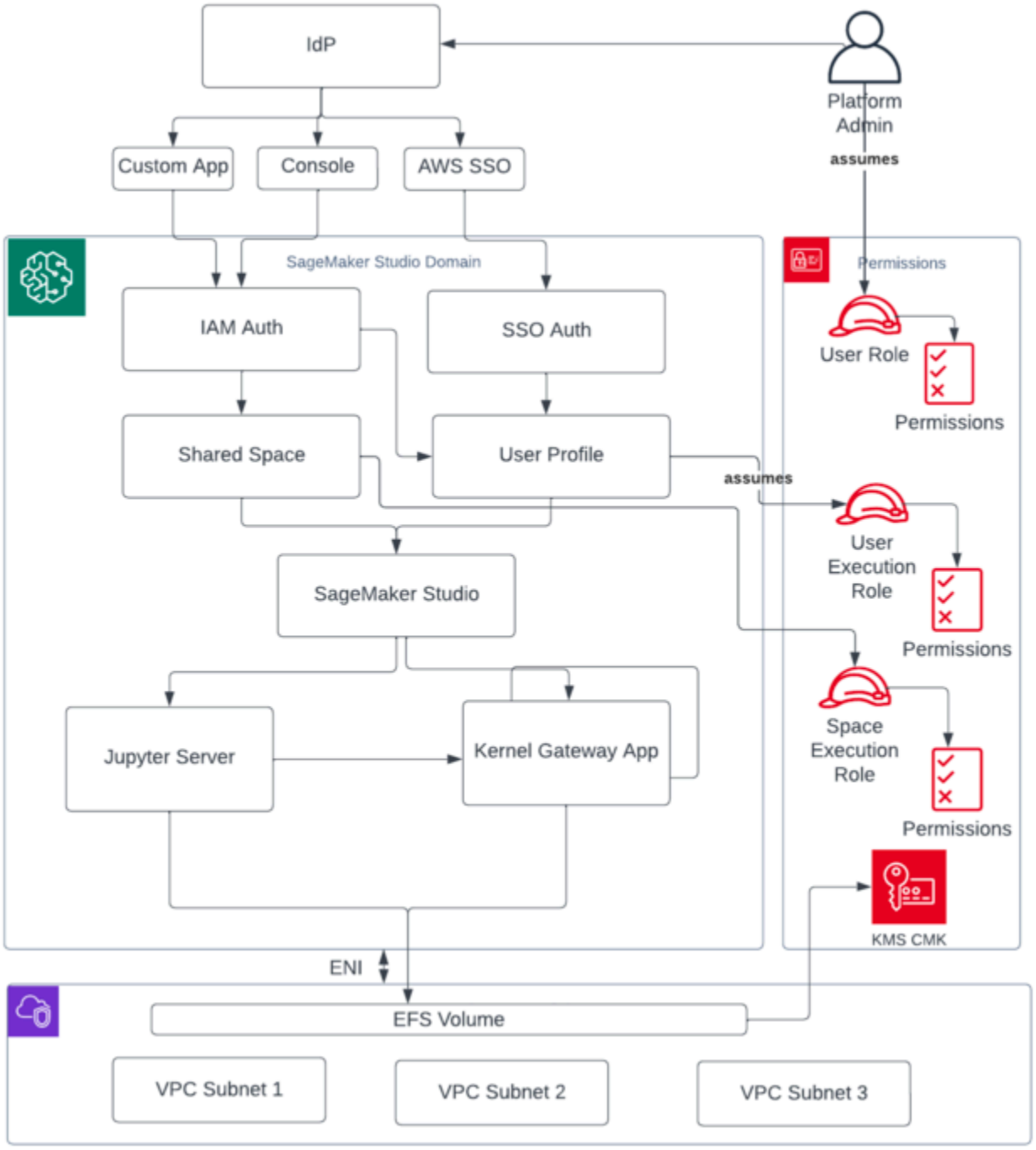
最後，您可以使用 將 SageMaker AI Studio 資源的自助部署實作到多個帳戶中AWS Service Catalog。如需詳細資訊，請參閱管理多個 AWS 帳戶 和 中的 AWS Service Catalog 產品 AWS 區域。

網域管理

Amazon SageMaker AI 網域包含：

- 相關聯的 Amazon Elastic File System (Amazon EFS) 磁碟區
- 授權使用者清單
- 各種安全性、應用程式、政策和 Amazon Virtual Private Cloud (AmazonVPC) 組態

下圖提供構成 SageMaker AIStudio網域之各種元件的高階檢視：

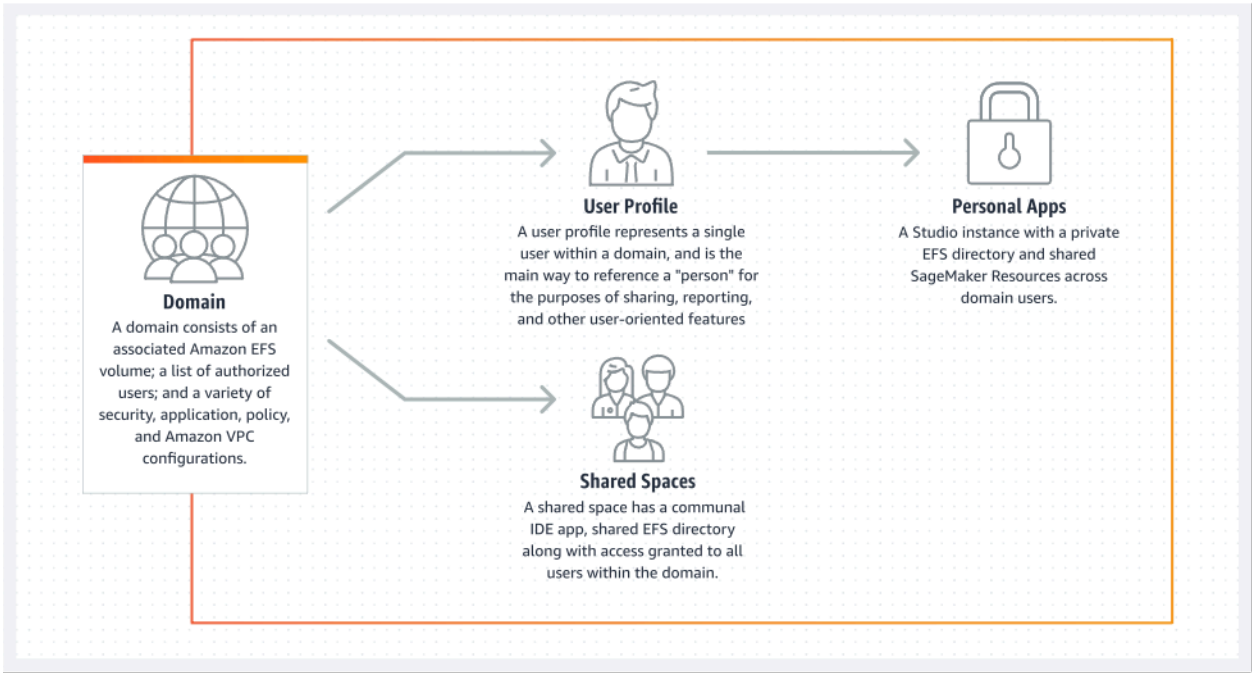


構成 SageMaker AI Studio 網域的各種元件的高階檢視

多個網域和共用空間

Amazon SageMaker AI 現在支援 AWS 區域 為每個帳戶在單一 中建立多個 SageMaker AI 網域。每個網域可以有自己的網域設定，例如身分驗證模式和聯網設定，例如 VPC和 子網路。使用者設定檔無法跨網域共用。如果人工使用者是多個團隊的一部分，以網域分隔，請為每個網域中的使用者建立使用者設定檔。請參閱多個網域概觀，以了解現有網域的回填標籤。

在IAM身分驗證模式中設定的每個網域都可以利用共用空間，在使用者之間進行近乎即時的協同合作。透過共用空間，使用者可以存取共用的 Amazon EFS目錄，以及使用者介面的共用JupyterServer應用程式，而且可以近乎即時地共同編輯。共用空間建立的資源自動標記可讓管理員追蹤專案層級的成本。共用 JupyterServer UI 也會篩選實驗和模型登錄項目等資源，因此只會顯示與共用 ML 工作相關的項目。下圖提供每個網域內私有應用程式和共用空間的概觀。



單一網域中的私有應用程式和共用空間概觀

在網域中設定共用空間

共用空間通常為特定 ML 工作或專案建立，其中單一網域的成員需要近乎即時地存取相同的基礎檔案儲存體和 IDE。使用者可以近乎即時地存取、讀取、編輯和共用筆記本，這為他們提供最快的路徑，以開始與對等進行反覆運算。

若要建立共用空間，您必須先指定空間預設執行角色，該角色將管理任何使用空間的使用者的許可。撰寫本文時，網域內的所有使用者都可以存取其網域中的所有共用空間。請參閱建立共用空間，以取得將共用空間新增至現有網域的最新文件。

為IAM聯合設定您的網域

在為 SageMaker AI Studio 網域設定 AWS Identity and Access Management (IAM) 聯合之前，您需要在 IdP IAM 中設定聯合使用者角色（例如平台管理員），如身分管理一節所述。

如需使用 IAM選項設定 SageMaker AI Studio 的詳細說明，請參閱使用 IAM Identity Center 加入 Amazon SageMaker 網域。

為單一登入 (SSO) 聯合設定您的網域

若要使用單一登入 (SSO) 聯合，您需要 AWS IAM Identity Center 在管理AWS Organizations帳戶中啟用 ，並在您需要執行 SageMaker AI Studio 的相同區域中啟用 。網域設定步驟類似於IAM聯合步驟，但您在身分驗證區段中選取 AWS IAM Identity Center(IdC) 除外。

如需詳細說明，請參閱使用 IAM Identity Center 加入 Amazon SageMaker 網域。

SageMaker AI Studio 使用者設定檔

使用者描述檔代表網域中的單一使用者，是參考「人員」的主要方式，用於共用、報告和其他使用者導向功能。當使用者加入 toSageMaker AI Studio 時，就會建立此實體。如果管理員透過電子郵件邀請某個人員或從 IdC 匯入，則會自動建立使用者設定檔。使用者設定檔是個別使用者設定的主要擁有者，並參考使用者的私有 Amazon Elastic File System (Amazon EFS) 主目錄。我們建議為每個 SageMaker AI Studio 應用程式的實體使用者建立使用者設定檔。每個使用者在 Amazon 上都有自己的專用目錄EFS，且使用者設定檔無法跨相同帳戶中的網域共用。

共用 SageMaker AI Studio 網域的每個使用者設定檔都會取得專用運算資源（例如 SageMaker AI Amazon Elastic Compute Cloud (AmazonEC2) 執行個體 (s))，以執行筆記本。配置給使用者一的運算執行個體與配置給使用者二的運算執行個體完全隔離。同樣地，分配給某個 AWS 帳戶中使用者的運算資源與分配給另一個帳戶中使用者的運算資源完全不同。每個使用者可以在隔離的 Docker 容器中執行最多四個應用程式（應用程式)，或在相同執行個體類型上執行映像。

Jupyter 伺服器應用程式

當您透過存取預先簽章URL或使用 IdC AWS IAM 登入來為使用者啟動 Amazon SageMaker AI Studio 筆記本時，Jupyter Server 應用程式會在 SageMaker AI 服務受管VPC執行個體中啟動。每個使用者都會在私有應用程式中取得自己的專用 Jupyter Server 應用程式。根據預設，適用於 SageMaker AI Studio 筆記本的 Jupyter Server 應用程式會在專用`ml.t3.medium`執行個體上執行（保留為系統執行個體類型）。此執行個體的運算不會向客戶收費。

Jupyter Kernel Gateway 應用程式

核心閘道應用程式可以透過 API或 SageMaker AI Studio 介面建立，並在所選的執行個體類型上執行。此應用程式可以使用其中一個預先設定熱門資料科學的內建 SageMaker AI Studio 映像，以及深度學習套件，例如 TensorFlow、Apache MXNet和 來執行PyTorch。

使用者可以在相同的 SageMaker Studio 中啟動和執行多個 Jupyter 筆記本核心、終端機工作階段和互動式主控台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。

若要建立其他應用程式，您需要使用不同的執行個體類型。使用者描述檔只能有一個執行中執行個體的任何執行個體類型。例如，使用者可以在同一個執行個體上使用 SageMaker AI Studio 內建資料科學映像執行簡單的筆記本，並使用內建 TensorFlow 映像執行另一個筆記本。使用者需支付執行個體執行時間的費用。為了避免使用者在未主動執行 SageMaker AI Studio 時產生成本，使用者需要關閉執行個體。如需詳細資訊，請參閱關閉並更新 Studio 應用程式。

