AWS Livre blanc

SageMaker Bonnes pratiques en matière d'administration du studio

SageMaker Bonnes pratiques en matière d'administration du studio: AWS Livre blanc

Table of Contents

SageMaker Bonnes pratiques en matière d'administration du studio

Date de publication : 25 avril 2023 (Révisions du document)

Résumé

Amazon SageMaker Studio fournit une interface visuelle Web unique dans laquelle vous pouvez effectuer toutes les étapes de développement du machine learning (ML), ce qui améliore la productivité des équipes de data science. SageMaker Studio vous offre un accès, un contrôle et une visibilité complets sur chaque étape requise pour créer, entraîner et évaluer des modèles.

Dans ce livre blanc, nous abordons les meilleures pratiques sur des sujets tels que le modèle d'exploitation, la gestion des domaines, la gestion des identités, la gestion des autorisations, la gestion du réseau, la journalisation, la surveillance et la personnalisation. Les meilleures pratiques décrites ici sont destinées au déploiement d'Enterprise SageMaker Studio, y compris les déploiements multi-locataires. Ce document est destiné aux administrateurs de plateformes ML, aux ingénieurs ML et aux architectes ML.

Êtes-vous Well-Architected ?

Le AWSWell-Architected Framework vous aide à comprendre les avantages et les inconvénients des décisions que vous prenez lors de la création de systèmes dans le cloud. Les six piliers du cadre vous permettent d'apprendre les meilleures pratiques architecturales pour concevoir et exploiter des systèmes fiables, sécurisés, efficaces, rentables et durables. À l'aide du AWS Well-Architected Tool, disponible gratuitement dans le AWS Management Console, vous pouvez évaluer votre charge de travail par rapport à ces meilleures pratiques en répondant à une série de questions pour chaque pilier.

Dans le Machine Learning Lens, nous nous concentrons sur la manière de concevoir, déployer et structurer vos charges de travail d'apprentissage automatique dans leAWS Cloud. Cet objectif s'ajoute aux meilleures pratiques décrites dans le Well-Architected Framework.

Introduction

Lorsque vous administrez SageMaker Studio en tant que plate-forme de ML, vous avez besoin de conseils sur les meilleures pratiques pour prendre des décisions éclairées afin de vous aider

à adapter votre plate-forme de ML à mesure que vos charges de travail augmentent. Pour le provisionnement, l'opérationnalisation et le dimensionnement de votre plateforme ML, tenez compte des points suivants :

- Choisissez le bon modèle d'exploitation et organisez vos environnements de machine learning pour atteindre vos objectifs commerciaux.
- Choisissez comment configurer l'authentification de domaine SageMaker Studio pour les identités des utilisateurs et tenez compte des limites au niveau du domaine.
- Décidez comment fédérer l'identité et l'autorisation de vos utilisateurs à la plateforme ML pour des contrôles d'accès et des audits précis.
- Envisagez de configurer des autorisations et des garde-fous pour les différents rôles de vos personas ML.
- Planifiez la topologie de votre réseau de cloud privé virtuel (VPC) en tenant compte de la sensibilité de votre charge de travail ML, du nombre d'utilisateurs, des types d'instances, des applications et des tâches lancées.
- Classez et protégez vos données au repos et en transit grâce au chiffrement.
- Réfléchissez à la manière de consigner et de surveiller les différentes interfaces de programmation d'applications (API) et les activités des utilisateurs à des fins de conformité.
- Personnalisez l'expérience du bloc-notes SageMaker Studio avec vos propres images et scripts de configuration du cycle de vie.

Modèle d'exploitation

Un modèle opérationnel est un cadre qui réunit les personnes, les processus et les technologies pour aider une organisation à générer de la valeur commerciale de manière évolutive, cohérente et efficace. Le modèle opérationnel ML fournit un processus de développement de produits standard pour les équipes de l'organisation. Il existe trois modèles de mise en œuvre du modèle opérationnel, en fonction de la taille, de la complexité et des facteurs commerciaux :

- Équipe de science des données centralisée — Dans ce modèle, toutes les activités de science des données sont centralisées au sein d'une seule équipe ou organisation. Ce modèle est similaire au modèle du centre d'excellence (COE), dans lequel toutes les unités commerciales font appel à cette équipe pour des projets de science des données.
- Équipes de science des données décentralisées — Dans ce modèle, les activités de science des données sont réparties entre différentes fonctions ou divisions commerciales, ou basées sur différentes gammes de produits.
- Équipes de data science fédérées — Dans ce modèle, les fonctions de services partagés telles que les référentiels de code, les pipelines d'intégration continue et de livraison continue (CI/CD), etc. sont gérées par l'équipe centralisée, et chaque unité commerciale ou fonction au niveau du produit est gérée par des équipes décentralisées. Cela est similaire au modèle hub and spoke, dans lequel chaque unité commerciale dispose de ses propres équipes de science des données ; toutefois, ces équipes coordonnent leurs activités avec l'équipe centralisée.

Avant de décider de lancer votre premier domaine de studio pour des cas d'utilisation en production, réfléchissez à votre modèle d'exploitation et aux AWS meilleures pratiques en matière d'organisation de votre environnement. Pour plus d'informations, reportez-vous à la section Organisation de votre AWS environnement à l'aide de plusieurs comptes.

La section suivante fournit des conseils sur l'organisation de votre structure de compte pour chacun des modèles opérationnels.

Structure de compte recommandée

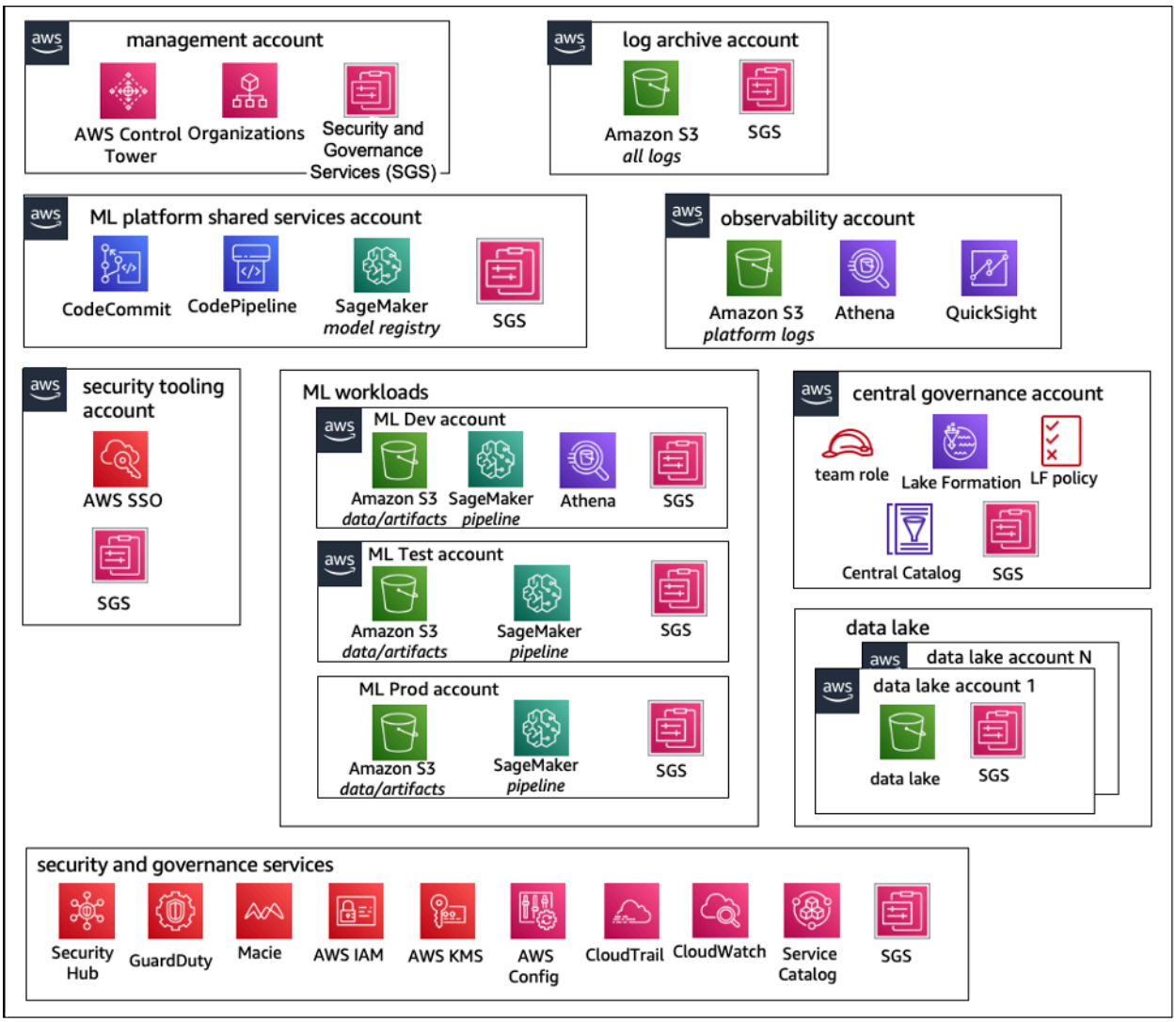
Dans cette section, nous présentons brièvement un modèle de structure de compte opérationnel que vous pouvez utiliser au départ et modifier en fonction des exigences opérationnelles de votre organisation. Quel que soit le modèle d'exploitation que vous choisissez, nous vous recommandons de mettre en œuvre les meilleures pratiques courantes suivantes :

- AWS Control TowerÀ utiliser pour la configuration, la gestion et la gouvernance de vos comptes.
- Centralisez vos identités auprès de votre fournisseur d'identité (IdP) AWSet d'IAM Identity Center avec un compte Securitiy Tooling à administrateur délégué et sécurisez l'accès aux charges de travail.
- Exécutez les charges de travail ML en isolant les charges de travail de développement, de test et de production au niveau du compte.
- Diffusez les journaux de charge de travail ML vers un compte d'archive de journaux, puis filtrez et appliquez une analyse des journaux dans un compte d'observabilité.
- Gérez un compte de gouvernance centralisé pour le provisionnement, le contrôle et l'audit de l'accès aux données.
- Intégrez des services de sécurité et de gouvernance (SGS) dotés de dispositifs de prévention et de détection appropriés dans chaque compte afin de garantir la sécurité et la conformité, conformément aux exigences de votre organisation et de votre charge de travail.

Modèle de structure de compte centralisé

Dans ce modèle, l'équipe de la plateforme ML est chargée de fournir :

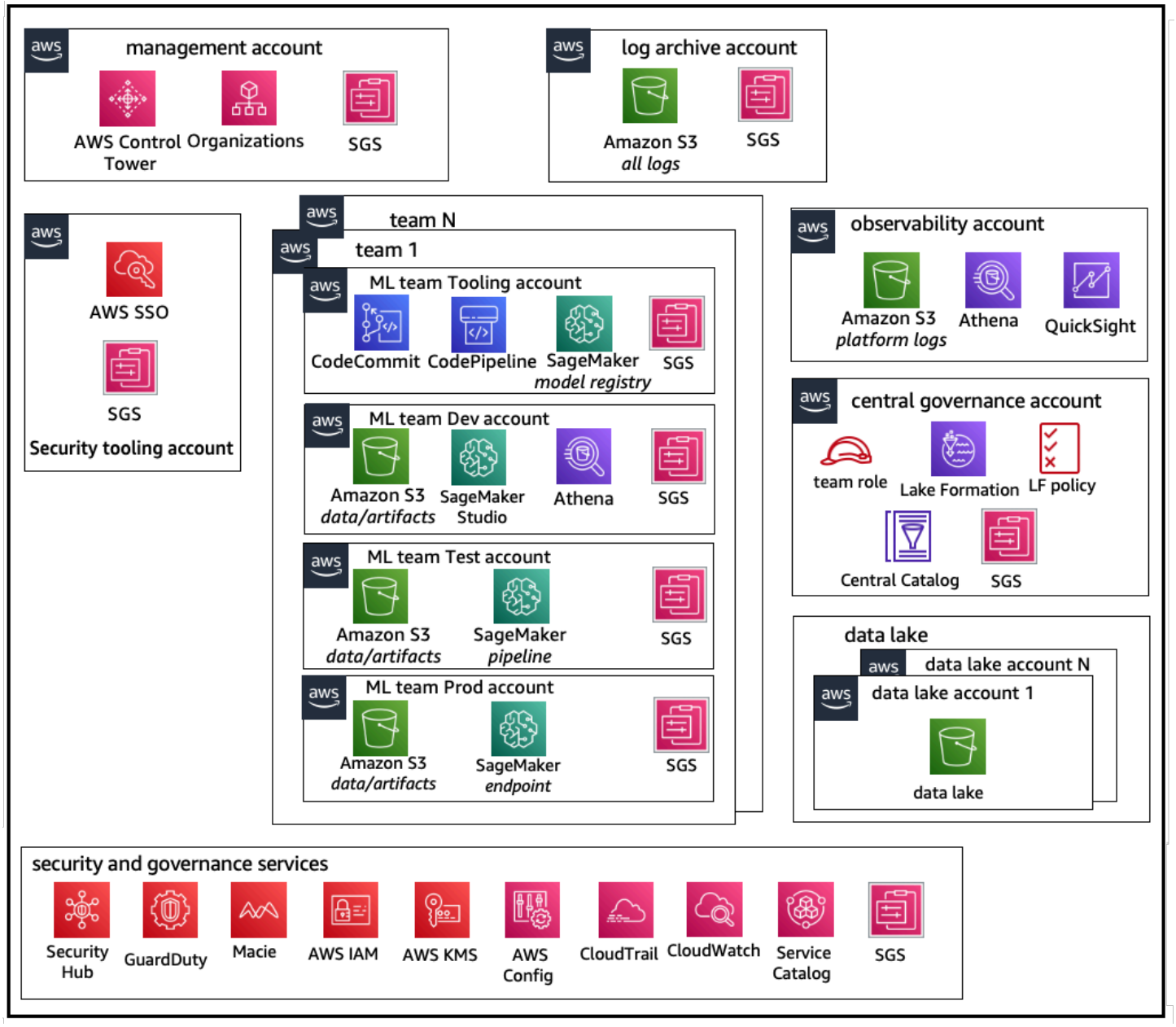
- Un compte d'outillage de services partagés qui répond aux exigences des équipes de science des données en matière d'opérations d'apprentissage automatique (MLops).
- Des comptes de développement, de test et de production de charges de travail ML partagés entre les équipes de data science.
- Des politiques de gouvernance garantissant que la charge de travail de chaque équipe de data science fonctionne de manière isolée.
- Bonnes pratiques courantes.



Structure de compte du modèle d'exploitation centralisé

Modèle de structure de compte décentralisé

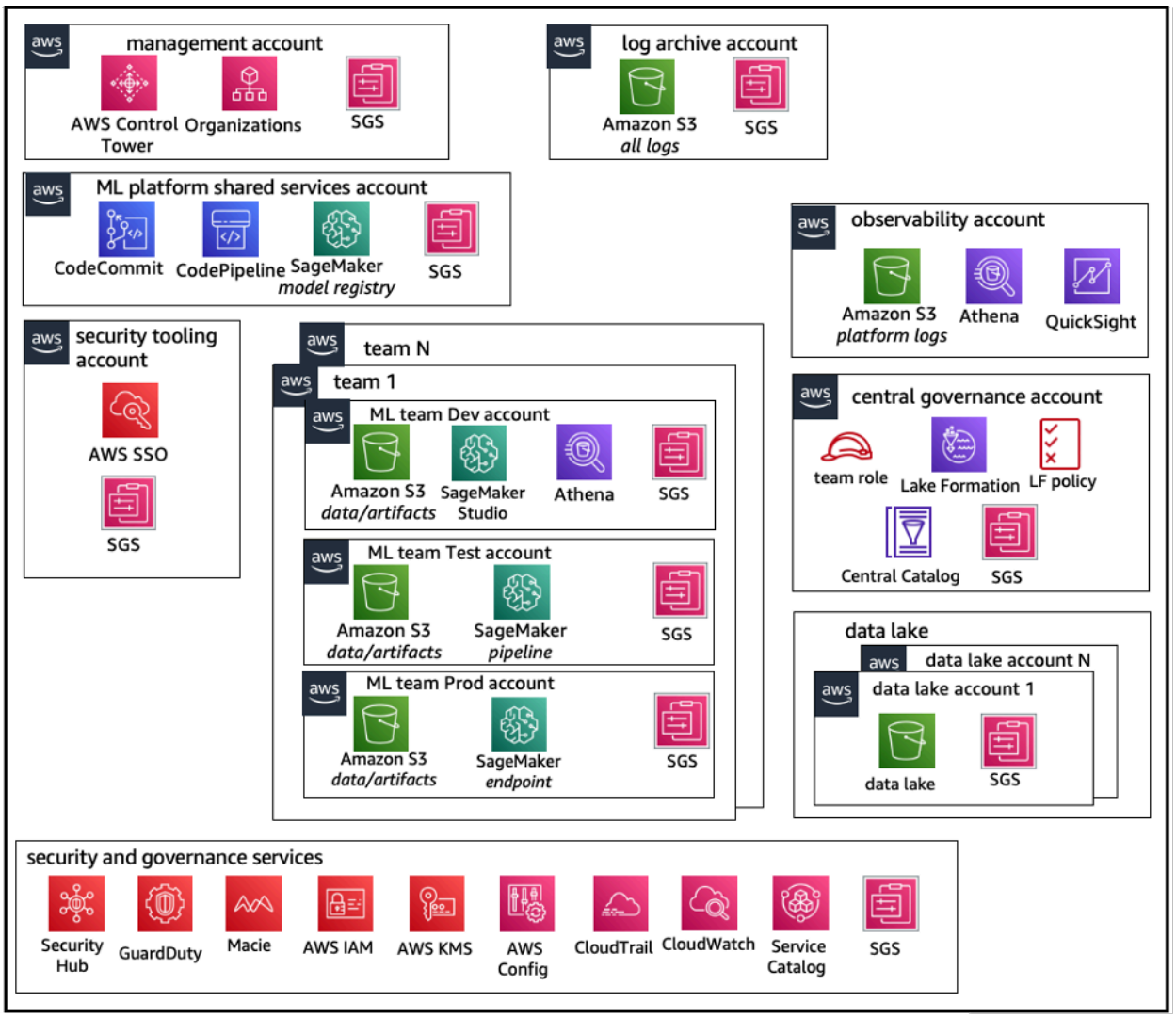
Dans ce modèle, chaque équipe de ML fonctionne de manière indépendante pour approvisionner, gérer et gouverner les comptes et les ressources de ML. Cependant, nous recommandons aux équipes de machine learning d'utiliser une approche centralisée d'observabilité et de gouvernance des données afin de simplifier la gouvernance des données et la gestion des audits.



Structure de compte du modèle opérationnel décentralisé

Modèle de structure de compte fédéré

Ce modèle est similaire au modèle centralisé ; toutefois, la principale différence est que chaque équipe de science des données/ML dispose de son propre ensemble de comptes de charge de travail de développement/test/production qui permettent une isolation physique robuste de ses ressources de ML et permettent également à chaque équipe d'évoluer indépendamment sans impact sur les autres équipes.



Structure de compte du modèle d'exploitation fédéré

Plateforme ML, mutualisation

La mutualisation est une architecture logicielle dans laquelle une seule instance logicielle peut desservir plusieurs groupes d'utilisateurs distincts. Un locataire est un groupe d'utilisateurs qui partagent un accès commun avec des privilèges spécifiques à l'instance logicielle. Par exemple, si vous créez plusieurs produits ML, chaque équipe produit ayant des exigences d'accès similaires peut être considérée comme un locataire ou une équipe.

Bien qu'il soit possible de mettre en œuvre plusieurs équipes au sein d'une instance SageMaker Studio (telle qu'un SageMakerdomaine), évaluez ces avantages par rapport à des compromis tels que le rayon de diffusion, l'attribution des coûts et les limites de niveau de compte lorsque vous

regroupez plusieurs équipes dans un seul domaine SageMaker Studio. Pour en savoir plus sur ces compromis et les meilleures pratiques, consultez les sections suivantes.

Si vous avez besoin d'une isolation absolue des ressources, envisagez d'implémenter des domaines SageMaker Studio pour chaque locataire dans un compte différent. En fonction de vos exigences en matière d'isolation, vous pouvez implémenter plusieurs secteurs d'activité (LOB) sous forme de domaines multiples au sein d'un même compte et d'une seule région. Utilisez des espaces partagés pour une collaboration en temps quasi réel entre les membres d'une même équipe/d'un même lob. Avec plusieurs domaines, vous continuerez à utiliser les politiques et autorisations de gestion des identités et des accès (IAM) pour garantir l'isolation des ressources.

SageMaker les ressources créées à partir d'un domaine sont automatiquement étiquetées avec le nom de ressource Amazon (ARN) du domaine et le profil utilisateur ou l'ARN de l'espace pour isoler facilement les ressources. Pour des exemples de politiques, reportez-vous à la documentation sur l'isolation des ressources du domaine. Vous pouvez y voir la référence détaillée indiquant quand utiliser une stratégie multi-comptes ou multidomaines, ainsi que les comparaisons de fonctionnalités dans la documentation, et vous pouvez consulter des exemples de scripts pour compléter les balises des domaines existants dans le référentiel. GitHub

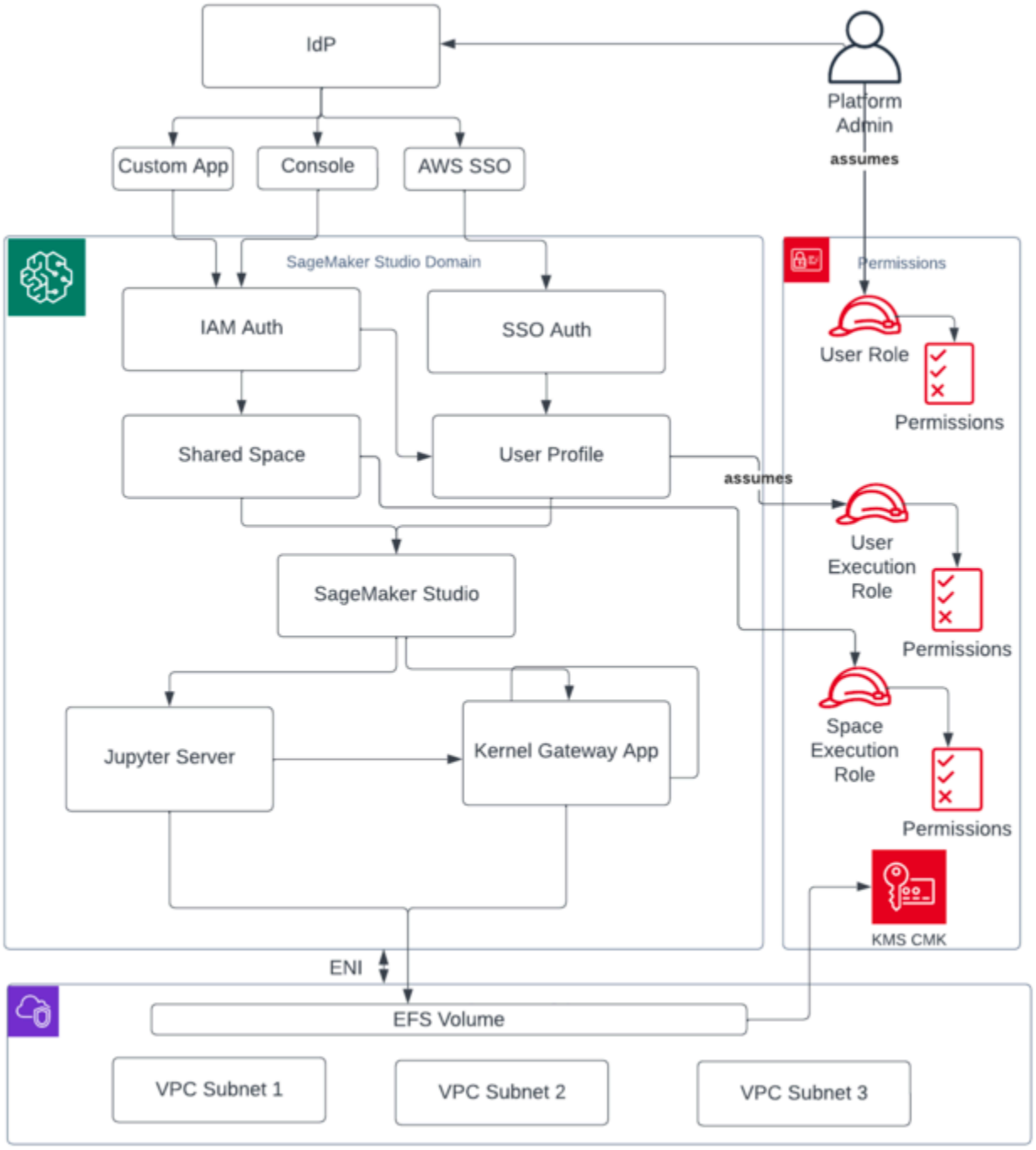
Enfin, vous pouvez implémenter un déploiement en libre-service des ressources SageMaker Studio sur plusieurs comptes à l'aide AWS Service Catalogde. Pour plus d'informations, reportez-vous à la section Gérer les AWS Service Catalog produits en plusieurs Comptes AWS et Régions AWS.