每次從 SageMaker AI Studio 界面關閉並重新開啟核心閘道應用程式時，該應用程式都會在新的執行個體上啟動。這表示套件的安裝不會透過重新啟動相同的應用程式而持續存在。同樣地，如果使用者變更筆記本上的執行個體類型，其已安裝的套件和工作階段變數也會遺失。不過，您可以使用自己的映像和生命週期指令碼等功能，將使用者自己的套件帶入 SageMaker AI Studio，並透過執行個體切換和新的執行個體啟動來保留這些套件。

Amazon Elastic File System 磁碟區

建立網域時，會建立單一 Amazon Elastic File System (AmazonEFS) 磁碟區供網域內的所有使用者使用。每個使用者設定檔都會在 Amazon EFS磁碟區中收到私有主目錄，用於存放使用者的筆記本、GitHub 儲存庫和資料檔案。網域中的每個空間都會在 Amazon EFS磁碟區中收到私有目錄，可供多個使用者設定檔存取。使用者透過檔案系統許可來分隔資料夾的存取。 SageMaker AI Studio 會為每個使

用者設定檔或空間建立全域唯一使用者 ID，並將其套用為可攜式作業系統界面 (POSIX) user/group ID for the user's home directory on EFS, which prevents other users/spaces來存取其資料。

備份與復原

現有的EFS磁碟區無法連接至新的 SageMaker AI 網域。在生產設定中，確定 Amazon EFS磁碟區已備份 （ 備份至另一個EFS磁碟區，或備份至 Amazon Simple Storage Service (Amazon S3))。如果不小心刪除了磁碟EFS區，管理員必須向下撕下並重新建立 SageMaker AI Studio 網域。程序如下：

透過 、、 和 `DescribeSpace`API呼叫，備份使用者設定檔`ListUserProfiles`、空格`ListSpaces`和相關聯EFS使用者 IDs(UIDs) `DescribeUserProfile`的清單。

1. 建立新的 SageMaker AI Studio 網域。
2. 建立使用者設定檔和空格。
3. 對於每個使用者設定檔，從 EFS/Amazon S3 上的備份複製檔案。
4. 或者，刪除舊 SageMaker AI Studio 網域上的所有應用程式和使用者設定檔。

如需詳細說明，請參閱附錄的 SageMaker AI Studio 網域備份和復原一節。

> **Note**
>
> 您也可以透過 來達成此目的`LifecycleConfigurations`，在每次使用者啟動其應用程式時，將資料備份到 S3 和從 S3 備份。

Amazon EBS磁碟區

Amazon Elastic Block Store (Amazon EBS) 儲存磁碟區也會連接至每個 SageMaker AI Studio Notebook 執行個體。它會用作在執行個體上執行的容器或映像的根磁碟區。雖然 Amazon EFS儲存體為持久性，但連接至容器的 Amazon EBS磁碟區是暫時的。如果客戶刪除應用程式，則儲存在 Amazon EBS磁碟區本機的資料不會保留。

保護對預先簽署 的存取 URL

當 SageMaker AI Studio 使用者開啟筆記本連結時， SageMaker AI Studio 會驗證聯合身分使用者IAM的政策，以授權存取，並產生和解析URL預先簽章的使用者。由於 SageMaker AI 主控台在網際網路網

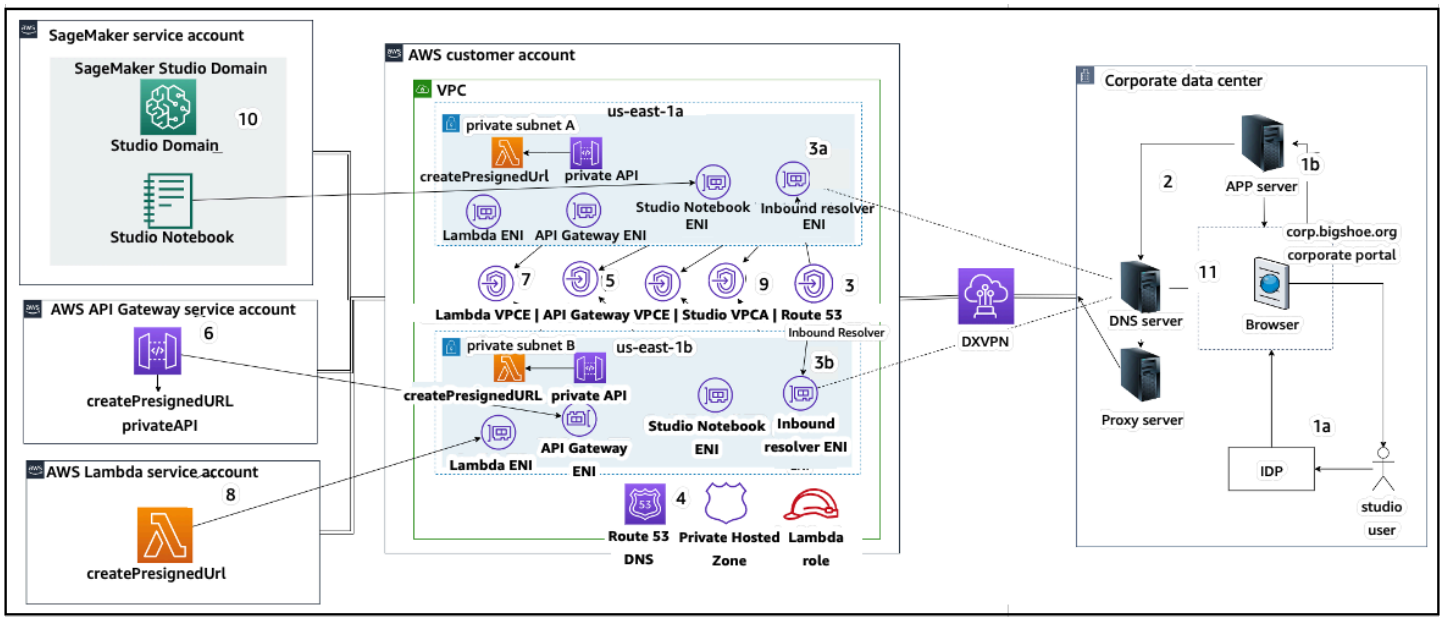
域上執行，因此預先簽章URL會顯示在瀏覽器工作階段中。這會顯示不想要的威脅向量，用於資料遭竊，並在未強制執行適當的存取控制時存取客戶資料。

Studio 支援對預先簽署URL的資料遭竊強制執行存取控制的幾種方法：

- 使用IAM政策條件進行用戶端 IP 驗證`aws:sourceIp`
- 使用 IAM條件進行用戶端VPC驗證`aws:sourceVpc`
- 使用IAM政策條件進行用戶端VPC端點驗證`aws:sourceVpce`

當您從 SageMaker AI 主控台存取 SageMaker AI Studio 筆記本時，唯一可用的選項是搭配IAM政策條件使用用戶端 IP 驗證`aws:sourceIp`。不過，您可以使用瀏覽器流量路由產品，例如 Zscaler，以確保人力資源網際網路存取的規模和合規性。這些流量路由產品會產生自己的來源 IP，其 IP 範圍不受企業客戶控制。這使得這些企業客戶無法使用 `aws:sourceIp`條件。

若要使用IAM政策條件來使用用戶端VPC端點驗證`aws:sourceVpce`，建立預先簽章URL需要源自於部署 SageMaker AI Studio VPC的相同客戶，且URL需透過客戶 上的 SageMaker AI Studio VPC端點解決預先簽章的問題VPC。可以使用DNS轉送規則 (Zscaler 和公司)，DNS然後使用 Amazon Route 53 傳入解析程式在客戶VPC端點中完成企業網路使用者URL存取期間預先簽署的此解析，如下列架構所示：



透過公司網路URL使用VPC端點存取預先簽署的 Studio

如需 step-by-step設定上述架構的指引，請參閱 Secure Amazon SageMaker AI Studio 預先簽章第 1 URLs部分：基礎基礎設施。

SageMaker AI 網域配額和限制

- SageMaker AI Studio SSO 網域聯合僅支援 區域，跨佈建 AWS Identity Center 之 AWS 組織的成員帳戶。
- 使用 AWS Identity Center 設定的網域目前不支援共用空間。
- VPC 和子網路組態無法在建立網域之後變更。不過，您可以使用不同的VPC子網路組態建立新的網域。
- 建立網域後，無法在 IAM和 SSO 模式之間切換網域存取。您可以使用不同的身分驗證模式建立新的網域。
- 每個使用者啟動的每個執行個體類型限制四個核心閘道應用程式。
- 每個使用者只能啟動每個執行個體類型的一個執行個體。
- 網域內消耗的資源有限制，例如執行個體類型啟動的執行個體數量，以及可建立的使用者設定檔數量。如需服務限制的完整清單，請參閱服務配額頁面。
- 客戶可以提交具有業務理由的企業支援案例，以提高預設資源限制，例如受帳戶層級護欄約束的網域或使用者設定檔數量。
- 每個帳戶的並行應用程式數量硬性限制為 2，500 個應用程式。網域和使用者設定檔限制取決於此硬性限制。例如， 帳戶可以具有具有 1，000 個使用者描述檔的單一網域，或具有 20 個網域，每個網域具有 50 個使用者描述檔。

身分管理

本節討論公司目錄中的人力資源使用者如何聯合 AWS 帳戶 並存取 SageMaker AI Studio。首先，我們將簡短描述使用者、群組和角色的映射方式，以及使用者聯合的運作方式。

使用者、群組和角色

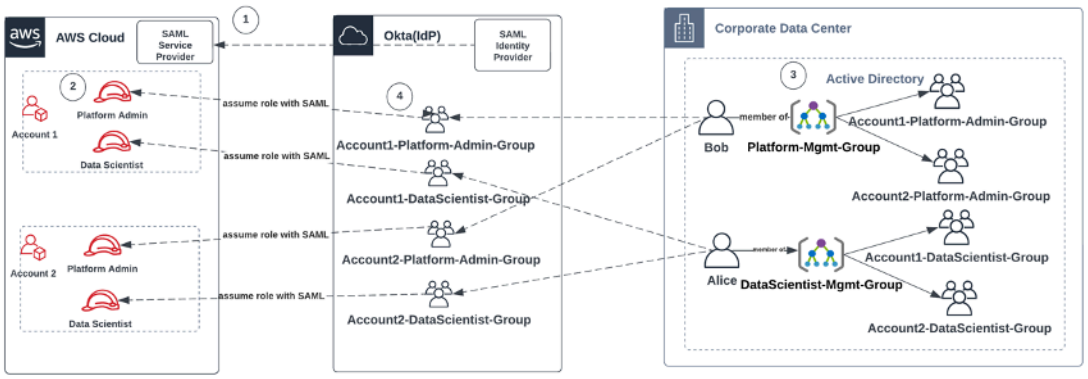
在 中 AWS，使用使用者、群組和角色來管理資源許可。客戶可以透過 或在企業目錄中管理其使用者和群組IAM，例如 Active Directory (AD)，透過 Okta 等外部 IdP 啟用，這可讓他們向雲端和內部部署中執行的各種應用程式驗證使用者。

如 AWS Security Pillar Identity Management 一節所述，最佳實務是在中央 IdP 中管理使用者身分，因為這有助於輕鬆整合您的後端人力資源程序，並有助於管理人力資源使用者的存取權。

IdPs 例如 Okta 可讓最終使用者對一或多個 進行身分驗證，AWS 帳戶 並使用 SSO 搭配安全聲明標記語言 () 存取特定角色SAML。IdP 管理員能夠將角色從 下載 AWS 帳戶 到 IdP，並將這些角色指派給使用者。登入 時 AWS，最終使用者會看到一個 AWS 畫面，顯示一或多個 中指派給他們的清單 AWS 角色 AWS 帳戶。他們可以選取要擔任的登入角色，定義其在該已驗證工作階段期間的權限。

對於您要提供存取權的每個特定帳戶和角色組合， 群組必須存在於 IdP 中。您可以考慮這些群組為AWS 角色特定的群組。任何身為這些角色特定群組成員的使用者都會獲得單一權限：存取特定 中的一個特定角色 AWS 帳戶。不過，此單一授權程序不會透過將每個使用者指派給特定 AWS 角色群組來擴展以管理使用者存取。為了簡化管理，我們建議您也為組織中需要不同 AWS 權限集的所有不同使用者集建立多個群組。