Gestion de domaine

Un SageMaker domaine Amazon comprend :

- Un volume Amazon Elastic File System (Amazon EFS) associé
- Liste des utilisateurs autorisés
- Une variété de configurations de sécurité, d'applications, de politiques et d'Amazon Virtual Private Cloud (Amazon VPC)

Le schéma suivant fournit une vue d'ensemble des différents composants qui constituent un SageMakerStudio domaine :

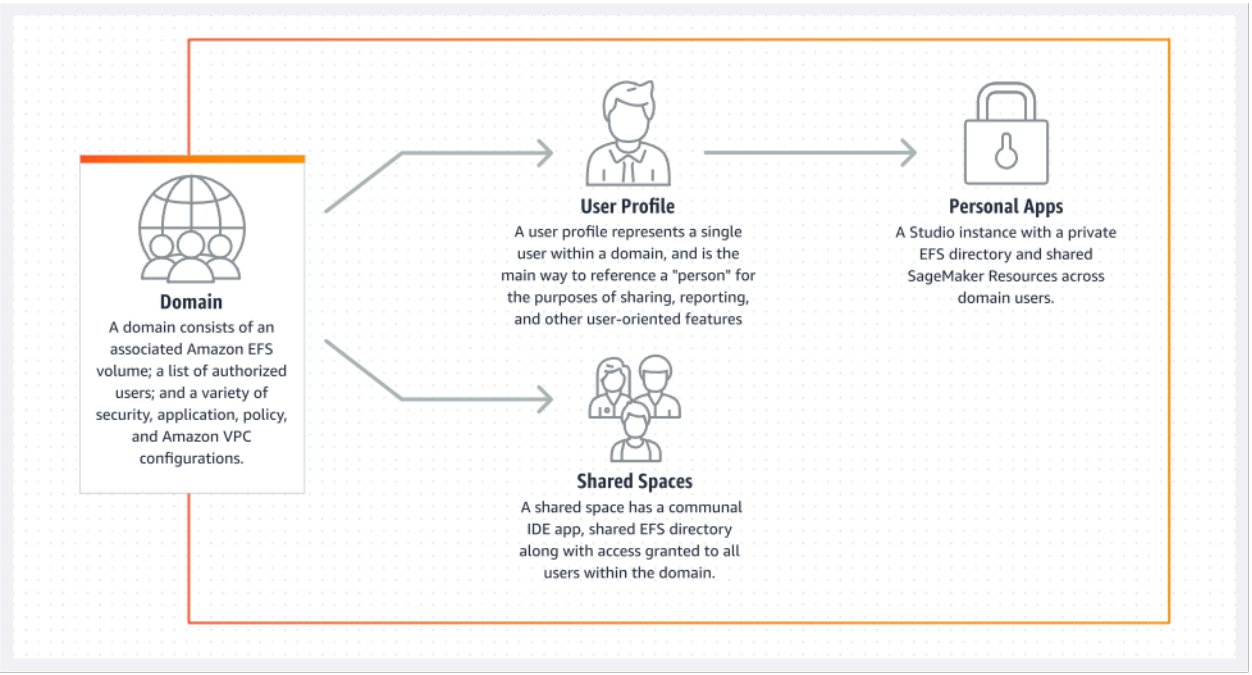


Vue de haut niveau des différents composants constituant un domaine SageMaker Studio

Domaines multiples et espaces partagés

Amazon prend SageMaker désormais en charge la création de plusieurs SageMaker domaines en un seul Région AWS pour chaque compte. Chaque domaine peut avoir ses propres paramètres de domaine, tels que le mode d'authentification, et ses propres paramètres réseau, tels que le VPC et les sous-réseaux. Un profil utilisateur ne peut pas être partagé entre les domaines. Si un utilisateur humain fait partie de plusieurs équipes séparées par des domaines, créez un profil utilisateur pour l'utilisateur dans chaque domaine. Reportez-vous à la présentation des domaines multiples pour en savoir plus sur le remblayage des balises pour les domaines existants.

Chaque domaine configuré en mode d'authentification IAM peut utiliser un espace partagé pour une collaboration en temps quasi réel entre les utilisateurs. Avec un espace partagé, les utilisateurs ont accès à un répertoire Amazon EFS partagé et à une JupyterServerapplication partagée pour l'interface utilisateur, et peuvent co-modifier en temps quasi réel. Le balisage automatique des ressources créées par les espaces partagés permet aux administrateurs de suivre les coûts au niveau du projet. L' JupyterServer interface utilisateur partagée filtre également les ressources telles que les expériences et les entrées de registre des modèles afin que seuls les éléments pertinents pour le projet de machine learning partagé soient affichés. Le schéma suivant fournit une vue d'ensemble des applications privées et des espaces partagés au sein de chaque domaine.



Vue d'ensemble des applications privées et des espaces partagés au sein d'un même domaine

Configurez des espaces partagés dans votre domaine

Les espaces partagés sont généralement créés pour une entreprise ou un projet de machine learning particulier dans le cadre duquel les membres d'un même domaine ont besoin d'un accès en temps quasi réel au même stockage de fichiers sous-jacent et au même IDE. L'utilisateur peut accéder à ses blocs-notes, les lire, les modifier et les partager en temps quasi réel, ce qui lui permet de commencer à itérer avec ses pairs le plus rapidement possible.

Pour créer un espace partagé, vous devez d'abord désigner un rôle d'exécution par défaut qui régira les autorisations de tout utilisateur utilisant l'espace. Au moment de la rédaction de cet article, tous les utilisateurs d'un domaine auront accès à tous les espaces partagés de leur domaine. Reportez-vous à la section Créer un espace partagé pour obtenir la dernière documentation sur l'ajout d'espaces partagés à un domaine existant.

Configuration de votre domaine pour la fédération IAM

Avant de configurer la fédération AWS Identity and Access Management (IAM) pour votre domaine SageMaker Studio, vous devez configurer un rôle d'utilisateur de fédération IAM (tel qu'un administrateur de plateforme) dans votre IdP, comme indiqué dans la section Gestion des identités.

Pour obtenir des instructions détaillées sur la configuration de SageMaker Studio avec l'option IAM, reportez-vous à la section Intégration au SageMaker domaine Amazon à l'aide d'IAM Identity Center.

Configurez votre domaine pour la fédération d'authentification unique (SSO)

Pour utiliser la fédération d'authentification unique (SSO), vous devez l'activer AWS IAM Identity Center dans votre compte de AWS Organizationsgestion dans la même région que celle dans laquelle vous devez exécuter SageMaker Studio. Les étapes de configuration du domaine sont similaires aux étapes de fédération IAM, sauf que vous sélectionnez AWS IAM Identity Center(iDC) dans la section Authentification.

Pour obtenir des instructions détaillées, reportez-vous à la section Intégration au SageMaker domaine Amazon à l'aide d'IAM Identity Center.

SageMaker Profil utilisateur du studio

Un profil utilisateur représente un utilisateur unique au sein d'un domaine et constitue le principal moyen de référencer une « personne » à des fins de partage, de reporting et d'autres fonctionnalités orientées vers l'utilisateur. Cette entité est créée lorsqu'un utilisateur intègre toSageMaker Studio. Si un administrateur invite une personne par e-mail ou l'importe depuis iDC, un profil utilisateur est automatiquement créé. Un profil utilisateur est le principal détenteur des paramètres d'un utilisateur individuel et contient une référence au répertoire personnel Amazon Elastic File System (Amazon EFS) privé de l'utilisateur. Nous recommandons de créer un profil utilisateur pour chaque utilisateur physique de l'application SageMaker Studio. Chaque utilisateur possède son propre répertoire dédié sur Amazon EFS, et les profils utilisateur ne peuvent pas être partagés entre les domaines d'un même compte.

Chaque profil utilisateur partageant le domaine SageMaker Studio reçoit des ressources de calcul dédiées (telles que des instances SageMaker Amazon Elastic Compute Cloud (Amazon EC2)) pour exécuter des blocs-notes. Les instances de calcul allouées à l'utilisateur 1 sont complètement isolées de celles allouées à l'utilisateur 2. De même, les ressources informatiques allouées aux utilisateurs d'un AWS compte sont complètement distinctes de celles allouées aux utilisateurs d'un autre compte. Chaque utilisateur peut exécuter jusqu'à quatre applications (applications) dans des conteneurs Docker isolés ou des images sur le même type d'instance.

Application Jupyter Server

Lorsque vous lancez un bloc-notes Amazon SageMaker Studio pour un utilisateur en accédant à l'URL pré-signée ou en vous connectant à l'aide d'AWS IAM iDC, l'application Jupyter Server est lancée dans l'instance VPC gérée par le service. SageMaker Chaque utilisateur dispose de sa propre application Jupyter Server dédiée dans une application privée. Par défaut, l'application Jupyter Server pour ordinateurs portables SageMaker Studio est exécutée sur une `ml.t3.medium` instance dédiée (réservée en tant que type d'instance système). Le calcul pour cette instance n'est pas facturé au client.

L'application Jupyter Kernel Gateway

L'application Kernel Gateway peut être créée via l'API ou l'interface SageMaker Studio, et elle s'exécute sur le type d'instance choisi. Cette application peut être exécutée à l'aide de l'une des images SageMaker Studio intégrées préconfigurées avec les logiciels de science des données les plus courants et de deep learning tels qu'TensorFlowApache MXnet et. PyTorch

Les utilisateurs peuvent démarrer et exécuter plusieurs noyaux de bloc-notes Jupyter, sessions de terminal et consoles interactives au sein de la même application SageMaker Studio Image/kernel Gateway. Les utilisateurs peuvent également exécuter jusqu'à quatre applications ou images Kernel Gateway sur la même instance physique, chacune étant isolée par son conteneur/image.

Pour créer des applications supplémentaires, vous devez utiliser un autre type d'instance. Un profil utilisateur ne peut avoir qu'une seule instance en cours d'exécution, quel que soit le type d'instance. Par exemple, un utilisateur peut exécuter à la fois un bloc-notes simple utilisant l'image de science des données intégrée à SageMaker Studio et un autre bloc-notes utilisant l' TensorFlow image intégrée, sur la même instance. Les utilisateurs sont facturés en fonction de la durée d'exécution de l'instance. Pour éviter des coûts lorsque l'utilisateur n'exécute pas activement SageMaker Studio, il doit arrêter l'instance. Pour plus d'informations, reportez-vous à la section Arrêter et mettre à jour les applications Studio.

Chaque fois que vous arrêtez et rouvrez une application Kernel Gateway depuis l'interface SageMaker Studio, cette application est démarrée sur une nouvelle instance. Cela signifie que l'installation du package n'est pas maintenue lors des redémarrages de la même application. De même, si un utilisateur change le type d'instance sur un bloc-notes, les packages installés et les variables de session sont perdus. Cependant, vous pouvez utiliser des fonctionnalités telles que l'ajout de votre propre image et des scripts de cycle de vie pour transférer les packages de l'utilisateur dans SageMaker Studio et les conserver via des changements d'instance et le lancement de nouvelles instances.

Volume Amazon Elastic File System

Lorsqu'un domaine est créé, un seul volume Amazon Elastic File System (Amazon EFS) est créé pour être utilisé par tous les utilisateurs du domaine. Chaque profil utilisateur reçoit un répertoire personnel privé dans le volume Amazon EFS pour stocker les blocs-notes, les GitHub référentiels et les fichiers de données de l'utilisateur. Chaque espace d'un domaine reçoit un répertoire privé dans le volume Amazon EFS auquel plusieurs profils d'utilisateurs peuvent accéder. L'accès aux dossiers est séparé par utilisateur, via les autorisations du système de fichiers. SageMaker Studio crée un ID utilisateur unique global pour chaque profil utilisateur ou espace, et l'applique en tant qu'identifiant d'utilisateur/de groupe POSIX (Portable Operating System Interface) pour le répertoire personnel de l'utilisateur sur EFS, ce qui empêche les autres utilisateurs/espaces d'accéder à ses données.

Sauvegarde et restauration

Un volume EFS existant ne peut pas être rattaché à un nouveau SageMaker domaine. Dans un environnement de production, assurez-vous que le volume Amazon EFS est sauvegardé (sur un autre volume EFS ou sur Amazon Simple Storage Service (Amazon S3)). Si un volume EFS est supprimé accidentellement, l'administrateur doit démonter et recréer le domaine SageMaker Studio. Procédez comme suit :

Sauvegardez la liste des profils utilisateur, des espaces et des identifiants utilisateur (UID) EFS associés via les appels `ListUserProfilesDescribeUserProfile`,`List Spaces`, et `DescribeSpace` API.

1. Créez un nouveau domaine SageMaker Studio.
2. Créez les profils utilisateur et les espaces.
3. Pour chaque profil utilisateur, copiez les fichiers de la sauvegarde sur EFS/Amazon S3.
4. Supprimez éventuellement toutes les applications et tous les profils utilisateur de l'ancien domaine SageMaker Studio.

Pour obtenir des instructions détaillées, reportez-vous à la section annexe relative à la sauvegarde et à la restauration du domaine SageMaker Studio.

> **Note**
>
> Cela peut également être réalisé en `LifecycleConfigurations` sauvegardant les données depuis et vers S3 chaque fois qu'un utilisateur démarre son application.

Volume Amazon EBS

Un volume de stockage Amazon Elastic Block Store (Amazon EBS) est également attaché à chaque instance de SageMaker Studio Notebook. Il est utilisé comme volume racine du conteneur ou de l'image exécutée sur l'instance. Tant que le stockage Amazon EFS est persistant, le volume Amazon EBS attaché au conteneur est temporaire. Les données stockées localement sur le volume Amazon EBS ne seront pas conservées si le client supprime l'application.