若要說明中央 IdP 設定，請考慮具有 AD 設定的企業，其中使用者和群組會同步到 IdP 目錄。在 中 AWS，這些 AD 群組會對應至IAM角色。工作流程的主要步驟如下：



加入 AD 使用者、AD 群組和IAM角色的工作流程

1. 在 中 AWS，設定每個 AWS 帳戶 與 IdP 的SAML整合。
2. 在 中 AWS，在每個 中設定角色， AWS 帳戶 並同步到 IdP。
3. 在公司 AD 系統中：
   a. 為每個帳戶角色建立 AD 群組並同步至 IdP （例如，`Account1-Platform-Admin-Group`（也稱為 AWS 角色群組））。
   b. 在每個角色層級建立管理群組 （例如 `Platform-Mgmt-Group`)，並將 AWS 角色群組指派為成員。
   c. 將使用者指派給該管理群組，以允許存取 AWS 帳戶 角色。
4. 在 IdP 中，將 AWS 角色群組 （例如 `Account1-Platform-Admin-Group`) 映射至 AWS 帳戶 角色 （例如 Account1 中的平台管理員）。
5. 當 Data Scientist Alice 登入 Idp 時，他們會收到 AWS 一個聯合應用程式 UI，有兩個選項可供選擇：「帳戶 1 Data Scientist」和「帳戶 2 Data Scientist」。
6. Alice 選擇「帳戶 1 資料科學家」選項，它們連接到 AWS 帳戶 1 (SageMaker AI 主控台） 中的授權應用程式。

如需設定SAML帳戶聯合的詳細資訊，請參閱 Okta AWS 的帳戶聯合設定方式 SAML2.0。

使用者聯合

SageMaker AI Studio 的身分驗證可以使用 IAM或 IAM IdC 完成。如果使用者是透過 管理IAM，他們可以選擇 IAM 模式。如果企業使用外部 IdP，他們可以透過 IAM或 IAM IdC 聯合。請注意，無法更新現有 SageMaker AI Studio 網域的身分驗證模式，因此在建立生產 SageMaker AI Studio 網域之前做出決策至關重要。

如果 SageMaker AI Studio 設定為 IAM 模式， SageMaker AI Studio 使用者會透過預先簽章存取應用程式URL，該預先簽章會在透過瀏覽器存取時自動將使用者登入 SageMaker AI Studio 應用程式。

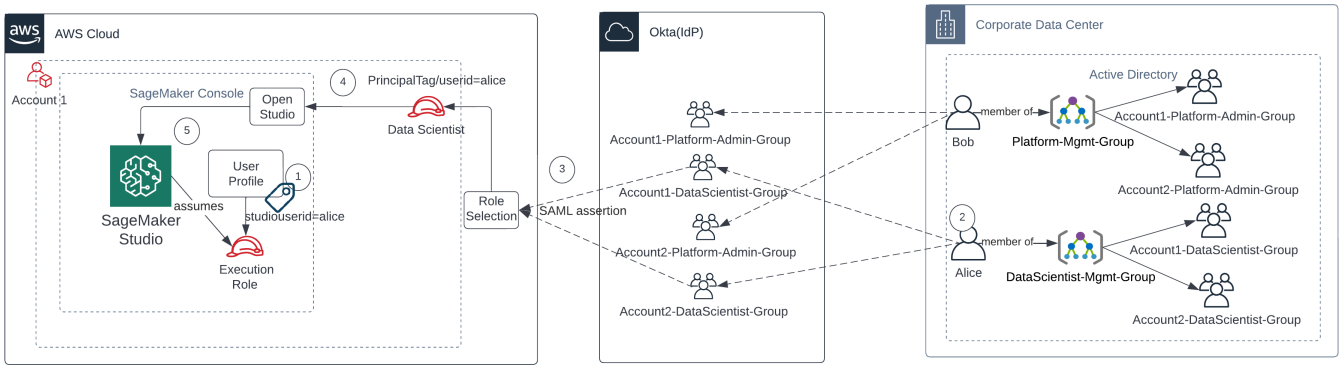
IAM 使用者

對於IAM使用者，管理員會為每個使用者建立 SageMaker AI Studio 使用者設定檔，並將使用者設定檔與 IAM角色建立關聯，以允許使用者在 Studio 中執行的必要動作。若要限制 AWS 使用者僅存取其 SageMaker AI Studio 使用者設定檔，管理員應標記 SageMaker AI Studio 使用者設定檔，並將IAM政策連接至使用者，以允許他們只有在標籤值與 AWS 使用者名稱相同時存取。政策陳述式如下所示：

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

AWS IAM 或 帳戶聯合

AWS 帳戶 聯合方法可讓客戶從其 IdP 聯合至 SageMaker AI SAML 主控台，例如 Okta。若要限制使用者僅存取其使用者設定檔，管理員應標記 SageMaker AI Studio 使用者設定檔、`PrincipalTags`新增 IdP，並將他們設定為傳輸標籤。下圖說明聯合身分使用者 (Data Scientist Alice) 如何獲得授權存取自己的 SageMaker AI Studio 使用者設定檔。



在IAM聯合模式中存取 SageMaker AI Studio

1. Alice SageMaker AI Studio 使用者設定檔會加上其使用者 ID 的標籤，並與執行角色相關聯。
2. Alice 驗證 IdP (Okta)。

3. IdP 會驗證 Alice 並發佈聲明，SAML其中包含兩個角色（帳戶 1 和 2 的資料科學家） Alice 是其中的成員。Alice 為帳戶 1 選取資料科學家角色。
4. Alice 已登入帳戶 1 SageMaker AI 主控台，並擔任 Data Scientist 的角色。Alice 從 Studio 應用程式執行個體清單中開啟其 Studio 應用程式執行個體。
5. 擔任角色工作階段中的 Alice 主體標籤會根據選取的 SageMaker AI Studio 應用程式執行個體使用者設定檔標籤進行驗證。如果設定檔標籤有效，則會啟動 SageMaker AI Studio 應用程式執行個體，並擔任執行角色。

如果您想要在使用者加入時自動建立 SageMaker AI 執行角色和政策，以下是完成此操作的一種方法：

1. 在每個帳戶和 Studio `SageMaker AI-Account1-Group` 網域層級設定 AD 群組，例如 。
2. 當您需要將使用者加入 SageMaker AI Studio 時，將 SageMaker AI-Account1-Group 新增至使用者的群組成員資格。

設定可接聽`SageMaker AI-Account1-Group`成員資格事件的自動化程序，並使用 AWS APIs 根據其 AD 群組成員資格來建立角色、政策、標籤和 SageMaker AI Studio 使用者設定檔。將角色連接至使用者設定檔。如需範例政策，請參閱 防止 SageMaker AI Studio 使用者存取其他使用者設定檔。

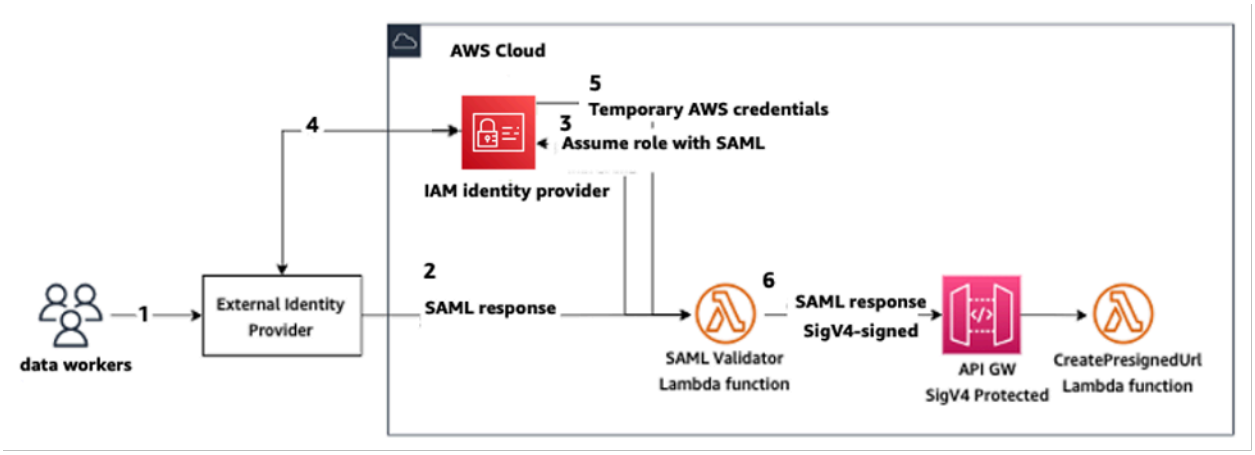
SAML 使用 進行身分驗證 AWS Lambda

在 IAM 模式下，使用者也可以使用SAML聲明驗證 SageMaker AI Studio。在此架構中，客戶有現有的 IdP，他們可以建立SAML應用程式，讓使用者存取 Studio（而非 AWS Identity Federation 應用程式）。客戶的 IdP 已新增至 IAM。 AWS Lambda 函數有助於使用 IAM和 驗證SAML聲明STS，然後直接叫用API閘道或 Lambda 函數，以建立預先簽章的網域 URL。

此解決方案的優點是 Lambda 函數可以自訂邏輯以存取 SageMaker AI Studio。例如：

- 如果使用者描述檔不存在，則自動建立使用者描述檔。
- 剖析SAML屬性，將角色或政策文件附加或移除至 SageMaker AI Studio 執行角色。
- 新增生命週期組態 (LCC) 和新增標籤，以自訂使用者設定檔。

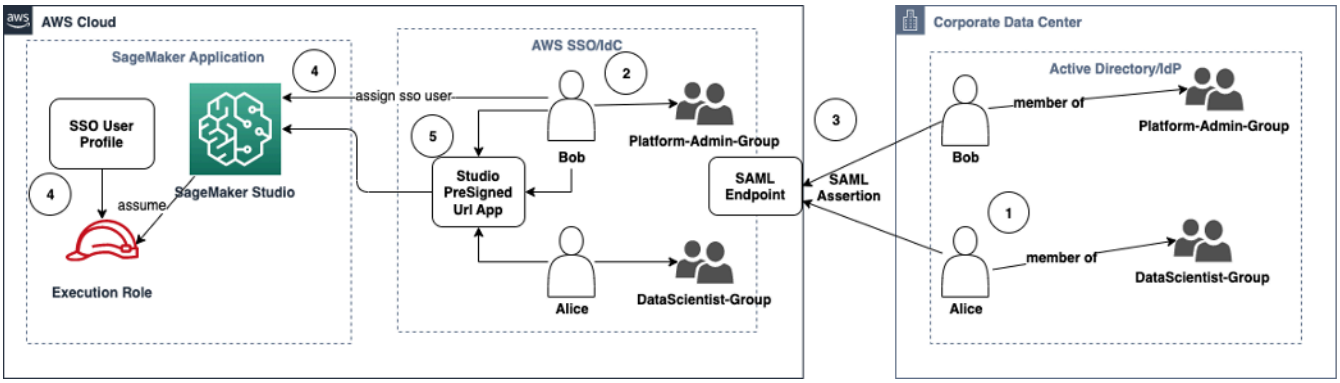
總而言之，此解決方案會將 SageMaker AI Studio 公開為具有用於身分驗證和授權的自訂邏輯的 SAML2.0 應用程式。如需實作詳細資訊，請參閱使用SageMaker SAML聲明的 Studio 存取附錄一節。



使用自訂SAML應用程式存取 SageMaker AI Studio

AWS IAM IdC 聯合

IdC 聯合方法可讓客戶從其 IdP 直接聯合到 SageMaker AI Studio 應用程式 SAML （例如 Okta)。下圖說明聯合身分使用者如何獲得存取自己 SageMaker AI Studio 執行個體的授權。



在 IAM IdC 模式下存取 SageMaker AI Studio

1. 在公司 AD 中，使用者是 AD 群組的成員，例如平台管理員群組和資料科學家群組。
2. Identity Provider (IdP) 的 AD 使用者和 AD 群組會同步至 AWS IAM Identity Center，並分別做為指派的單一登入使用者和群組使用。
3. IdP 會將SAML宣告發佈至 AWS IdC SAML端點。
4. 在 SageMaker AI Studio 中，IdC 使用者會指派給 SageMaker Studio 應用程式。此指派可以使用 IdC 群組完成， SageMaker AI Studio 將套用至每個 IdC 使用者層級。建立此指派時， SageMaker AI Studio 會建立 IdC 使用者設定檔並連接網域執行角色。

5. 使用者使用 IdC 中作為雲端應用程式URL託管的安全預先簽章來存取 SageMaker AI Studio 應用程式。 SageMaker AI Studio 會擔任連接至其 IdC 使用者設定檔的執行角色。

網域身分驗證指南

以下是選擇網域身分驗證模式時的一些考量：

1. 如果您希望使用者無法存取 AWS 管理主控台 並直接檢視 SageMaker AI Studio UI，請使用具有 IdC 的 AWS IAM單一登入模式。
2. 如果您希望使用者無法存取 AWS 管理主控台 並直接在 IAM 模式下檢視 SageMaker AI Studio UI，您可以使用後端的 Lambda 函數產生URL使用者描述檔的預先簽章，並將他們重新導向至 SageMaker AI Studio UI。
3. 在 IdC 模式中，每個使用者都會對應至單一使用者設定檔。
4. 所有使用者設定檔都會在 IdC 模式中自動指派預設執行角色。如果您希望使用者獲指派不同的執行角色，您將需要使用 UpdateUserProfile 更新使用者設定檔API。
5. 如果您想要將IAM模式 （使用產生的預先簽章 URL) 中的 SageMaker AI Studio UI 存取限制在VPC端點，而不周遊網際網路，您可以使用自訂DNS解析程式。請參閱 Secure Amazon SageMaker AI Studio 預先簽署的第 1 URLs部分：基礎基礎設施部落格文章。

許可管理

本節討論設定常用IAM角色、政策和護欄以佈建和操作 SageMaker AI Studio 網域的最佳實務。

IAM 角色和政策

最佳實務是，您可能想要先識別相關的人員和應用程式，稱為涉及 ML 生命週期的委託人，以及授予他們所需的 AWS 許可。由於 SageMaker AI 是一項受管服務，因此您也需要考慮服務主體，這些服務主體是可以代表使用者API撥打電話 AWS 的服務。下圖說明您可能想要建立的不同IAM角色，對應至組織中的不同角色。

SageMaker AI IAM角色

這些角色會詳細說明，以及IAMpermissions一些需要的特定範例。

- ML Admin 使用者角色 — 這是為資料科學家佈建環境的委託人，方法是建立 Studio 網域和使用者設定檔 (`sagemaker:CreateDomain`、`sagemaker:CreateUserProfile`)、為使用者建立 AWS Key Management Service (AWS KMS) 金鑰、為資料科學家建立 S3 儲存貯體，以及建立 Amazon ECR 儲存庫來存放容器。他們也可以為使用者設定預設組態和生命週期指令碼、建置自訂映像並將其連接至 SageMaker AI Studio 網域，以及提供 Service Catalog 產品，例如自訂專案、Amazon EMR 範本。

  例如，由於此委託人不會執行訓練任務，因此不需要啟動 SageMaker AI 訓練或處理任務的許可。如果他們使用基礎設施做為程式碼範本，例如 CloudFormation 或 Terraform 來佈建網域和使用者，則佈建服務會擔任此角色，以代表管理員建立資源。此角色可能具有使用 的 SageMaker AI 唯讀存取權 AWS 管理主控台。

  此使用者角色還需要特定EC2許可，才能在私有 內啟動網域VPC、加密EFS磁碟區的KMS許可，以及為 Studio 建立服務連結角色的許可 (`iam:CreateServiceLinkedRole`)。我們將在文件稍後說明這些精細許可。

- Data Scientist 使用者角色 — 此主體是登入 SageMaker AI Studio、探索資料、建立處理和訓練任務和管道的使用者，以此類推。使用者需要的主要許可是啟動 SageMaker AI Studio 的許可，其餘政策可由 SageMaker AI 執行服務角色管理。
- SageMaker AI 執行服務角色：因為 SageMaker AI 是受管服務，所以會代表使用者啟動任務。此角色通常在允許的許可方面最為廣泛，因為許多客戶選擇使用單一執行角色來執行訓練任務、處理任務或模型託管任務。雖然這是開始使用的簡單方法，但由於客戶在其旅程中成熟，他們通常會將筆記本執行角色分割為不同API動作的不同角色，特別是在部署環境中執行這些任務時。

  您在建立角色時將角色與 SageMaker AI Studio 網域建立關聯。不過，由於客戶可能需要與網域中的不同使用者設定檔相關聯的不同角色的彈性（例如，根據其任務函數），您也可以將個別IAM角色與每個使用者設定檔建立關聯。我們建議您將單一實體使用者映射至單一使用者設定檔。如果您在建立時未將角色連接至使用者設定檔，預設行為也會將網域執行角色與使用者設定檔建立關聯SageMaker AIStudio。

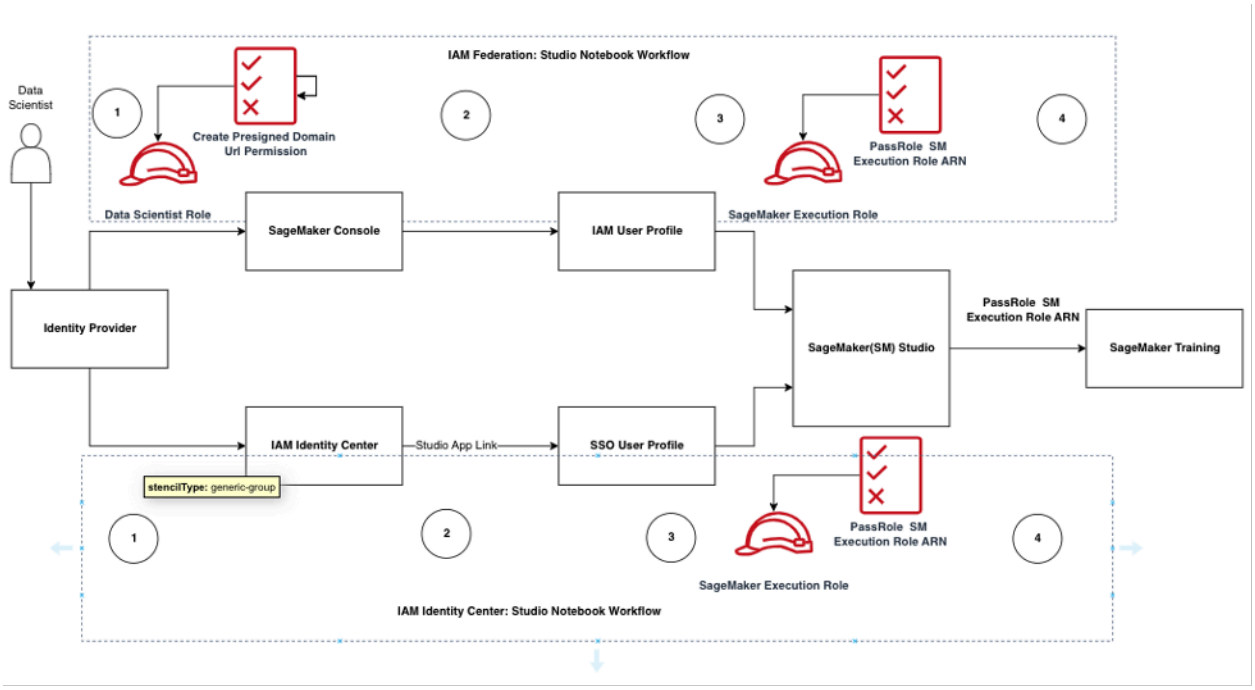
  如果多個資料科學家和 ML 工程師在專案上合作，且需要共用許可模型來存取資源，我們建議您建立團隊層級 SageMaker AI 服務執行角色，以跨團隊成員共用IAM許可。在您需要在每個使用者層級鎖定許可的執行個體中，您可以建立個別使用者層級 SageMaker AI 服務執行角色；不過，您需要注意您的服務限制。

SageMaker AI Studio 筆記本授權工作流程

本節討論 SageMaker AI Studio Notebook 授權如何用於資料科學家直接從 SageMaker AI Studio Notebook 建置和訓練模型時需要執行的各種活動。 SageMaker AI 網域支援兩種授權模式：

- IAM 聯合
- IAM 身分中心

接下來，本文會逐步引導您完成每個模式的資料科學家授權工作流程。



Studio 使用者的身分驗證和授權工作流程

IAM 聯合： SageMaker Studio Notebook 工作流程

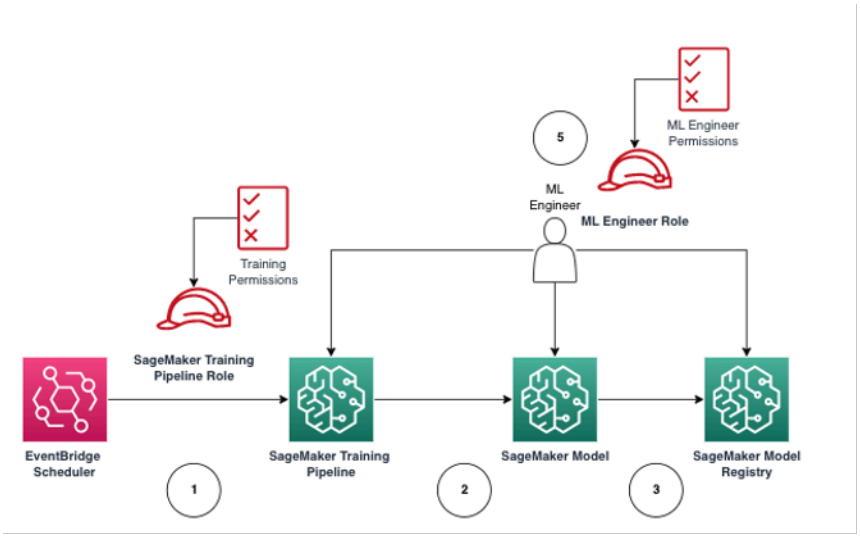
1. Data Scientist 會向企業身分提供者進行身分驗證，並在 SageMaker AI 主控台中擔任 Data Scientist 使用者角色（使用者聯合角色）。此聯合角色具有 SageMaker AI 執行角色的`iam:PassRole`API 許可，可將角色 Amazon Resource Name (ARN) 傳遞至 SageMaker Studio。
2. Data Scientist 從與 SageMaker AI 執行角色相關聯的 Studio IAM使用者設定檔中選取 Open Studio 連結
3. 假設使用者的 SageMaker 執行角色許可，會啟動 SageMaker Studio IDE服務。此角色具有 SageMaker AI 執行角色的`iam:PassRole`API許可，可將角色傳遞ARN給 SageMaker AI 訓練服務。
4. 當 Data Scientist 在遠端運算節點中啟動訓練任務時， SageMaker AI 執行角色ARN會傳遞給 SageMaker AI 訓練服務。這會使用此項目建立新的角色工作階段，ARN並執行訓練任務。如果您需要進一步縮小訓練任務的許可範圍，您可以建立訓練特定角色，並在呼叫訓練 ARN時傳遞該角色 API。

IAM Identity Center： SageMaker AI Studio 筆記本工作流程

1. Data Scientist 會向企業身分提供者進行身分驗證，然後按一下AWSIAM身分中心。Data Scientist 會與使用者身分中心入口網站一起顯示。
2. Data Scientist 按一下從其 IdC 使用者設定檔建立的 SageMaker AI Studio 應用程式連結，該設定檔與 SageMaker AI 執行角色相關聯。
3. 假設使用者的 SageMaker AI 執行角色許可，則啟動 SageMaker AI Studio IDE服務。此角色具有 SageMaker AI 執行角色的`iam:PassRole`API許可，可將角色傳遞ARN給 SageMaker AI 訓練服務。
4. 當資料科學家在遠端運算節點中啟動訓練任務時， SageMaker AI 執行角色ARN會傳遞給 SageMaker AI 訓練服務。執行角色會使用此 ARN建立新的角色工作階段ARN，並執行訓練任務。如果您需要進一步縮小訓練任務的許可範圍，您可以建立訓練特定的角色，並在呼叫訓練 ARN時傳遞該角色API。

部署環境： SageMaker AI 訓練工作流程

在部署環境中，例如系統測試和生產，任務是透過自動排程器和事件觸發程序執行，而對這些環境的人工存取會受到 SageMaker AI Studio Notebooks 的限制。本節討論IAM角色如何在部署環境中使用 SageMaker AI 訓練管道。