Sécurisation de l'accès à l'URL pré-signée

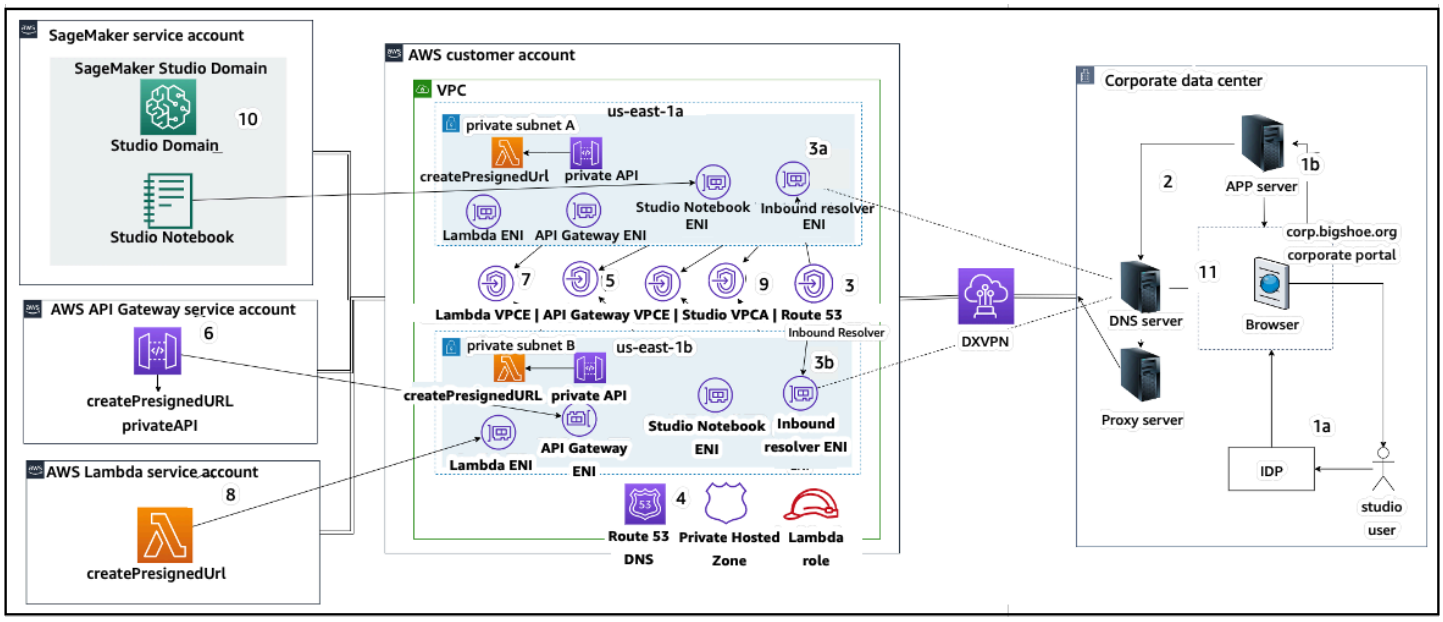
Lorsqu'un utilisateur de SageMaker Studio ouvre le lien du bloc-notes, SageMaker Studio valide la politique IAM de l'utilisateur fédéré pour autoriser l'accès, puis génère et résout l'URL pré-signée pour l'utilisateur. Comme la SageMaker console s'exécute sur un domaine Internet, cette URL générée et pré-signée est visible dans la session du navigateur. Cela constitue un vecteur de menace indésirable pour le vol de données et l'accès aux données des clients lorsque les contrôles d'accès appropriés ne sont pas appliqués.

Studio prend en charge plusieurs méthodes pour renforcer les contrôles d'accès contre le vol de données d'URL pré-signées :

- Validation de l'adresse IP du client à l'aide de la condition de politique IAM `aws:sourceIp`
- Validation du VPC client à l'aide de la condition IAM `aws:sourceVpc`
- Validation du point de terminaison VPC du client à l'aide de la condition de politique IAM `aws:sourceVpce`

Lorsque vous accédez aux blocs-notes SageMaker Studio depuis la SageMaker console, la seule option disponible consiste à utiliser la validation de l'adresse IP du client avec la condition de politique IAM. `aws:sourceIp` Cependant, vous pouvez utiliser des produits de routage du trafic par navigateur tels que Zscaler pour garantir l'évolutivité et la conformité de l'accès Internet de votre personnel. Ces produits de routage du trafic génèrent leur propre adresse IP source, dont la plage d'adresses IP n'est pas contrôlée par l'entreprise cliente. Il est donc impossible pour ces entreprises clientes d'utiliser `aws:sourceIp` cette condition.

Pour utiliser la validation du point de terminaison VPC client à l'aide de la condition de politique IAM`aws:sourceVpce`, la création d'une URL pré-signée doit provenir du même VPC client où SageMaker Studio est déployé, et la résolution de l'URL pré-signée doit se faire via un point de terminaison Studio VPC sur le SageMaker VPC du client. Cette résolution de l'URL pré-signée pendant le temps d'accès pour les utilisateurs du réseau d'entreprise peut être réalisée à l'aide de règles de transfert DNS (à la fois dans Zscaler et dans le DNS d'entreprise), puis vers le point de terminaison VPC du client à l'aide d'un résolveur entrant Amazon Route 53, comme illustré dans l'architecture suivante :



Accès à l'URL pré-signée de Studio avec un point de terminaison VPC via le réseau d'entreprise

Pour step-by-step obtenir des conseils sur la configuration de l'architecture précédente, reportez-vous à la section relative aux URL présignées d'Amazon SageMaker Studio, partie 1 : Infrastructure de base.

SageMaker quotas et limites de domaines

- SageMaker La fédération SSO du domaine Studio est prise en charge uniquement dans la région, sur tous les comptes membres de l'AWSorganisation dans laquelle AWS Identity Center est approvisionné.
- Les espaces partagés ne sont actuellement pas pris en charge avec les domaines configurés avec AWS Identity Center.
- La configuration du VPC et du sous-réseau ne peut pas être modifiée après la création du domaine. Vous pouvez toutefois créer un nouveau domaine avec une configuration de VPC et de sous-réseau différente.
- L'accès au domaine ne peut pas être commuté entre les modes IAM et SSO après la création du domaine. Vous pouvez créer un nouveau domaine avec un mode d'authentification différent.
- Il existe une limite de quatre applications de passerelle de noyau par type d'instance lancées pour chaque utilisateur.
- Chaque utilisateur ne peut lancer qu'une seule instance de chaque type d'instance.

- Les ressources consommées au sein d'un domaine sont limitées, telles que le nombre d'instances lancées par type d'instance et le nombre de profils utilisateur pouvant être créés. Reportez-vous à la page des quotas de service pour obtenir la liste complète des limites de service.
- Les clients peuvent soumettre un dossier de support d'entreprise avec une justification commerciale pour augmenter les limites de ressources par défaut, telles que le nombre de domaines ou les profils d'utilisateurs, sous réserve de garanties au niveau du compte.
- La limite stricte du nombre d'applications simultanées par compte est de 2 500 applications. Les limites des domaines et des profils utilisateurs dépendent de cette limite stricte. Par exemple, un compte peut avoir un seul domaine avec 1 000 profils d'utilisateurs, ou 20 domaines avec 50 profils d'utilisateur chacun.

Gestion des identités

Cette section explique comment les utilisateurs du personnel d'un annuaire d'entreprise se fédérent dans SageMaker Studio Comptes AWS et y accèdent. Tout d'abord, nous allons décrire brièvement comment les utilisateurs, les groupes et les rôles sont mappés, ainsi que le fonctionnement de la fédération d'utilisateurs.

Utilisateurs, groupes et rôles

Dans AWS, les autorisations relatives aux ressources sont gérées à l'aide d'utilisateurs, de groupes et de rôles. Les clients peuvent gérer leurs utilisateurs et leurs groupes soit via IAM, soit dans un annuaire d'entreprise tel qu'Active Directory (AD), activé via un IdP externe tel qu'Okta, qui leur permet d'authentifier les utilisateurs auprès de diverses applications exécutées dans le cloud et sur site.

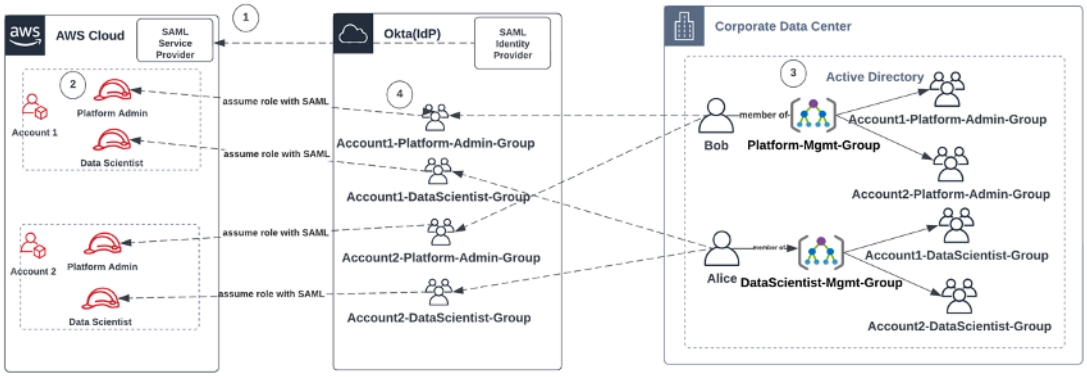
Comme indiqué dans la section Gestion des identités du pilier de AWS sécurité, il est recommandé de gérer les identités de vos utilisateurs dans un IdP central, car cela permet de s'intégrer facilement à vos processus RH principaux et de gérer l'accès aux utilisateurs de votre personnel.

IdPs tels qu'Okta, permettent aux utilisateurs finaux de s'authentifier auprès d'un ou de plusieurs rôles Comptes AWS et d'accéder à des rôles spécifiques à l'aide du SSO avec le langage SAML (Security Assertation Markup Language). Les administrateurs d'IdP ont la possibilité de télécharger des rôles depuis l' Comptes AWS IdP et de les attribuer aux utilisateurs. Lorsqu'ils se connectent à AWS, les utilisateurs finaux voient apparaître un AWS écran qui affiche une liste des AWS rôles qui leur ont été assignés dans un ou plusieurs rôles Comptes AWS. Ils peuvent sélectionner le rôle à assumer pour la connexion, qui définit leurs autorisations pour la durée de cette session authentifiée.

Un groupe doit exister dans IdP pour chaque combinaison de comptes et de rôles spécifique à laquelle vous souhaitez donner accès. Vous pouvez considérer ces groupes comme des groupes AWS spécifiques à un rôle. Tout utilisateur membre de ces groupes spécifiques à un rôle se voit octroyer un droit unique : l'accès à un rôle spécifique dans un rôle spécifique Compte AWS. Toutefois, ce processus d'autorisation unique ne permet pas de gérer l'accès des utilisateurs en attribuant à chaque utilisateur des groupes de AWS rôles spécifiques. Pour simplifier l'administration, nous vous recommandons également de créer un certain nombre de groupes pour tous les groupes d'utilisateurs distincts de votre organisation qui nécessitent différents ensembles de AWS droits.

Pour illustrer la configuration centrale de l'IdP, imaginons une entreprise dotée d'une configuration AD, dans laquelle les utilisateurs et les groupes sont synchronisés avec le répertoire IdP. Dans

AWS, ces groupes AD sont mappés aux rôles IAM. Les principales étapes du flux de travail sont les suivantes :



Flux de travail pour l'intégration des utilisateurs AD, des groupes AD et des rôles IAM

1. Dans AWS, configurez l'intégration SAML pour chacun d'entre vous Comptes AWS avec votre IdP.
2. Dans AWS, configurez des rôles dans chacun d'eux Compte AWS et synchronisez-les avec IdP.
3. Dans le système AD d'entreprise :
   a. Créez un groupe AD pour chaque rôle de compte et synchronisez-le avec IdP (par exemple, `Account1-Platform-Admin-Group` (alias AWS Role Group)).
   b. Créez un groupe de gestion à chaque niveau de personnalité (par exemple,`Platform-Mgmt-Group`) et assignez des groupes de AWS rôles en tant que membres.
   c. Affectez des utilisateurs à ce groupe de gestion pour autoriser l'accès aux Compte AWS rôles.
4. Dans IdP, associez des groupes de AWS rôles (tels que`Account1-Platform-Admin-Group`) à Compte AWS des rôles (tels que Platform Admin dans Account1).
5. Lorsque la data scientist Alice se connecte à Idp, une interface utilisateur de l'application AWS Federation lui est présentée, avec deux options parmi lesquelles choisir : « Account 1 Data Scientist » et « Account 2 Data Scientist ».
6. Alice choisit l'option « Data Scientist du compte 1 », et ils sont connectés à leur application autorisée dans le AWS compte 1 (SageMaker console).

Pour obtenir des instructions détaillées sur la configuration de la fédération de comptes SAML, reportez-vous à la section Comment configurer SAML 2.0 pour AWS la fédération de comptes d'Okta.

Fédération d'utilisateurs

L'authentification pour SageMaker Studio peut être effectuée à l'aide d'IAM ou d'IAM iDC. Si les utilisateurs sont gérés via IAM, ils peuvent choisir le mode IAM. Si l'entreprise utilise un IdP externe, elle peut se fédérer via IAM ou IAM iDC. Notez que le mode d'authentification ne peut pas être mis à jour pour un domaine SageMaker Studio existant. Il est donc essentiel de prendre la décision avant de créer un domaine SageMaker Studio de production.

Si SageMaker Studio est configuré en mode IAM, les utilisateurs de SageMaker Studio accèdent à l'application via une URL pré-signée qui connecte automatiquement un utilisateur à l'application SageMaker Studio lorsqu'il y accède via un navigateur.

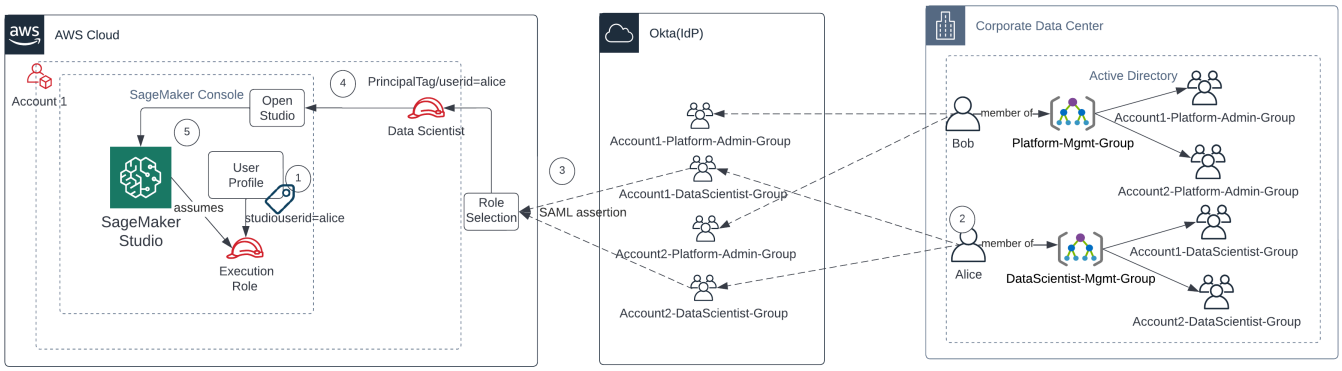
Utilisateurs IAM

Pour les utilisateurs IAM, l'administrateur crée des profils utilisateur SageMaker Studio pour chaque utilisateur et associe le profil utilisateur à un rôle IAM qui autorise les actions nécessaires que l'utilisateur doit effectuer depuis Studio. Pour empêcher un AWS utilisateur d'accéder uniquement à son profil utilisateur SageMaker Studio, l'administrateur doit étiqueter le profil utilisateur SageMaker Studio et associer à l'utilisateur une politique IAM qui lui permet d'accéder uniquement si la valeur de la balise est identique au nom AWS d'utilisateur. La déclaration de politique se présente comme suit :

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

AWS IAM ou fédération de comptes

La méthode Compte AWS de fédération permet aux clients de se fédérer dans la SageMaker console à partir de leur IdP SAML, tel qu'Okta. Pour empêcher les utilisateurs d'accéder uniquement à leur profil utilisateur, l'administrateur doit baliser le profil utilisateur de SageMaker Studio, ajouter `PrincipalTags` l'IdP et le définir comme balises transitives. Le schéma suivant montre comment l'utilisateur fédéré (Data Scientist Alice) est autorisé à accéder à son propre profil utilisateur SageMaker Studio.



Accès à SageMaker Studio en mode de fédération IAM

1. Le profil utilisateur d'Alice SageMaker Studio est balisé avec son ID utilisateur et associé au rôle d'exécution.
2. Alice s'authentifie auprès de l'IdP (Okta).
3. IdP authentifie Alice et publie une assertion SAML avec les deux rôles (Data Scientist pour les comptes 1 et 2) dont Alice est membre. Alice sélectionne le rôle de data scientist pour le compte 1.
4. Alice est connectée à la SageMaker console Account 1, assumant le rôle de Data Scientist. Alice ouvre son instance d'application Studio à partir de la liste des instances d'application Studio.
5. La balise principale Alice dans la session de rôle assumé est validée par rapport à la balise de profil utilisateur de l'instance d'application SageMaker Studio sélectionnée. Si la balise de profil est valide, l'instance de l'application SageMaker Studio est lancée en assumant le rôle d'exécution.

Si vous souhaitez automatiser la création de rôles et de politiques SageMaker d'exécution dans le cadre de l'intégration des utilisateurs, voici un moyen d'y parvenir :

1. Configurez un groupe AD, par exemple `SageMaker-Account1-Group` au niveau de chaque compte et du domaine Studio.

2. Ajoutez SageMaker -Account1-Group à l'adhésion au groupe de l'utilisateur lorsque vous devez intégrer un utilisateur à Studio. SageMaker

Configurez un processus d'automatisation qui écoute l'événement `SageMaker-Account1-Group` d'adhésion et utilisez des AWS API pour créer le rôle, les politiques, les balises et le profil utilisateur de SageMaker Studio en fonction de leur appartenance au groupe AD. Associez le rôle au profil utilisateur. Pour un exemple de politique, reportez-vous àEmpêcher les utilisateurs de SageMaker Studio d'accéder à d'autres profils utilisateur.

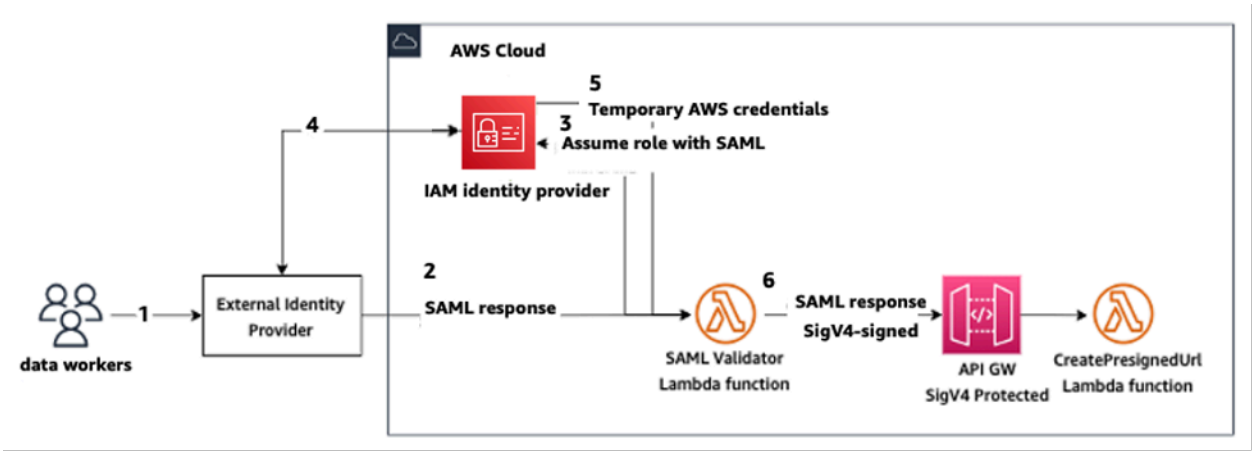
Authentification SAML à l'aide de AWS Lambda

En mode IAM, les utilisateurs peuvent également être authentifiés dans SageMaker Studio à l'aide d'assertions SAML. Dans cette architecture, le client dispose d'un IdP existant, dans lequel il peut créer une application SAML permettant aux utilisateurs d'accéder à Studio (au lieu de l'application AWS Identity Federation). L'IdP du client est ajouté à IAM. Une AWS Lambda fonction permet de valider l'assertion SAML à l'aide d'IAM et de STS, puis invoque directement une passerelle API ou une fonction Lambda pour créer l'URL de domaine pré-signée.

L'avantage de cette solution est que la fonction Lambda peut personnaliser la logique d'accès à SageMaker Studio. Par exemple :

- Créez automatiquement un profil utilisateur s'il n'en existe pas.
- Attachez ou supprimez des rôles ou des documents de politique au rôle d'exécution de SageMaker Studio en analysant les attributs SAML.
- Personnalisez le profil utilisateur en ajoutant la configuration du cycle de vie (LCC) et en ajoutant des balises.

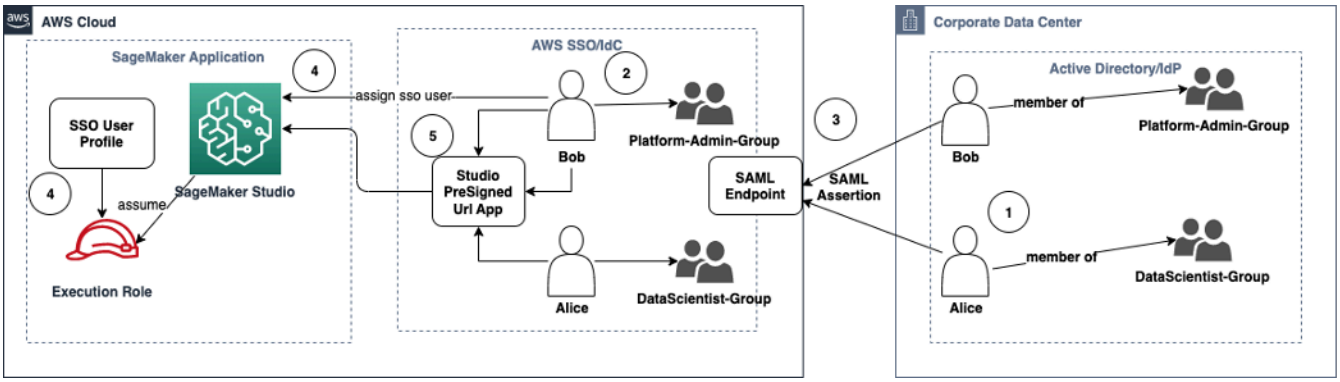
En résumé, cette solution exposera SageMaker Studio en tant qu'application SAML2.0 avec une logique personnalisée pour l'authentification et l'autorisation. Reportez-vous à la section de l'annexe Accès au SageMaker studio à l'aide de l'assertion SAML pour les détails de mise en œuvre.



Accès à SageMaker Studio à l'aide d'une application SAML personnalisée

Fédération AWS IAM iDC

La méthode de fédération iDC permet aux clients de se fédérer directement dans l'application SageMaker Studio à partir de leur IdP SAML (tel qu'Okta). Le schéma suivant montre comment l'utilisateur fédéré est autorisé à accéder à sa propre instance de SageMaker Studio.



Accès à SageMaker Studio en mode IAM iDC

1. Dans l'AD d'entreprise, l'utilisateur est membre de groupes AD tels que le groupe Platform Admin et le groupe Data Scientist.
2. L'utilisateur AD et les groupes AD du fournisseur d'identité (IdP) sont synchronisés avec AWS IAM Identity Center et disponibles en tant qu'utilisateurs et groupes d'authentification unique pour les attributions, respectivement.
3. L'IdP publie une assertion SAML sur le point de terminaison SAML AWS iDC.
4. Dans le SageMaker Studio, l'utilisateur iDC est affecté à l'application SageMaker Studio. Cette attribution peut être effectuée à l'aide d'iDC Group et SageMaker Studio s'appliquera à chaque

niveau d'utilisateur iDC. Lorsque cette attribution est créée, SageMaker Studio crée un profil utilisateur iDC et attache le rôle d'exécution du domaine.

5. L'utilisateur accède à l'application SageMaker Studio à l'aide de l'URL présignée sécurisée hébergée sous forme d'application cloud à partir de l'iDC. SageMaker Studio assume le rôle d'exécution associé à son profil utilisateur iDC.

Conseils d'authentification de domaine

Voici quelques points à prendre en compte lors du choix du mode d'authentification d'un domaine :

1. Si vous souhaitez que vos utilisateurs n'accèdent pas à l'interface utilisateur de SageMaker Studio AWS Management Console et ne la consultent pas directement, utilisez le mode d'authentification unique avec AWS IAM iDC.
2. Si vous souhaitez que vos utilisateurs n'accèdent pas à l'interface utilisateur de SageMaker Studio AWS Management Console et ne la consultent pas directement en mode IAM, vous pouvez le faire en utilisant une fonction Lambda dans le backend pour générer une URL présignée pour le profil utilisateur et en les redirigeant vers l'interface utilisateur de Studio. SageMaker
3. En mode iDC, chaque utilisateur est mappé à un profil utilisateur unique.
4. Le rôle d'exécution par défaut est automatiquement attribué à tous les profils utilisateur en mode iDC. Si vous souhaitez que différents rôles d'exécution soient attribués à vos utilisateurs, vous devez mettre à jour les profils utilisateurs à l'aide de l'UpdateUserProfileAPI.
5. Si vous souhaitez restreindre l'accès à l'interface utilisateur de SageMaker Studio en mode IAM (à l'aide de l'URL présignée générée) à un point de terminaison VPC, sans passer par Internet, vous pouvez utiliser un résolveur DNS personnalisé. Reportez-vous au billet de blog SageMaker consacré aux URL présignées d'Amazon Studio, partie 1 : Infrastructure fondamentale.

Gestion des autorisations

Cette section décrit les meilleures pratiques pour configurer les rôles, les politiques et les garde-fous IAM couramment utilisés pour le provisionnement et l'exploitation du domaine Studio. SageMaker

Rôles et politiques IAM

La meilleure pratique consiste à identifier d'abord les personnes et les applications pertinentes, connues sous le nom de responsables impliqués dans le cycle de vie du machine learning, et les AWS autorisations que vous devez leur accorder. Comme SageMaker c'est le cas pour un service géré, vous devez également prendre en compte les principes de service, qui sont AWS des services qui peuvent effectuer des appels d'API au nom d'un utilisateur. Le schéma suivant illustre les différents rôles IAM que vous souhaiterez peut-être créer, correspondant aux différents personnages de l'organisation.

SageMaker Rôles IAM

Ces rôles sont décrits en détail, ainsi que quelques exemples de permissions IAMspécifiques dont ils auront besoin.

- Rôle utilisateur d'administrateur ML : il s'agit d'un directeur qui fournit l'environnement aux scientifiques des données en créant des domaines de studio et des profils utilisateur (`sagemaker:CreateDomain`,`sagemaker:CreateUserProfile`), en créant AWS Key Management Service des clés pour les utilisateurs, en créant des compartiments S3 pour les scientifiques des données et en créant des référentiels Amazon ECR pour héberger des conteneurs. AWS KMS Ils peuvent également définir des configurations par défaut et des scripts de cycle de vie pour les utilisateurs, créer et joindre des images personnalisées au domaine SageMaker Studio, et fournir des produits Service Catalog tels que des projets personnalisés et des modèles Amazon EMR.