SageMaker 受管生產環境中的 AI 訓練工作流程

1. Amazon EventBridge 排程器會觸發 SageMaker AI 訓練管道任務。
2. SageMaker AI 訓練管道任務會擔任 SageMaker AI 訓練管道角色來訓練模型。
3. 訓練過的 SageMaker AI 模型會註冊到 SageMaker AI 模型登錄檔中。

4. ML 工程師擔任 ML 工程師使用者角色來管理訓練管道和 SageMaker AI 模型。

資料許可

SageMaker AI Studio 使用者存取任何資料來源的能力受與其 SageMaker AI IAM執行角色相關聯的許可所規範。附加的政策可以授權他們從特定 Amazon S3 儲存貯體或字首讀取、寫入或刪除，並連接至 Amazon RDS 資料庫。

存取 AWS Lake Formation 資料

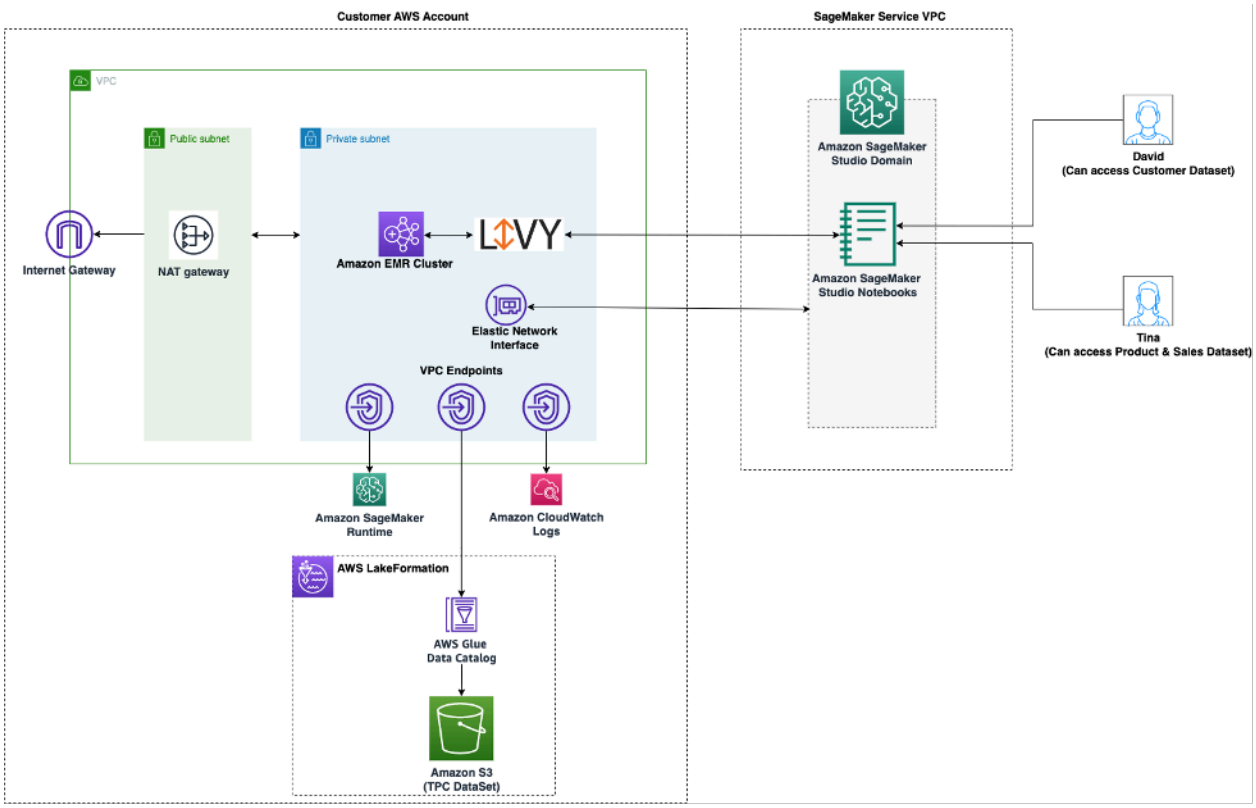
許多企業已開始使用 管理的資料湖AWS Lake Formation，為其使用者啟用精細的資料存取。作為這類受管資料的範例，管理員可以遮蔽某些使用者的敏感資料欄，同時仍然啟用相同基礎資料表的查詢。

若要使用 SageMaker AI Studio 的 Lake Formation，管理員可以將 SageMaker AI IAM執行角色註冊為 `DataLakePrincipals`。如需詳細資訊，請參閱 Lake Formation 許可參考。授權後，有三種主要方法可從 SageMaker AI Studio 存取和寫入受管資料：

1. 從 SageMaker AI Studio 筆記本，使用者可以利用查詢引擎，例如 Amazon Athena 或建置在 boto3 上的程式庫，將資料直接提取到筆記本。AWS SDK for Pandas （ 先前稱為 awswrangler) 是熱門的程式庫。以下是程式碼範例，示範其無縫程度：

```
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 原生連線至 AmazonEMR，大規模讀取和寫入資料。透過使用 Apache Livy 和 Amazon EMR執行期角色， SageMaker AI Studio 已建置原生連線，可讓您將 SageMaker AI 執行IAM角色 （ 或其他授權角色 ） 傳遞至 Amazon EMR叢集以進行資料存取和處理。如需說明，請參閱從 Studio 連線至 Amazon EMR叢集。 up-to-date



從 SageMaker Studio 存取 Lake Formation 管理資料的架構

3. 使用 SageMaker AI Studio 原生連線進行AWS Glue 互動式工作階段，以大規模讀取和寫入資料。SageMaker AI Studio 筆記本具有內建核心，可讓使用者以互動方式在 上執行命令AWS Glue。這可實現 Python、Spark 或 Ray 後端的可擴展使用，這些後端可以從受管資料來源大規模無縫讀取和寫入資料。核心允許使用者傳遞其 SageMaker 執行或其他授權IAM角色。如需詳細資訊，請參閱使用 AWS Glue 互動式工作階段準備資料。

常見護欄

本節討論使用IAM政策、資源政策、VPC端點政策和服務控制政策 ()，在 ML 資源上套用控管最常用的護欄SCPs。

限制筆記本對特定執行個體的存取

此服務控制政策可用來限制資料科學家在建立 Studio 筆記本時可存取的執行個體類型。請注意，任何使用者都需要「系統」執行個體，才能建立託管 SageMaker AI Studio 的預設 Jupyter Server 應用程式。

```
{
```

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

限制不合規的 SageMaker AI Studio 網域

對於 SageMaker AI Studio 網域，以下服務控制政策可用於強制執行流量來存取客戶資源，使其不會透過公有網際網路，而是透過客戶的 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"
                }
            }
        }
    ]
}
```

限制啟動未經授權的 SageMaker AI 映像

下列政策可防止使用者在其網域內啟動未經授權的 SageMaker AI 映像：f

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

僅透過 SageMaker AI VPC端點啟動筆記本

除了 SageMaker AI 控制平面的VPC端點之外， SageMaker AI 還支援VPC端點讓使用者連線到 SageMaker AI Studio 筆記本或 SageMaker AI 筆記本執行個體。如果您已經為 SageMaker AI Studio/筆記本執行個體設定VPC端點，則下列IAM條件索引鍵只會允許透過 SageMaker AI Studio VPC端點或 SageMaker AI API端點建立 SageMaker 與 AI Studio 筆記本的連線。

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

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

限制 SageMaker AI Studio 筆記本存取有限的 IP 範圍

公司通常會將 SageMaker AI Studio 存取限制在特定允許的公司 IP 範圍。下列具有 `SourceIP` 條件索引鍵IAM的政策可以限制這一點。

```
{
    "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"
                    ]
                }
```

```
            }
        }
    ]
}
```

防止 SageMaker AI Studio 使用者存取其他使用者設定檔

身為管理員，當您建立使用者設定檔時，請確定設定檔已使用 SageMaker AI Studio 使用者名稱加上標籤索引鍵 `studiouserid`。委託人（連接至使用者的使用者或角色）也應該有具有金鑰的標籤 `studiouserid`（此標籤可以命名為任何項目，且不限於 `studiouserid`)。

接著，將下列政策連接至使用者在啟動 SageMaker AI Studio 時將擔任的角色。

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

強制標記

資料科學家需要使用 SageMaker AI Studio 筆記本來探索資料，以及建置和訓練模型。將標籤套用至筆記本有助於監控用量和控制成本，並確保擁有權和可稽核性。

對於 SageMaker AI Studio 應用程式，請確保已標記使用者設定檔。標籤會自動從使用者設定檔傳播到應用程式。若要使用標籤強制建立使用者設定檔（透過 CLI和 支援SDK)，請考慮將此政策新增至 管理員角色：

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

對於其他資源，例如訓練任務和處理任務，您可以使用下列政策來強制標記：

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

SageMaker AI Studio 中的根存取

在 SageMaker AI Studio 中，筆記本在 Docker 容器中執行，根據預設，該容器沒有主機執行個體的根存取權。同樣地，除了預設執行身分使用者之外，容器內的所有其他使用者 ID 範圍都會在主機執行個體本身上重新映射為非特殊權限使用者IDs。因此，權限提升的威脅僅限於筆記本容器本身。

建立自訂映像時，您可能想要為使用者提供更嚴格的控制項的非根許可；例如，避免以根目錄執行不理想的程序，或安裝公開可用的套件。在這種情況下，您可以建立映像，以在 Dockerfile 中以非根使用者身分執行。無論您是將使用者建立為根或非根，您都需要確保 UID/GID of the user is identical to the UID/GID中的 AppImageConfig 適用於自訂應用程式的 ，這會為 SageMaker AI 建立使用自訂映像執行應用程式的組態。例如，如果您的 Dockerfile 是為非根使用者建置的，例如：

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

`AppImageConfig` 檔案需要在其 GID中提及相同的 UID和 `KernelGatewayConfig`：

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

自訂映像的可接受 UID/GID 值為 Studio 映像的 0/0 和 1000/100。如需建置自訂映像和相關`AppImageConfig`設定的範例，請參閱此 Github 儲存庫。

為了避免使用者篡改，請勿將 `CreateAppImageConfig`、 `UpdateAppImageConfig`或 `DeleteAppImageConfig`許可授予 SageMaker AI Studio 筆記本使用者。

網路管理

若要設定 SageMaker AI Studio 網域，您需要指定VPC網路、子網路和安全群組。指定 VPC和 子網路時，請務必IPs考量下列各節討論的使用量和預期成長。

VPC 網路規劃

與 SageMaker AI Studio 網域相關聯的客戶VPC子網路必須使用適當的無類別網域間路由 (CIDR) 範圍建立，取決於下列因素：

- 使用者數量。
- 每個使用者的應用程式數量。
- 每個使用者的唯一執行個體類型數目。
- 每個使用者的平均訓練執行個體數量。
- 預期成長百分比。

SageMaker AI 和參與 AWS 服務針對下列使用案例，將彈性網路介面 (ENI) 注入客戶VPC子網路：

- Amazon EFS ENI會為 SageMaker AI 網域的EFS掛載目標注入（連接至 SageMaker AI 網域的每個子網路/可用區域一個 IP)。
- SageMaker AI Studio 會ENI為使用者設定檔或共用空間所使用的每個唯一執行個體注入 。例如：
  - 如果使用者描述檔執行預設的 Jupyter 伺服器應用程式（一個「系統」執行個體）、資料科學應用程式和 Base Python 應用程式（都在`ml.t3.medium`執行個體上執行），Studio 會注入兩個 IP 地址。
  - 如果使用者設定檔執行預設的 Jupyter 伺服器應用程式（一個「系統」執行個體）、Tensorflow GPU 應用程式 (`ml.g4dn.xlarge`執行個體上） 和資料 wrangler 應用程式 (`ml.m5.4xlarge`執行個體上），Studio 會注入三個 IP 地址。
- 針對跨網域VPC子網路/可用區域ENI的每個VPC端點注入（SageMaker AI VPC端點IPs為四個；參與服務VPC端點IPs為 ~6 個ECR，例如 S3、 和 CloudWatch。）
- 如果以相同的VPC組態啟動 SageMaker AI 訓練和處理任務，每個任務每個執行個體都需要兩個 IP 地址。

> **Note**
>
> VPC SageMaker AI Studio 的設定，例如子網路和VPC僅限 流量，不會自動傳遞至從 SageMaker AI Studio 建立的訓練/處理任務。呼叫 Create*Job 時，使用者需要視需要設定 VPC設定和網路隔離APIs。如需詳細資訊，請參閱在無網際網路模式下執行訓練和推論容器。

案例：資料科學家在兩種不同的執行個體類型上執行實驗

在此案例中，假設 SageMaker AI 網域已設定為VPC僅限 流量模式。有VPC端點設定，例如 SageMaker AI API、 SageMaker AI 執行期、Amazon S3 和 Amazon ECR。

資料科學家正在 Studio 筆記本上執行實驗，在兩種不同的執行個體類型（例如 `ml.t3.medium`和 `ml.m5.large`) 上執行，並在每個執行個體類型中啟動兩個應用程式。

假設資料科學家也同時在`ml.m5.4xlarge`執行個體上執行具有相同VPC組態的訓練任務。

在此案例中， SageMaker AI Studio 服務將插入ENIs如下：

表 1：ENIs注入客戶VPC進行實驗案例

實體	目標	ENI 注入	備註	Level
EFS 掛載目標	VPC 子網路	三個	三個 AZs/subnet	網域
VPC 端點	VPC 子網路	30	三個 AZs/subnet，VPCE每個 10 個	網域
Jupyter 伺服器	VPC 子網路	一	每個執行個體一個 IP	使用者
KernelGateway 應用程式	VPC 子網路	Two	每個執行個體類型一個 IP	使用者
培訓	VPC 子網路	Two	IPs 每個訓練執行個體兩個	使用者

實體	目標	ENI 注入	備註	Level
			如果使用 EFA ，則IPs每個訓練執行個體 5 個	

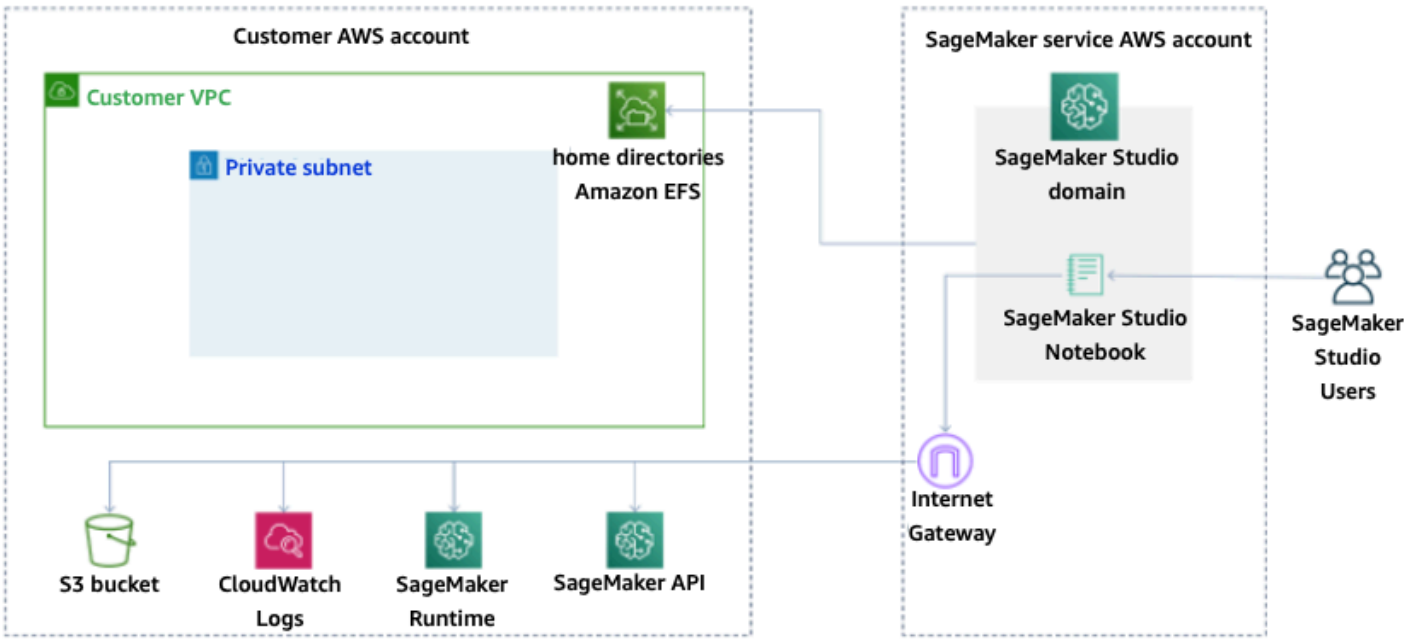
在此案例中，客戶總共有 38 個IPs耗用量VPC，其中 33 IPs 個是在網域層級與使用者共用，而 5 個IPs是在使用者層級耗用。如果您有 100 名使用者在此網域中具有類似的使用者描述檔，同時執行這些活動，則除了網域層級 IP 消耗之外，還會IPs在使用者層級消耗 5 x 100 = 500，也就是IPs每個子網路 11 個，總共 511 個IPs。在此案例中，您需要CIDR使用 /22 建立VPC子網路，以配置 1024 個 IP 地址，並擁有成長空間。

VPC 網路選項

SageMaker AI Studio 網域支援使用下列其中一個選項來設定VPC網路：

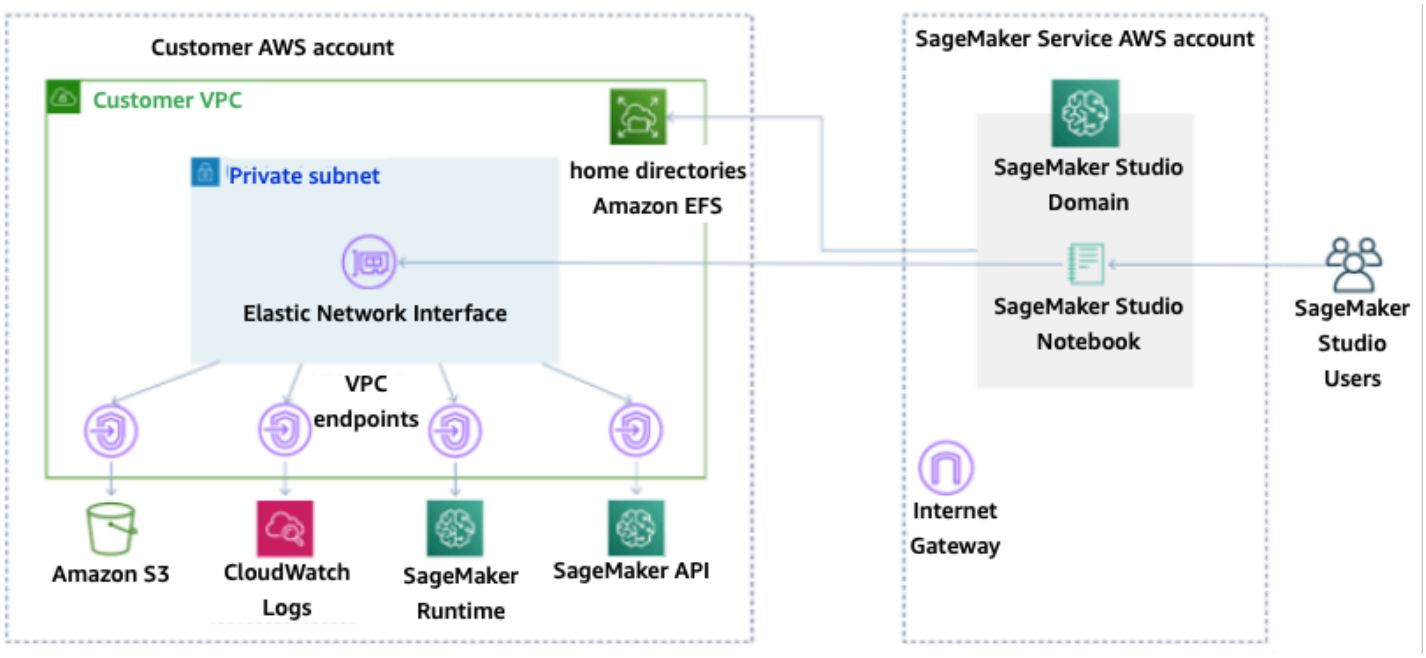
- 僅限公有網際網路
- 僅限 VPC

僅限公有網際網路選項允許 SageMaker AI API服務透過 AI 服務帳戶管理VPC的 中佈建的網際網路閘道使用公有 SageMaker 網際網路，如下圖所示：



預設模式：透過 SageMaker AI 服務帳戶存取網際網路

VPC 唯一的選項會停用 SageMaker AI 服務帳戶VPC管理之 的網際網路路由，並允許客戶設定要透過VPC端點路由的流量，如下圖所示：



VPC 僅限 模式：無法透過 SageMaker AI 服務帳戶存取網際網路

對於VPC僅以 模式設定的網域，為每個使用者設定檔設定安全群組，以確保完全隔離基礎執行個體。AWS 帳戶中的每個網域都可以有自己的VPC組態和網際網路模式。如需設定VPC網路組態的詳細資訊，請參閱 中的VPC將 SageMaker AI Studio 筆記本連接至外部資源。

限制

- 建立 SageMaker AI Studio 網域之後，您就無法將新的子網路與網域建立關聯。
- 網路VPC類型 (僅限公有網際網路或VPC僅限 ） 無法變更。

資料保護

在建構 ML 工作負載之前，應該已備妥會影響安全性的基礎實務。例如，資料分類提供一種方法，根據敏感程度對資料進行分類，而加密會透過讓資料無法辨識存取來保護資料。這些方法很重要，因為它們支援防止處理不當或遵守法規義務等目標。

SageMaker AI Studio 提供多種功能來保護靜態和傳輸中的資料。不過，如AWS 共享責任模型所述，客戶有責任控制 AWS 在全球基礎設施上託管的內容。在本節中，我們會說明客戶如何使用這些功能來保護其資料。

保護靜態資料

為了保護您的 SageMaker AI Studio 筆記本以及模型建置資料和模型成品， SageMaker AI 會加密筆記本，以及來自訓練和批次轉換任務的輸出。 SageMaker AI 預設會使用 AWS Amazon S3 的 Managed Key 來加密這些筆記本。Amazon S3 的此 AWS 受管金鑰無法共用以進行跨帳戶存取。針對跨帳戶存取，請在建立 SageMaker AI 資源時指定客戶受管金鑰，以便共用以跨帳戶存取。

使用 SageMaker AI Studio，資料可以存放在下列位置：

- S3 儲存貯體 – 啟用可共用筆記本時， SageMaker AI Studio 會在 S3 儲存貯體中共用筆記本快照和中繼資料。
- EFS 磁碟區 – SageMaker AI Studio 會將磁碟EFS區連接至您的網域，以存放筆記本和資料檔案。即使刪除網域，此EFS磁碟區仍會持續存在。
- EBS 磁碟區 – EBS 連接至筆記本執行所在的執行個體。此磁碟區會在執行個體期間持續存在。

使用 加密靜態 AWS KMS

- 您可以傳遞AWS KMS 金鑰來加密連接至筆記本、訓練、調校、批次轉換任務和端點的EBS磁碟區。
- 如果您未指定KMS金鑰， SageMaker AI 會使用系統管理的KMS金鑰來加密作業系統 (OS) 磁碟區和 ML 資料磁碟區。
- 基於合規原因而需要使用KMS金鑰加密的敏感資料，應該存放在 ML 儲存磁碟區或 Amazon S3 中，這兩者都可以使用您指定的KMS金鑰進行加密。

保護傳輸中的資料

SageMaker AI Studio 可確保 ML 模型成品和其他系統成品在傳輸中和靜態時加密。對 SageMaker AI API和主控台的請求是透過安全的 (SSL) 連線提出。有些內部網路傳輸中資料 (服務平台內部) 未加密。其中包含：

- 服務控制平面與訓練任務執行個體 (不是客戶資料) 之間的命令與控制通訊。
- 分散式處理和訓練任務中節點之間的通訊 （網路內）。

不過，您可以選擇加密訓練叢集中節點之間的通訊。啟用包含所有容器的流量加密可能會增加訓練時間，特別是使用分散式深入學習演算法時。

根據預設，Amazon SageMaker AI 會在 Amazon 中執行訓練任務VPC，以協助確保資料的安全。您可以新增另一個層級的安全性，透過設定私有來保護訓練容器和資料VPC。此外，您可以將 SageMaker AI Studio 網域設定為VPC僅以 模式執行，並設定VPC端點以透過私有網路路由流量，而不會透過網際網路輸出流量。

資料保護護欄

加密靜態 SageMaker AI 託管磁碟區

使用下列政策在託管 SageMaker AI 端點以進行線上推論時強制執行加密：

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

```
 }
```

加密模型監控期間使用的 S3 儲存貯體

Model Monitoring 會擷取傳送至 SageMaker AI 端點的資料，並將其存放在 S3 儲存貯體中。當您設定資料擷取組態時，您需要加密 S3 儲存貯體。目前沒有對此的補償性控制。