  Comme ce directeur n'exécutera pas de tâches de formation, par exemple, il n'a pas besoin d'autorisations pour lancer des tâches de SageMaker formation ou de traitement. S'ils utilisent

l'infrastructure comme modèles de code, tels que CloudFormation Terraform, pour approvisionner des domaines et des utilisateurs, ce rôle sera assumé par le service de provisionnement pour créer les ressources au nom de l'administrateur. Ce rôle peut disposer d'un accès en lecture seule pour SageMaker utiliser le. AWS Management Console

Ce rôle d'utilisateur aura également besoin de certaines autorisations EC2 pour lancer le domaine dans un VPC privé, d'autorisations KMS pour chiffrer le volume EFS, ainsi que d'autorisations pour créer un rôle lié à un service pour `iam:CreateServiceLinkedRole` Studio (). Nous décrirons ces autorisations détaillées plus loin dans le document.

- Rôle d'utilisateur du data scientist : ce principe est celui de l'utilisateur qui se connecte à SageMaker Studio, explore les données, crée des tâches et des pipelines de traitement et de formation, etc. L'autorisation principale dont l'utilisateur a besoin est l'autorisation de lancer SageMaker Studio, et le reste des politiques peut être géré par le rôle de service SageMaker d'exécution.
- SageMaker rôle de service d'exécution : étant donné qu'il SageMaker s'agit d'un service géré, il lance des tâches pour le compte d'un utilisateur. Ce rôle est souvent le plus large en termes d'autorisations autorisées, car de nombreux clients choisissent d'utiliser un seul rôle d'exécution pour exécuter des tâches de formation, des tâches de traitement ou des tâches d'hébergement de modèles. Bien qu'il s'agisse d'un moyen simple de démarrer, les clients évoluant au fil de leur parcours, ils divisent souvent le rôle d'exécution du bloc-notes en rôles distincts pour différentes actions d'API, en particulier lorsqu'ils exécutent ces tâches dans des environnements déployés.

  Vous associez un rôle au domaine SageMaker Studio lors de sa création. Toutefois, comme les clients peuvent avoir besoin de la flexibilité d'avoir différents rôles associés aux différents profils d'utilisateur du domaine (par exemple, en fonction de leur fonction), vous pouvez également associer un rôle IAM distinct à chaque profil utilisateur. Nous vous recommandons de mapper un seul utilisateur physique à un profil utilisateur unique. Si vous n'associez aucun rôle à un profil utilisateur lors de sa création, le comportement par défaut consiste également à associer le rôle d'exécution du SageMakerStudio domaine au profil utilisateur.

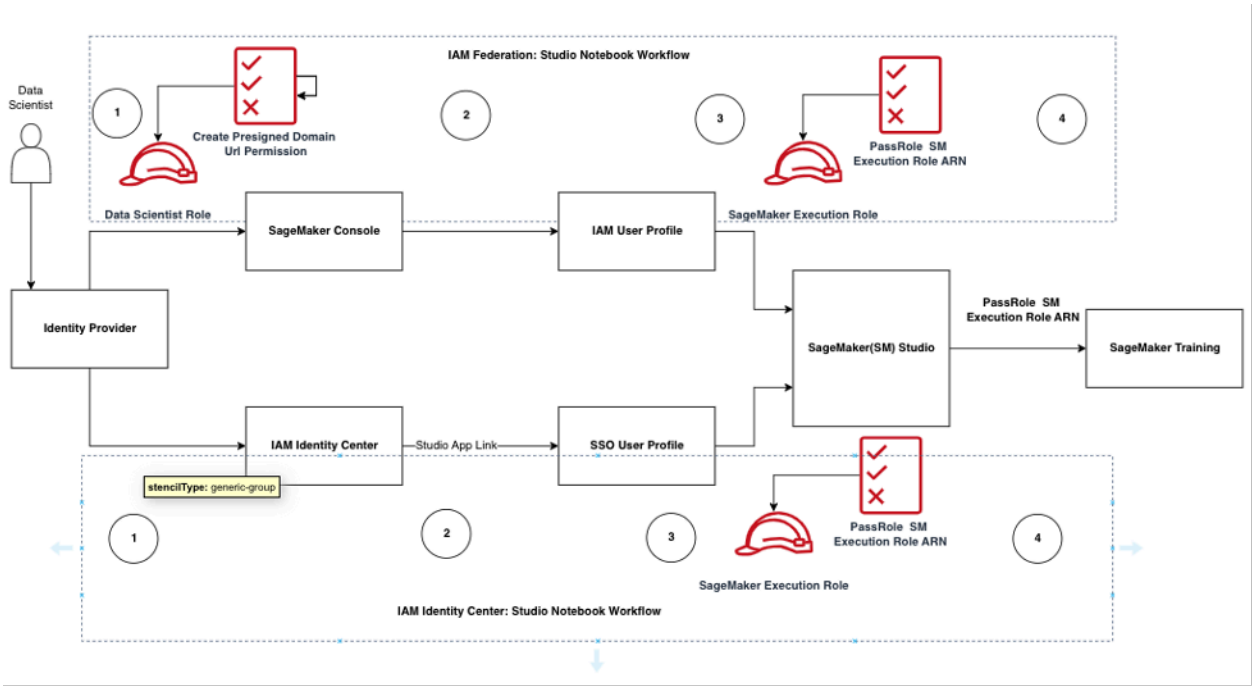
  Dans les cas où plusieurs data scientists et ingénieurs ML travaillent ensemble sur un projet et ont besoin d'un modèle d'autorisation partagé pour accéder aux ressources, nous vous recommandons de créer un rôle d'exécution de SageMaker service au niveau de l'équipe afin de partager les autorisations IAM entre les membres de votre équipe. Dans les cas où vous devez verrouiller les autorisations à chaque niveau d'utilisateur, vous pouvez créer un rôle d'exécution de SageMaker service individuel au niveau de l'utilisateur ; vous devez toutefois tenir compte de vos limites de service.

SageMaker Flux de travail d'autorisation de Studio Notebook

Cette section explique comment fonctionne l'autorisation du SageMaker Studio Notebook pour les différentes activités que le Data Scientist doit effectuer pour créer et entraîner le modèle directement à partir du SageMaker Studio Notebook. Le SageMaker domaine prend en charge deux modes d'autorisation :

- Fédération IAM
- IAM Identity Center

Ensuite, ce paper explique le flux de travail d'autorisation du Data Scientist pour chacun de ces modes.



Flux de travail d'authentification et d'autorisation pour les utilisateurs de Studio

Fédération IAM : flux de travail de SageMaker Studio Notebook

1. Un Data Scientist s'authentifie auprès de son fournisseur d'identité d'entreprise et assume le rôle d'utilisateur Data Scientist (le rôle de fédération d'utilisateurs) dans la SageMaker console. Ce rôle de fédération dispose `iam:PassRole` d'une autorisation d'API sur le rôle d' SageMaker exécution pour transmettre le rôle Amazon Resource Name (ARN) à SageMaker Studio.

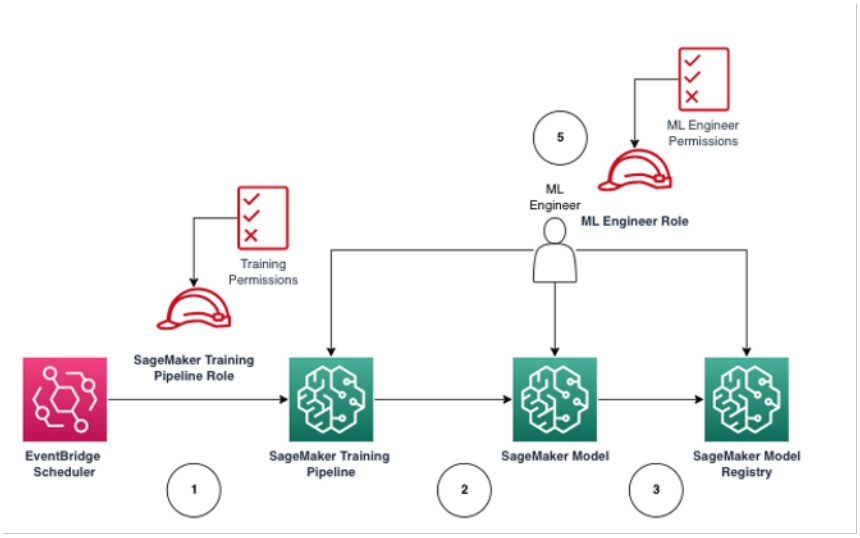
2. Le Data Scientist sélectionne le lien Open Studio dans son profil utilisateur Studio IAM associé au rôle SageMaker d'exécution.
3. Le service SageMaker Studio IDE est lancé, en supposant les autorisations de rôle SageMaker d'exécution du profil utilisateur. Ce rôle dispose `iam:PassRole` d'une autorisation API sur le rôle SageMaker d'exécution pour transmettre l'ARN du rôle au service de SageMaker formation.
4. Lorsque Data Scientist lance la tâche de formation dans le ou les nœuds de calcul distants, le rôle SageMaker d'exécution ARN est transmis au service de SageMaker formation. Cela crée une nouvelle session de rôle avec cet ARN et exécute la tâche de formation. Si vous devez définir davantage l'autorisation pour une tâche de formation, vous pouvez créer un rôle spécifique à la formation et transmettre l'ARN de ce rôle lorsque vous appelez l'API de formation.

IAM Identity Center : flux de travail de SageMaker Studio Notebook

1. Le data scientist s'authentifie auprès de son fournisseur d'identité d'entreprise et clique sur AWS IAM Identity Center. Le Data Scientist reçoit le portail Identity Center pour l'utilisateur.
2. Le Data Scientist clique sur le lien de l'application SageMaker Studio créé à partir de son profil utilisateur iDC, qui est associé au rôle SageMaker d'exécution.
3. Le service SageMaker Studio IDE est lancé, en supposant les autorisations de rôle SageMaker d'exécution du profil utilisateur. Ce rôle dispose `iam:PassRole` d'une autorisation API sur le rôle SageMaker d'exécution pour transmettre l'ARN du rôle au service de SageMaker formation.
4. Lorsque le Data Scientist lance la tâche de formation dans un ou plusieurs nœuds de calcul distants, le rôle SageMaker d'exécution ARN est transmis au service de SageMaker formation. L'ARN du rôle d'exécution crée une nouvelle session de rôle avec cet ARN et exécute la tâche de formation. Si vous devez limiter davantage l'autorisation pour les tâches de formation, vous pouvez créer un rôle spécifique à la formation et transmettre l'ARN de ce rôle lorsque vous appelez l'API de formation.

Environnement déployé : flux SageMaker de formation

Dans les environnements déployés tels que les tests et la production de systèmes, les tâches sont exécutées via un planificateur automatique et des déclencheurs d'événements, et l'accès humain à ces environnements est restreint depuis les ordinateurs portables SageMaker Studio. Cette section explique comment les rôles IAM fonctionnent avec le pipeline de SageMaker formation dans l'environnement déployé.



SageMaker flux de formation dans un environnement de production géré

1. Amazon EventBridge Scheduler déclenche la tâche du pipeline de SageMaker formation.
2. La SageMaker tâche du pipeline de SageMaker formation assume le rôle du pipeline de formation pour entraîner le modèle.
3. Le SageMaker modèle entraîné est enregistré dans le registre des SageMaker modèles.
4. Un ingénieur ML assume le rôle d'utilisateur de l'ingénieur ML pour gérer le pipeline et le SageMaker modèle de formation.

Autorisations relatives aux données

La possibilité pour les utilisateurs de SageMaker Studio d'accéder à n'importe quelle source de données est régie par les autorisations associées à leur rôle d'exécution SageMaker IAM. Les politiques associées peuvent les autoriser à lire, écrire ou supprimer des données de certains compartiments ou préfixes Amazon S3, et à se connecter aux bases de données Amazon RDS.

Accès aux données AWS Lake Formation

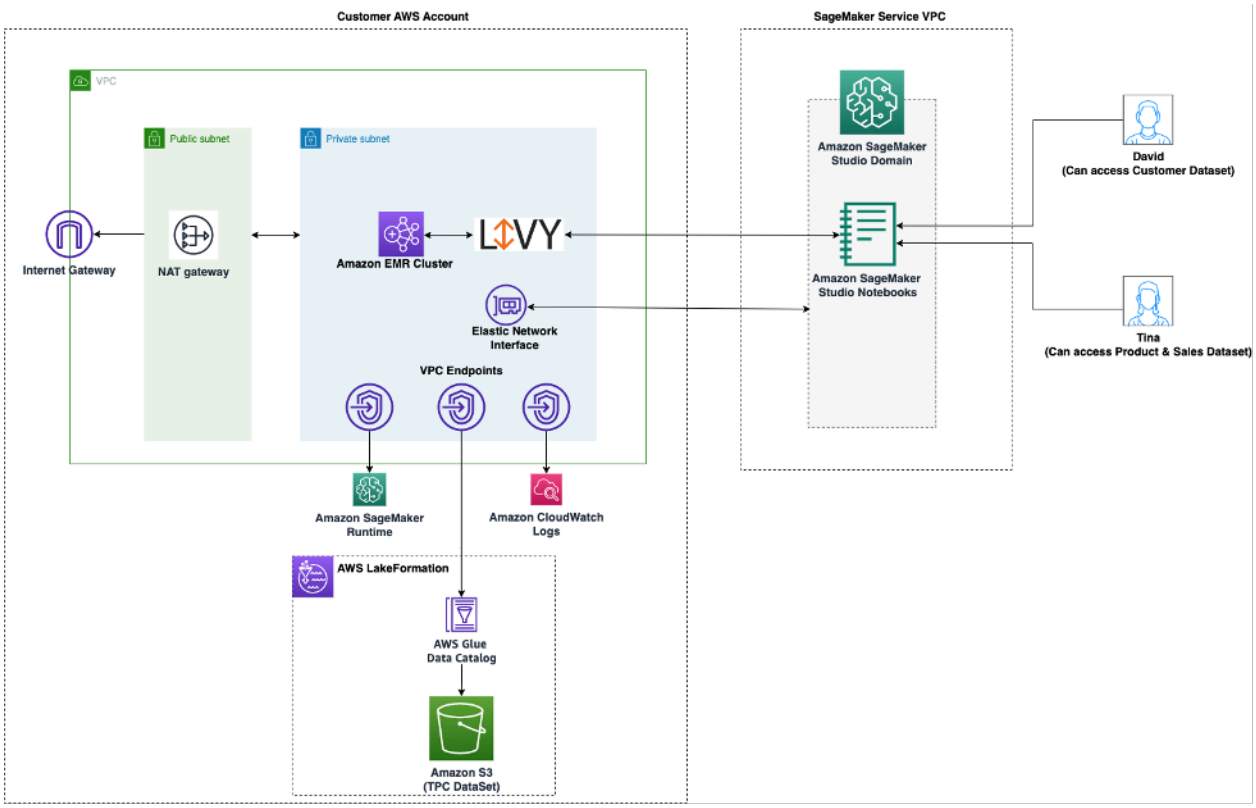
De nombreuses entreprises ont commencé à utiliser des lacs de données régis AWS Lake Formationpour permettre à leurs utilisateurs d'accéder aux données de manière précise. À titre d'exemple de telles données gouvernées, les administrateurs peuvent masquer des colonnes sensibles pour certains utilisateurs tout en autorisant les requêtes de la même table sous-jacente.