除了擷取端點輸出之外，模型監控服務還會檢查是否有偏離預先指定的基準。您需要加密輸出和用於監控偏離的中繼儲存磁碟區。

```
{
   "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 AI Studio 網域儲存磁碟區

強制加密連接到 Studio 網域的儲存磁碟區。此政策需要使用者提供 CMK 來加密連接到 Studio 網域的儲存磁碟區。

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

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

加密存放在 S3 中用來共用筆記本的資料

這是用來加密儲存貯體中儲存的任何資料的政策，用於在 SageMaker AI Studio 網域中的使用者之間共用筆記本：

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

限制

- 建立網域後，您就無法使用自訂 AWS KMS 金鑰更新連接的EFS磁碟區儲存體。

- 建立金鑰後，您就無法使用KMS金鑰更新訓練/處理任務或端點組態。

日誌記錄和監控

為了協助您偵錯編譯任務、處理任務、訓練任務、端點、轉換任務、筆記本執行個體和筆記本執行個體生命週期組態、演算法容器、模型容器或筆記本執行個體生命週期組態傳送至 stdout 或 stderr 的任何內容，也會傳送至 Amazon CloudWatch Logs。您可以使用 Amazon 監控 SageMaker AI Studio CloudWatch，它會收集原始資料並將其處理為可讀且近乎即時的指標。這些統計資料會保留 15 個月，因此您可以存取歷史資訊，並更深入了解 Web 應用程式或服務的效能。

使用 記錄 CloudWatch

由於資料科學程序本質上是實驗性和反覆性的，因此記錄活動至關重要，例如筆記本使用、訓練/處理任務執行時間、訓練指標，以及提供叫用延遲等指標的端點。根據預設， SageMaker AI 會將指標發佈至 CloudWatch 日誌，而這些日誌可以使用客戶管理的金鑰來加密 AWS KMS。

您也可以使用VPC端點將日誌傳送到 ， CloudWatch 而無需使用公有網際網路。您也可以設定留意特定閾值的警示，當滿足這些閾值時傳送通知或採取動作。如需詳細資訊，請參閱 Amazon CloudWatch 使用者指南。

SageMaker AI 會在 下為 Studio 建立單一日誌群組`/aws/sagemaker/studio`。每個使用者設定檔和應用程式都有自己的日誌串流在此日誌群組下，生命週期組態指令碼也有自己的日誌串流。例如，名為「studio-user」的使用者描述檔具有 Jupyter Server 應用程式和連接的生命週期指令碼，而 Data Science Kernel Gateway 應用程式具有下列日誌串流：

`/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`

若要讓 SageMaker AI CloudWatch 代表您將日誌傳送至 ，Training/Processing/Transform任務的發起人APIs將需要下列許可：

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

若要使用自訂 AWS KMS 金鑰加密這些日誌，您必須先修改金鑰政策，以允許 CloudWatch 服務加密和解密金鑰。建立日誌加密 AWS KMS 金鑰後，請修改金鑰政策以包含下列項目：

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

```
    ]
}
```

請注意，您可以隨時使用 ArnEquals，並提供您要加密之 CloudWatch 日誌的特定 Amazon Resource Name (ARN)。在這裡，我們顯示您可以使用此金鑰來加密 帳戶中的所有日誌，以簡化操作。此外，訓練、處理和模型端點會發佈有關執行個體CPU和記憶體使用率的指標、託管調用延遲等。您可以進一步設定 AmazonSNS，在特定閾值超過時通知管理員事件。訓練和處理的取用者APIs需要具有下列許可：

```
{
     "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 稽核 AWS CloudTrail

若要改善您的合規狀態，APIs請使用 稽核您的所有 AWS CloudTrail。根據預設，所有 SageMaker AI APIs都會使用 記錄AWS CloudTrail。您不需要任何其他IAM許可即可啟用 CloudTrail。

除了 `InvokeEndpoint`和 之外，所有 SageMaker AI 動作都會由 記錄`InvokeEndpointAsync`，CloudTrail 並記錄在 操作中。例如，對 `CreateTrainingJob`、 `CreateEndpoint`和`CreateNotebookInstance`動作的呼叫會在 CloudTrail 日誌檔案中產生項目。

每個 CloudTrail 事件項目都包含產生請求者的相關資訊。身分資訊可協助您判斷下列事項：

- 該請求是否使用根或 AWS IAM 使用者登入資料提出。
- 提出該請求時，是否使用了特定角色或聯合身分使用者的暫時安全憑證。
- 請求是否由其他 AWS 服務提出。如需範例事件，請參閱使用文件記錄 SageMaker AI API呼叫 CloudTrail。

根據預設， 會將使用者設定檔的 Studio 執行角色名稱 CloudTrail 記錄為每個事件的識別符。如果每個使用者都有自己的執行角色，這就有效。如果多個使用者共用相同的執行角色，您可以使用`sourceIdentity` 組態將 Studio 使用者設定檔名稱傳播到其中 CloudTrail。請參閱從 Amazon SageMaker AI Studio 監控使用者資源存取權以啟用 `sourceIdentity`功能。在共用空間中，所有動作都將空間ARN稱為來源，您無法透過 稽核`sourceIdentity`。

成本屬性

SageMaker AI Studio 內建功能，可協助管理員追蹤其個別網域、共用空間和使用者的花費。

自動化標記

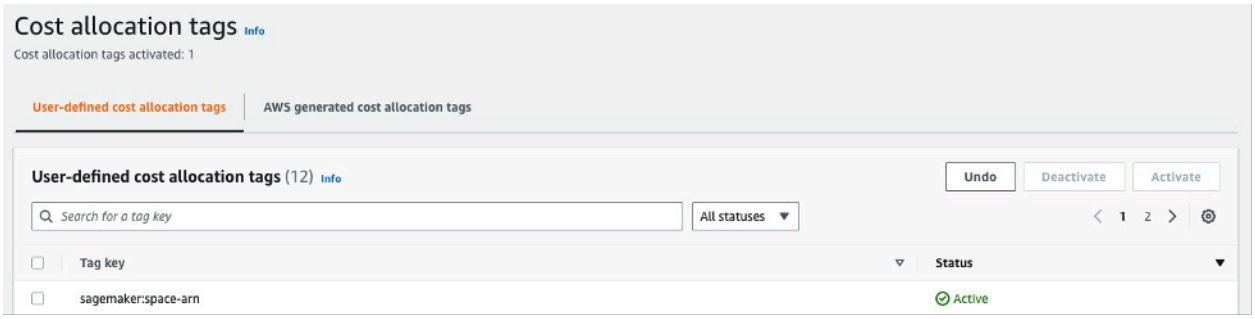
SageMaker AI Studio 現在會自動使用其各自的 來標記新 SageMaker 資源，例如訓練任務、處理任務和核心應用程式`sagemaker:domain-arn`。在更精細的層級上， SageMaker AI 也會使用`sagemaker:user-profile-arn`或 標記資源`sagemaker:space-arn`，以指定資源的主要建立者。

SageMaker AI 網域EFS磁碟區會以名為 `ManagedByAmazonSageMakerResource` 的索引鍵標記網域 的值ARN。它們沒有精細的標籤來了解每個使用者層級上的空間用量。不過，管理員可以將EFS磁碟區連接至EC2執行個體以進行自訂監控。

成本監控

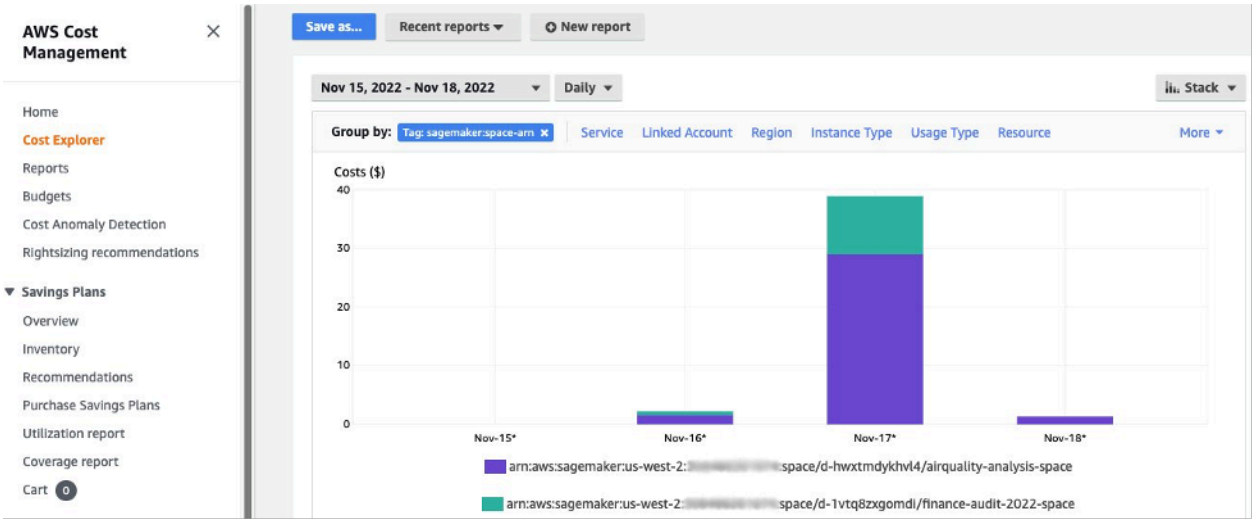
自動化標籤可讓管理員透過 AWS Cost Explorer和 等 out-of-the-box解決方案，以及根據成本和用量報告 () 中的資料建置的自訂解決方案，來追蹤AWS Budgets、報告和監控您的 ML 支出CURs。 AWS

若要使用連接的標籤進行成本分析，必須先在 AWS Billing 主控台的成本分配標籤區段中啟用標籤。標籤最多可能需要 24 小時才會出現在成本分配標籤面板中，因此您需要在啟用前先建立 SageMaker AI 資源。



在 Cost Explorer 上ARN啟用空間做為成本分配標籤

在您啟用成本分配標籤後， AWS 會開始追蹤您的標記資源，並在 24-48 小時後，標籤會在成本總管中顯示為可選取的篩選條件。



依範例網域的共用空間分組的成本

成本控制

當第一個 SageMaker AI Studio 使用者加入時， SageMaker AI 會為網域建立EFS磁碟區。由於筆記本和資料檔案存放在使用者的主目錄中，因此此EFS磁碟區會產生儲存成本。當使用者啟動 Studio 筆記本時，會針對執行筆記本的運算執行個體啟動它們。如需成本的詳細明細，請參閱 Amazon SageMaker AI 定價。

管理員可以透過指定使用者可以啟動的執行個體清單來控制運算成本，方法是使用通用護欄區段中所述IAM的政策。此外，我們建議客戶使用 SageMaker AI Studio 自動關閉擴充功能，透過自動關閉閒置的應用程式來節省成本。此伺服器延伸會定期輪詢每個使用者設定檔執行的應用程式，並根據管理員設定的逾時關閉閒置的應用程式。

若要為網域中的所有使用者設定此延伸，您可以使用生命週期組態，如自訂一節所述。此外，您也可以使用延伸檢查程式，確保網域的所有使用者都已安裝延伸模組。

自訂

生命週期組態

生命週期組態是由 SageMaker AI Studio 生命週期事件啟動的 shell 指令碼，例如啟動新的 SageMaker AI Studio 筆記本。您可以使用這些 shell 指令碼來自動化 SageMaker AI Studio 環境的自訂，例如安裝自訂套件、用於自動關閉非作用中筆記本應用程式的 Jupyter 延伸模組，以及設定 Git 組態。如需如何建置生命週期組態的詳細說明，請參閱此部落格：使用生命週期組態自訂 Amazon SageMaker AI Studio。

SageMaker AI Studio 筆記本的自訂映像

Studio 筆記本隨附一組預先建置的影像，其中包含 Amazon SageMaker AI Python SDK 和最新版本的IPython執行時間或核心。使用此功能，您可以將自己的自訂映像帶入 Amazon SageMaker AI 筆記本。然後，這些映像可供通過網域身分驗證的所有使用者使用。

開發人員和資料科學家可能需要針對數個不同的使用案例進行自訂映像：

- 存取特定或最新版本的熱門 ML 架構 TensorFlow，例如 PyTorchMXNet、 或其他。
- 將本機開發的自訂程式碼或演算法帶入 SageMaker AI Studio 筆記本，以進行快速迭代和模型訓練。
- 透過 存取資料湖或內部部署資料存放區APIs。管理員需要在映像中包含對應的驅動程式。
- 存取後端執行時間 （ 也稱為核心 ），但 IPython (例如 R、Julia 或其他) 除外。您也可以使用概述的方法來安裝自訂核心。

如需如何建置自訂映像的詳細指示，請參閱建立自訂 SageMaker AI 映像。

JupyterLab 擴充功能

透過 SageMaker AI Studio JuypterLab 3 筆記本，您可以利用不斷成長的開放原始碼 JupyterLab延伸社群。本節重點介紹一些自然適合 SageMaker AI 開發人員工作流程的案例，但建議您瀏覽可用的延伸模組，甚至建立您自己的延伸模組。

JupyterLab 3 現在可大幅簡化封裝和安裝擴充功能的程序。您可以透過 bash 指令碼安裝上述擴充功能。例如，在 SageMaker AI Studio 中，從 Studio 啟動器開啟系統終端機，然後執行下列命令。此外，您可以使用生命週期組態自動安裝這些擴充功能，以便在 Studio 重新啟動之間保留這些擴充功能。您可以為網域或個別使用者層級的所有使用者設定此項目。

例如，若要安裝 Amazon S3 檔案瀏覽器的擴充功能，請在系統終端機中執行下列命令，並確認重新整理瀏覽器：

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

如需延伸管理的詳細資訊，包括如何寫入適用於 JupyterLab 筆記本第 1 版和第 3 版的生命週期組態，以實現回溯相容性，請參閱安裝 JupyterLab 和 Jupyter 伺服器延伸。

Git 儲存庫

SageMaker AI Studio 已預先安裝 Jupyter Git 延伸模組，讓使用者輸入 Git 儲存庫URL的自訂內容、將其複製到您的EFS目錄、推送變更，以及檢視遞交歷史記錄。管理員可以在網域層級設定建議的 git 儲存庫，以便顯示為最終使用者的下拉式清單選項。如需說明，請參閱將建議的 Git 儲存庫連接至 Studio up-to-date。

如果儲存庫是私有的，則延伸項目會要求使用者使用標準 git 安裝，將其登入資料輸入終端機。或者，使用者可以將 SSH 登入資料存放在其個別EFS目錄，以便於管理。

Conda 環境

SageMaker AI Studio 筆記本使用 Amazon EFS作為持久性儲存層。資料科學家可以使用持久性儲存來建立自訂 conda 環境，並使用這些環境來建立核心。這些核心由 支援EFS，而且在核心、應用程式或 Studio 重新啟動之間會持續存在。Studio 會自動將所有有效的環境選為 KernelGateway核心。

對於資料科學家來說，建立 conda 環境的程序很簡單，但核心在核心選擇器上填入大約需要一分鐘的時間。若要建立環境，請在系統終端機中執行下列動作：

```
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
```

如需詳細說明，請參閱 Amazon Studio 筆記本中管理 Python 套件的四種方法中 Studio EFS磁碟區區段中的 Persist Conda 環境。 SageMaker

結論

在本白皮書中，我們檢閱了幾個跨領域的最佳實務，例如操作模型、網域管理、身分管理、許可管理、網路管理、記錄、監控和自訂，讓平台管理員能夠設定和管理 SageMaker AI Studio 平台。

附錄

多租戶比較

表 2 — 多租戶比較

多網域	多帳戶	單一網域內的屬性型存取控制 (ABAC)
使用標籤來達成資源隔離。SageMaker AI Studio 會自動使用網域ARN和使用者設定檔/空間 標記所有資源ARN。	每個租戶都在自己的帳戶中，因此有絕對的資源隔離。	使用標籤來達成資源隔離。使用者必須管理 建立的資源標記ABAC。
清單APIs不能受到標籤的限制。資源的 UI 篩選是在共用空間上完成，不過，列出透過 AWS CLI 或 Boto3 進行的API呼叫SDK會列出整個區域的資源。	也可以APIs隔離清單，因為租戶位於其專用帳戶中。	清單APIs不能受到標籤的限制。列出透過 AWS CLI 或 Boto3 進行的API呼叫，SDK會列出整個區域的資源。
SageMaker AI Studio 運算和每個租用戶的儲存成本，可以透過使用網域ARN做為成本分配標籤來輕鬆監控。	SageMaker 每個租戶的 AI Studio 運算和儲存成本易於使用專用帳戶進行監控。	SageMaker 每個租用戶的 AI Studio 運算成本必須使用自訂標籤計算。 SageMaker AI Studio 儲存成本無法監控每個網域，因為所有租戶共用相同的EFS磁碟區。
服務配額是在帳戶層級設定，因此單一租戶仍然可以使用所有資源。	服務配額可以在每個租戶的帳戶層級設定。	服務配額是在帳戶層級設定，因此單一租戶仍然可以使用所有資源。
擴展到多個租戶可以透過基礎設施做為程式碼 (IaC) 或服務目錄來實現。	擴展到多個租戶涉及組織和販賣多個帳戶。	擴展需要每個新租用戶的租用戶特定角色，且使用者設定檔需要手動標記租用戶名稱。

多網域	多帳戶	單一網域內的屬性型存取控制 (ABAC)
租用戶內使用者之間的協作可以透過共用空間進行。	租用戶內使用者之間的協作可以透過共用空間進行。	所有租用戶將可存取相同的共用空間以進行協作。

SageMaker AI Studio 網域備份和復原

如果意外EFS刪除，或由於聯網或身分驗證的變更而需要重新建立網域，請遵循這些指示。

選項 1：使用 從現有的 備份 EFS EC2

SageMaker Studio 網域備份

1. 列出 SageMaker Studio (CLI、) 中的使用者設定檔和空格SDK。
2. 在 UIDs上將使用者設定檔/空間映射至 EFS。
   a. 對於 users/spaces, describe the user profile/space(CLI、) 清單中的每個使用者SDK。
   b. 將使用者設定檔/空間映射至 `HomeEfsFileSystemUid`。
   c. `UserSettings['ExecutionRole']` 如果使用者具有不同的執行角色，請將使用者設定檔映射至 。
   d. 識別預設空間執行角色。
3. 建立新的網域並指定預設空間執行角色。
4. 建立使用者設定檔和空格。
   - 對於使用者清單中的每個使用者，使用執行角色映射建立使用者設定檔 (CLI、SDK)。
5. 為新的 EFS和 建立映射UIDs。
   a. 對於使用者清單中的每個使用者，描述使用者設定檔 (CLI、SDK)。
   b. 將使用者設定檔映射至 `HomeEfsFileSystemUid`。
6. 或者，刪除所有應用程式、使用者設定檔、空格，然後刪除網域。

EFS 備份

若要備份 EFS，請使用下列指示：

1. 啟動EC2執行個體，並將舊 SageMaker Studio 網域的傳入/傳出安全群組連接到新EC2執行個體（允許連接埠 2049 TCP上的NFS流量通過。請參閱 中的將 SageMaker Studio 筆記本VPC連接至外部資源。
2. 將 SageMaker Studio EFS磁碟區掛載至新的EC2執行個體。請參閱掛載EFS檔案系統。
3. 將檔案複製到EBS本機儲存體： `>sudo cp -rp /efs /studio-backup:`
   a. 將新的網域安全群組連接至EC2執行個體。
   b. 將新EFS磁碟區掛載至EC2執行個體。
   c. 將檔案複製到新EFS磁碟區。
   d. 對於使用者集合中的每個使用者：
      i. 建立目錄：`mkdir new_uid`。
      ii. 將檔案從舊UID目錄複製到新UID目錄。
      iii. 變更所有檔案的擁有權：`chown <new_UID>`針對所有檔案。

選項 2：EFS使用 S3 和生命週期組態從現有備份

1. 請參閱使用 Amazon Linux 2 將工作遷移至 Amazon SageMaker 筆記本執行個體。
2. 建立 S3 儲存貯體以進行備份（例如 `>studio-backup`。
3. 列出具有執行角色的所有使用者設定檔。
4. 在目前的 SageMaker Studio 網域中，在網域層級設定預設LCC指令碼。
   - 在 中LCC，`/home/sagemaker-user`將 中的所有內容複製到 S3 中的使用者設定檔字首（例如，`s3://studio-backup/studio-user1`)。
5. 重新啟動所有預設的 Jupyter Server 應用程式（以便LCC執行）。
6. 刪除所有應用程式、使用者設定檔和網域。
7. 建立新的 SageMaker Studio 網域。
8. 從使用者設定檔和執行角色清單中建立新的使用者設定檔。
9. LCC 在網域層級設定 ：
   - 在 中LCC，將 S3 中使用者設定檔字首中的所有內容複製到 `/home/sagemaker-user`
10為具有LCC組態的所有使用者建立預設 Jupyter Server 應用程式 (CLI、SDK)。

SageMaker 使用SAML聲明的 Studio 存取

解決方案設定：

1. 在外部 IdP 中建立SAML應用程式。
2. 在 中將外部 IdP 設定為身分提供者IAM。
3. 建立可由 IdP 存取的 `SAMLValidator` Lambda 函數 （透過函數URL或API閘道）。
4. 建立 `GeneratePresignedUrl` Lambda 函數和API閘道以存取函數。
5. 建立IAM使用者可以擔任的角色來叫用API閘道。此角色應以宣告SAML形式以下列格式做為屬性傳遞：
   - 屬性名稱： https://aws.amazon.com/SAML/Attributes/Role
   - 屬性值：`<IdentityProviderARN>`、 `<RoleARN>`
6. 將 SAML Assertion Consumer Service (ACS) 端點更新為`SAMLValidator`叫用 URL。

SAML 驗證器範例程式碼：

```
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
```

深入閱讀

- 在 （部落格） 上設定安全且受管良好的機器學習環境 AWSAWS
- 為團隊和群組設定 Amazon SageMaker AI Studio，並完全隔離資源 (AWS 部落格）
- 使用 AWS SSO 和 Okta Universal Directory 加入 Amazon SageMaker AI Studio (AWS 部落格）
- 如何為 AWS 帳戶聯合設定 SAML 2.0 (Okta 文件）
- 在 上建置 Secure Enterprise Machine Learning 平台 AWS (AWS 技術指南）
- 使用生命週期組態自訂 Amazon SageMaker AI Studio (AWS 部落格）
- 將您自己的自訂容器映像帶至 Amazon SageMaker AI Studio 筆記本 (AWS 部落格）
- 建置自訂 SageMaker AI 專案範本 – 最佳實務 (AWS 部落格）
- 使用 Amazon SageMaker AI Pipelines 進行多帳戶模型部署 (AWS 部落格）
- 第 1 部分： NatWest 群組如何建置可擴展、安全且永續的MLOps平台 (AWS 部落格）
- 安全的 Amazon SageMaker AI Studio 預先簽署第 1 URLs部分：基礎基礎設施 (AWS 部落格）

貢獻者

本文件的貢獻者包括：

- Ram Vittal，Amazon Web Services ML 解決方案架構師
- Amazon Web Services ML Solutions Architect Sean Morgan
- Durga Sury，Amazon Web Services ML Solutions 架構師

特別感謝下列貢獻想法、修訂和觀點的人員：

- Alessandro Cerè，Amazon Web Services AI/ML 解決方案架構師
- Sumit Thakur，Amazon Web Services SageMaker AI 產品領導者
- Han Zhang，Amazon Web Services 資深軟體開發工程師
- Amazon Web Services、Amazon Web Services 軟體開發工程師 Bhadrinath Pani

文件修訂

若要收到有關此白皮書更新的通知，請訂閱 RSS 摘要。

變更	描述	日期
白皮書已更新	斷開的鏈接固定和許多編輯變化。	2023 年 4 月 25 日
初始出版	白皮書已發佈。	2022 年 10 月 19 日

注意

客戶有責任自行對本文件中的資訊進行獨立評估。本文件：(a) 僅供參考，(b) 代表目前的AWS產品供應項目和做法，如有變更，恕不另行通知，且 (c) 不會向其關聯公司、供應商或授權人建立任何承諾或保證。AWS AWS產品或服務係依「原狀」提供，不含任何明示或暗示之擔保、陳述或條件。客戶的責任和責任由AWS協議控制，本文件不屬於與客戶之間AWS的任何協議的一部分，也不會修改。AWS

AWS 詞彙表

如需最新的 AWS 術語，請參閱《AWS 詞彙表 參考》中的 AWS 詞彙表。

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。