Pour utiliser Lake Formation from SageMaker Studio, les administrateurs peuvent enregistrer les rôles d'exécution SageMaker IAM en tant que`DataLakePrincipals`. Pour plus d'informations, reportez-vous à la section Lake Formation Permissions Reference. Une fois autorisé, il existe trois méthodes principales pour accéder aux données gouvernées et les écrire à partir de SageMaker Studio :

1. À partir d'un bloc-notes SageMaker Studio, les utilisateurs peuvent utiliser des moteurs de requêtes tels qu'Amazon Athena ou des bibliothèques basées sur boto3 pour extraire des données directement vers le bloc-notes. Le kit SDK AWS pour Pandas (précédemment connu sous le nom de awswrangler) est une bibliothèque populaire. Voici un exemple de code pour montrer à quel point cela peut être simple :

```
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. Utilisez la connectivité native de SageMaker Studio à Amazon EMR pour lire et écrire des données à grande échelle. Grâce à l'utilisation des rôles d'exécution Apache Livy et Amazon EMR SageMaker , Studio a développé une connectivité native qui vous permet de transmettre SageMaker votre rôle IAM d'exécution (ou un autre rôle autorisé) à un cluster Amazon EMR pour l'accès aux données et leur traitement. Reportez-vous à la section Connect to an Amazon EMR cluster depuis Studio pour obtenir up-to-date des instructions.



Architecture d'accès aux données gérées par Lake Formation depuis SageMaker Studio

3. Utilisez la connectivité native de SageMaker Studio pour les sessions AWS Glue interactives afin de lire et d'écrire des données à grande échelle. SageMaker Les blocs-notes Studio sont dotés de noyaux intégrés qui permettent aux utilisateurs d'exécuter des commandes de manière interactive. AWS Glue Cela permet une utilisation évolutive des backends Python, Spark ou Ray qui peuvent lire et écrire des données en toute fluidité à grande échelle à partir de sources de données gouvernées. Les noyaux permettent aux utilisateurs de transmettre leur rôle SageMaker d'exécution ou d'autres rôles IAM autorisés. Reportez-vous à la section Préparation des données à l'aide de sessions AWS Glue interactives pour plus d'informations.

Rambardes communes

Cette section décrit les garde-fous les plus couramment utilisés pour appliquer la gouvernance à vos ressources ML à l'aide de politiques IAM, de politiques de ressources, de politiques de point de terminaison VPC et de politiques de contrôle des services (SCP).

Limitez l'accès au bloc-notes à des instances spécifiques

Cette politique de contrôle des services peut être utilisée pour limiter les types d'instances auxquels les data scientists ont accès lors de la création de blocs-notes Studio. Notez que tout utilisateur aura besoin de l'instance « système » autorisée pour créer l'application Jupyter Server par défaut qui héberge SageMaker Studio.

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

Limiter les domaines SageMaker Studio non conformes

Pour les domaines SageMaker Studio, la politique de contrôle des services suivante peut être utilisée pour obliger le trafic à accéder aux ressources des clients afin qu'ils ne passent pas par l'Internet public, mais via le VPC du client :

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

Limiter le lancement d' SageMaker images non autorisées

La politique suivante empêche un utilisateur de lancer une SageMaker image non autorisée dans son domaine : f

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

Lancez des ordinateurs portables uniquement via des points de terminaison SageMaker VPC

Outre les points de terminaison VPC pour le plan de SageMaker contrôle, prend en charge les points de terminaison SageMaker VPC permettant aux utilisateurs de se connecter aux blocs-notes Studio ou aux SageMaker instances de blocs-notes. SageMaker Si vous avez déjà configuré un point de terminaison VPC pour une instance SageMaker Studio/Notebook, la clé de condition IAM suivante n'autorisera les connexions aux blocs-notes Studio que si elles sont établies via le point de terminaison SageMaker Studio SageMaker VPC ou via le point de terminaison API. SageMaker

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

Limiter l'accès aux blocs-notes SageMaker Studio à une plage d'adresses IP limitée

Les entreprises limitent souvent l'accès à SageMaker Studio à certaines plages d'adresses IP d'entreprise autorisées. La politique IAM suivante avec la clé de `SourceIP` condition peut limiter cela.

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

Empêcher les utilisateurs de SageMaker Studio d'accéder à d'autres profils utilisateur

En tant qu'administrateur, lorsque vous créez le profil utilisateur, assurez-vous que le profil est étiqueté avec le nom d'utilisateur SageMaker Studio avec la clé de balise`studiouserid`. Le principal (utilisateur ou rôle attaché à l'utilisateur) doit également avoir une étiquette avec la clé `studiouserid` (cette balise peut porter n'importe quel nom et n'est pas limitée à`studiouserid`).

Ensuite, associez la politique suivante au rôle que l'utilisateur assumera lors du lancement de SageMaker Studio.

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

```
                "sagemaker:CreatePresignedDomainUrl"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "sagemaker:ResourceTag/studiouserid": "${aws:PrincipalTag/
studiouserid}"
                }
            }
        }
    ]
}
```

Appliquer le balisage

Les data scientists doivent utiliser les blocs-notes SageMaker Studio pour explorer les données, ainsi que pour créer et entraîner des modèles. L'application de balises aux ordinateurs portables permet de surveiller l'utilisation et de contrôler les coûts, tout en garantissant la propriété et l'auditabilité.

Pour les applications SageMaker Studio, assurez-vous que le profil utilisateur est balisé. Les balises sont automatiquement propagées aux applications à partir du profil utilisateur. Pour imposer la création de profils utilisateur à l'aide de balises (prises en charge par le biais de la CLI et du SDK), pensez à ajouter cette politique au rôle d'administrateur :

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

Pour les autres ressources, telles que les tâches de formation et les tâches de traitement, vous pouvez rendre les balises obligatoires en appliquant la politique suivante :

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

Accès root dans SageMaker Studio

Dans SageMaker Studio, le bloc-notes s'exécute dans un conteneur Docker qui, par défaut, ne dispose pas d'un accès root à l'instance hôte. De même, à l'exception de l'exécution en tant qu'utilisateur par défaut, toutes les autres plages d'ID utilisateur à l'intérieur du conteneur sont mappées à nouveau en tant qu'ID utilisateur non privilégiés sur l'instance hôte elle-même. Par conséquent, la menace d'augmentation des privilèges est limitée au conteneur de blocs-notes lui-même.

Lorsque vous créez des images personnalisées, vous souhaiterez peut-être accorder à votre utilisateur des autorisations non root pour des contrôles plus stricts, par exemple en évitant d'exécuter des processus indésirables en tant que root ou en installant des packages accessibles au public. Dans ce cas, vous pouvez créer l'image à exécuter en tant qu'utilisateur non root dans le Dockerfile. Que vous créiez l'utilisateur en tant que root ou non root, vous devez vous assurer que l'UID/GID de l'utilisateur est identique à l'UID/GID de l'application personnalisée, qui crée la configuration AppImageConfigpour SageMaker exécuter une application à l'aide de l'image

personnalisée. Par exemple, si votre Dockerfile est conçu pour un utilisateur non root, tel que celui-ci :

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

Le `AppImageConfig` fichier doit mentionner le même UID et le même GID dans son dossier : `KernelGatewayConfig`

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

Les valeurs UID/GID acceptables pour les images personnalisées sont 0/0 et 1000/100 pour les images Studio. Pour des exemples de création d'images personnalisées et des `AppImageConfig` paramètres associés, consultez ce référentiel Github.

Pour éviter que les utilisateurs n'altèrent cela, n'accordez pas les `DeleteAppImageConfig` autorisations ou `CreateAppImageConfig` `UpdateAppImageConfig` les autorisations aux utilisateurs du bloc-notes SageMaker Studio.

Gestion du réseau

Pour configurer le domaine SageMaker Studio, vous devez spécifier le réseau VPC, les sous-réseaux et les groupes de sécurité. Lorsque vous spécifiez le VPC et les sous-réseaux, assurez-vous d'allouer des adresses IP en tenant compte du volume d'utilisation et de la croissance attendue, comme indiqué dans les sections suivantes.

Planification du réseau VPC

Les sous-réseaux VPC du client associés au domaine SageMaker Studio doivent être créés avec la plage de routage interdomaine sans classe (CIDR) appropriée, en fonction des facteurs suivants :

- Nombre d'utilisateurs
- Nombre d'applications par utilisateur.
- Nombre de types d'instances uniques par utilisateur.
- Nombre moyen d'instances de formation par utilisateur.
- Pourcentage de croissance attendu.

SageMaker et les AWS services participants injectent des interfaces réseau élastiques (ENI) dans le sous-réseau VPC du client pour les cas d'utilisation suivants :

- Amazon EFS injecte une ENI pour une cible de montage EFS pour le SageMaker domaine (une adresse IP par sous-réseau/zone de disponibilité attachée au SageMaker domaine).
- SageMaker Studio injecte un ENI pour chaque instance unique utilisée par un profil utilisateur ou un espace partagé. Par exemple :
  - Si un profil utilisateur exécute une application serveur Jupyter par défaut (une instance « système »), une application Data Science et une application Base Python (toutes deux exécutées sur une `ml.t3.medium` instance), Studio injecte deux adresses IP.
  - Si un profil utilisateur exécute une application de serveur Jupyter par défaut (une instance « système »), une application GPU Tensorflow (sur une `ml.g4dn.xlarge` instance) et une application Data Wrangler (sur une `ml.m5.4xlarge` instance), Studio injecte trois adresses IP.
- Une ENI pour chaque point de terminaison VPC dans les sous-réseaux VPC/zones de disponibilité du domaine est injectée (quatre adresses IP pour les points de terminaison VPC ; environ six adresses IP pour les SageMaker points de terminaison VPC des services participants tels que S3, ECR et.) CloudWatch

- Si les tâches de SageMaker formation et de traitement sont lancées avec la même configuration VPC, chaque tâche nécessite deux adresses IP par instance.

> **Note**
>
> Les paramètres VPC de SageMaker Studio, tels que les sous-réseaux et le trafic uniquement VPC, ne sont pas automatiquement transmis aux tâches de formation/de traitement créées à partir de Studio. SageMaker L'utilisateur doit configurer les paramètres VPC et l'isolation du réseau selon les besoins lorsqu'il appelle les API Create*Job. Reportez-vous à la section Exécuter des conteneurs d'entraînement et d'inférence en mode sans Internet pour plus d'informations.

Scénario : un data scientist réalise des expériences sur deux types d'instances différents

Dans ce scénario, supposons qu'un SageMaker domaine soit configuré en mode trafic uniquement VPC. Des points de terminaison VPC sont configurés, tels que l' SageMakerAPI, le SageMaker runtime, Amazon S3 et Amazon ECR.

Un data scientist réalise des expériences sur des blocs-notes Studio, s'exécute sur deux types d'instances différents (par exemple, `ml.t3.medium` et`ml.m5.large`) et lance deux applications dans chaque type d'instance.

Supposons que le data scientist exécute également simultanément une tâche de formation avec la même configuration VPC sur une `ml.m5.4xlarge` instance.

Dans ce scénario, le service SageMaker Studio injectera des ENI comme suit :

Tableau 1 — Des ENI injectés dans le VPC du client pour un scénario d'expérimentation

Entité	Cible	ENI injecté	Remarques	Niveau
Cible de montage EFS	Sous-réseaux VPC	Trois	Trois AZS/sous-réseaux	Domaine
Points de terminaison d'un VPC	Sous-réseaux VPC	30	Trois AZS/sous-réseaux avec 10 VPCE chacun	Domaine

Entité	Cible	ENI injecté	Remarques	Niveau
Serveur Jupyter	Sous-réseau VPC	Un	Une adresse IP par instance	Utilisateur
KernelGateway appli	Sous-réseau VPC	Deux	Une adresse IP par type d'instance	Utilisateur
Entraînement	Sous-réseau VPC	Deux	Deux adresses IP par instance de formation Cinq adresses IP par instance de formation si EFA est utilisé	Utilisateur

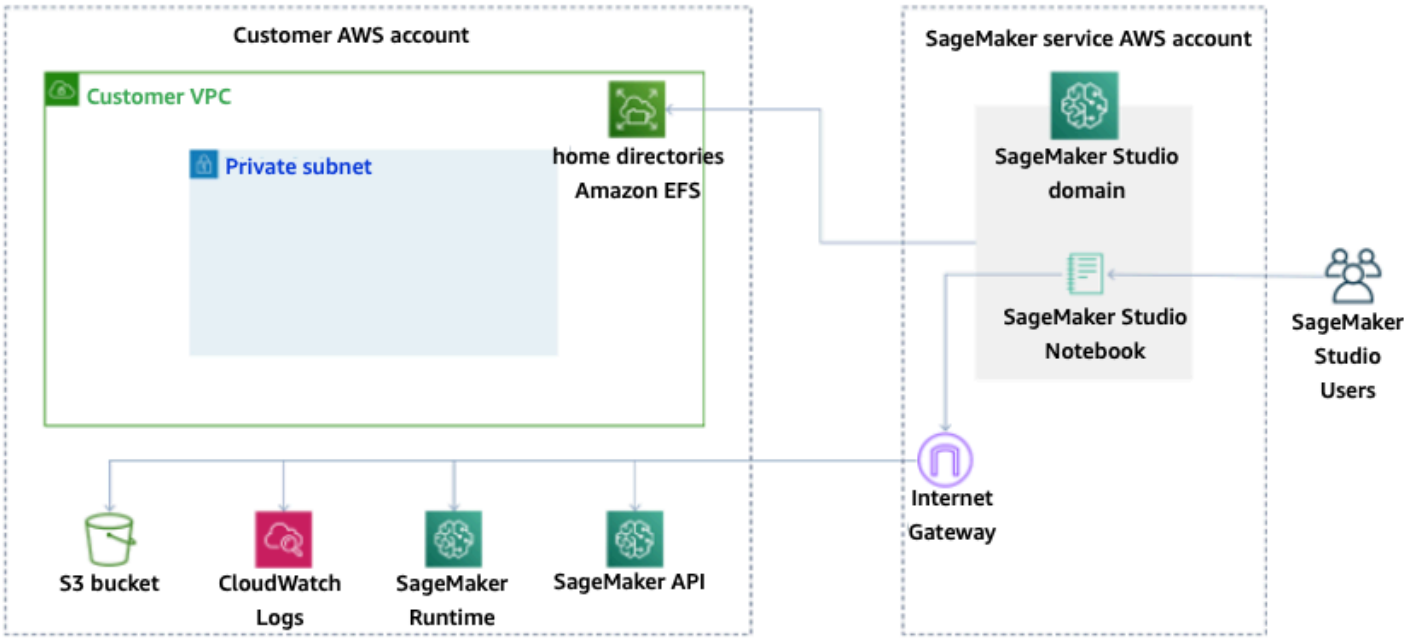
Dans ce scénario, 38 adresses IP sont consommées au total dans le VPC du client, 33 adresses IP étant partagées entre les utilisateurs au niveau du domaine et cinq adresses IP consommées au niveau utilisateur. Si 100 utilisateurs ayant des profils utilisateurs similaires dans ce domaine effectuent ces activités simultanément, vous consommerez cinq x 100 = 500 adresses IP au niveau utilisateur, en plus de la consommation d'adresses IP au niveau du domaine, qui est de 11 adresses IP par sous-réseau, pour un total de 511 adresses IP. Pour ce scénario, vous devez créer le sous-réseau VPC CIDR avec /22 qui allouera 1024 adresses IP, avec de la marge de croissance.

Options de réseau VPC

Un domaine SageMaker Studio prend en charge la configuration du réseau VPC avec l'une des options suivantes :

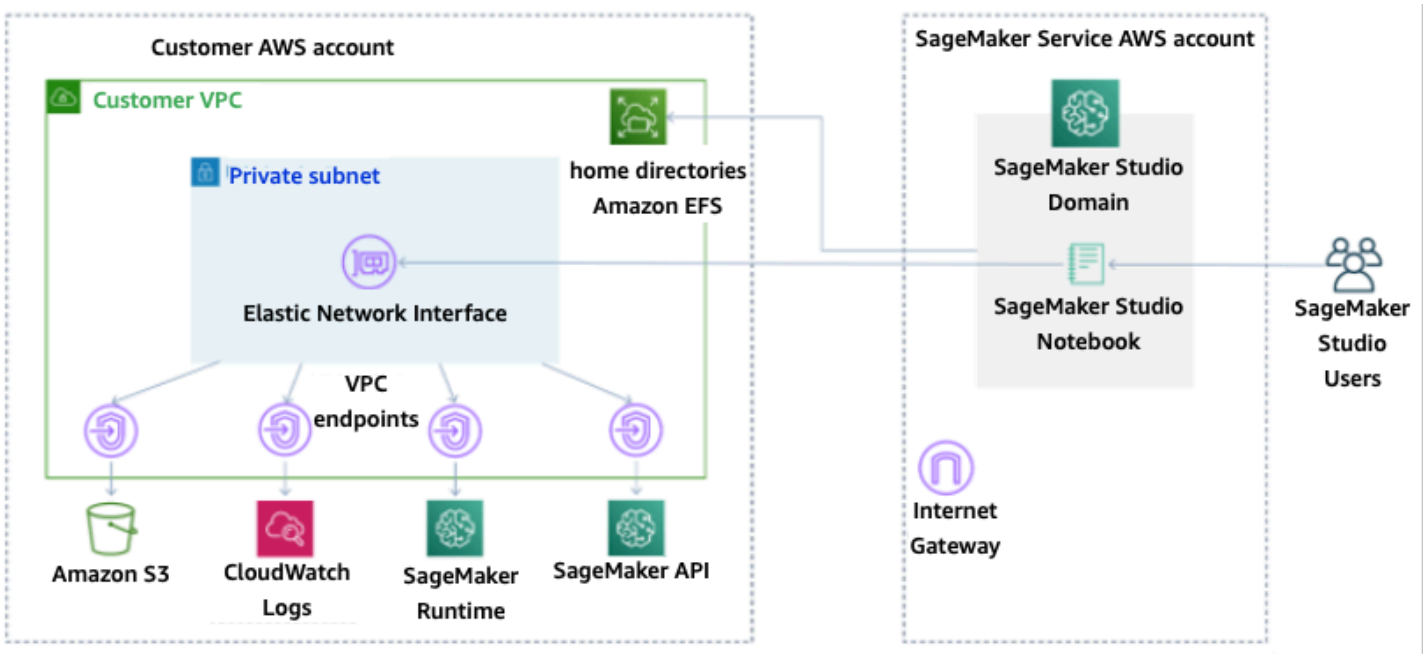
- Internet public uniquement
- VPC uniquement

L'option Internet public uniquement permet aux services d' SageMaker API d'utiliser l'Internet public via la passerelle Internet fournie dans le VPC, gérée par SageMaker le compte de service, comme le montre le schéma suivant :



Mode par défaut : accès à Internet via SageMaker un compte de service

L'option VPC uniquement désactive le routage Internet depuis le VPC géré par le compte de SageMaker service et permet au client de configurer le trafic à acheminer via les points de terminaison VPC, comme le montre le schéma suivant :



Mode VPC uniquement : pas d'accès à Internet via SageMaker un compte de service

Pour un domaine configuré en mode VPC uniquement, configurez un groupe de sécurité par profil utilisateur afin de garantir une isolation complète des instances sous-jacentes. Chaque domaine d'un AWS compte peut avoir sa propre configuration VPC et son propre mode Internet. Pour plus de détails concernant la configuration du réseau VPC, reportez-vous à Connect SageMaker Studio Notebooks d'un VPC à des ressources externes.

Limites

- Une fois qu'un domaine SageMaker Studio a été créé, vous ne pouvez pas associer de nouveaux sous-réseaux au domaine.
- Le type de réseau VPC (Internet public uniquement ou VPC uniquement) ne peut pas être modifié.

Protection des données

Avant de concevoir l'architecture d'une charge de travail ML, les pratiques fondamentales qui influencent la sécurité doivent être mises en place. Par exemple, la classification des données permet de classer les données en fonction de leur niveau de sensibilité, et le chiffrement protège les données en les rendant incompréhensibles pour tout accès non autorisé. Ces méthodes sont importantes, car elles répondent à des objectifs tels que la prévention des erreurs de manipulation ou le respect des obligations réglementaires.

SageMaker Studio propose plusieurs fonctionnalités pour protéger les données au repos et en transit. Cependant, comme décrit dans le modèle de responsabilité AWS partagée, les clients sont tenus de garder le contrôle sur le contenu hébergé sur l'infrastructure AWS mondiale. Dans cette section, nous décrivons comment les clients peuvent utiliser ces fonctionnalités pour protéger leurs données.

Protégez les données au repos

Pour protéger vos blocs-notes SageMaker Studio ainsi que vos données de création de modèles et vos artefacts, SageMaker chiffre les blocs-notes, ainsi que les résultats des tâches de formation et de transformation par lots. SageMaker les chiffre par défaut à l'aide de la clé AWS gérée pour Amazon S3. Cette clé gérée par AWS pour Amazon S3 ne peut pas être partagée pour l'accès entre comptes. Pour l'accès entre comptes, spécifiez votre clé gérée par le client lors de la création SageMaker des ressources afin qu'elle puisse être partagée pour un accès entre comptes.

Avec SageMaker Studio, les données peuvent être stockées dans les emplacements suivants :

- Compartiment S3 : lorsqu'un bloc-notes partageable est activé, SageMaker Studio partage les instantanés et les métadonnées du bloc-notes dans un compartiment S3.
- Volume EFS : SageMaker Studio attache un volume EFS à votre domaine pour stocker des blocs-notes et des fichiers de données. Ce volume EFS persiste même après la suppression du domaine.
- Volume EBS — EBS est attaché à l'instance sur laquelle le bloc-notes s'exécute. Ce volume est conservé pendant toute la durée de l'instance.

Chiffrement au repos avec AWS KMS

- Vous pouvez transmettre votre AWS KMSclé pour chiffrer un volume EBS connecté à des ordinateurs portables, à des formations, à des réglages, à des tâches de transformation par lots et à des terminaux.
- Si vous ne spécifiez pas de clé KMS, SageMaker chiffre à la fois les volumes du système d'exploitation (OS) et les volumes de données ML à l'aide d'une clé KMS gérée par le système.
- Les données sensibles qui doivent être chiffrées avec une clé KMS pour des raisons de conformité doivent être stockées dans le volume de stockage ML ou dans Amazon S3, tous deux pouvant être chiffrés à l'aide d'une clé KMS que vous spécifiez.

Protéger les données en transit

SageMaker Studio veille à ce que les artefacts du modèle ML et les autres artefacts du système soient chiffrés en transit et au repos. Les demandes adressées à la console et à l'API SageMaker sont envoyées par le biais d'une connexion sécurisée (SSL). Certaines données intra-réseau en transit (au sein de la plateforme de service) ne sont pas chiffrées. Cela comprend :

- Communications de commande et de contrôle entre le plan de contrôle de service et les instances de tâche d'entraînement (pas les données client).
- Communications entre les nœuds dans le cadre de tâches de traitement et de formation distribuées (intra-réseau).

Toutefois, vous pouvez choisir de chiffrer les communications entre les nœuds d'un cluster d'entraînement. L'activation du chiffrement du trafic entre conteneurs peut augmenter la durée de l'entraînement, surtout si vous utilisez des algorithmes de deep learning distribués.

Par défaut, Amazon SageMaker exécute des tâches de formation dans un Amazon VPC afin de garantir la sécurité de vos données. Pour protéger vos conteneurs d'entraînement et vos données, vous pouvez ajouter un autre niveau de sécurité en configurant un VPC privé. En outre, vous pouvez configurer votre domaine SageMaker Studio pour qu'il s'exécute en mode VPC uniquement et configurer des points de terminaison VPC pour acheminer le trafic sur un réseau privé sans le faire sortir par Internet.

Garde-corps de protection des données

Chiffrez les volumes SageMaker d'hébergement au repos

Utilisez la politique suivante pour appliquer le chiffrement lors de l'hébergement d'un SageMaker point de terminaison à des fins d'inférence en ligne :

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

Chiffrer les compartiments S3 utilisés lors de la surveillance des modèles

Model Monitoring capture les données envoyées à votre SageMaker terminal et les stocke dans un compartiment S3. Lorsque vous configurez la configuration de capture de données, vous devez chiffrer le compartiment S3. Il n'existe actuellement aucun contrôle compensatoire pour cela.

Outre la capture des résultats des terminaux, le service Model Monitoring vérifie la dérive par rapport à une base de référence prédéfinie. Vous devez chiffrer les sorties et les volumes de stockage intermédiaires utilisés pour surveiller la dérive.

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

```
        "Sid": "Encryption",
        "Effect": "Allow",
        "Action": [
            "sagemaker:CreateMonitoringSchedule",
            "sagemaker:UpdateMonitoringSchedule"
        ],
        "Resource": "*",
        "Condition": {
            "Null": {
                "sagemaker:VolumeKmsKey": "false",
                "sagemaker:OutputKmsKey": "false"
            }
        }
    }
  ]
}
```

Chiffrer un volume de stockage de domaine SageMaker Studio

Appliquez le chiffrement au volume de stockage attaché au domaine Studio. Cette politique oblige l'utilisateur à fournir une clé CMK pour chiffrer les volumes de stockage attachés aux domaines du studio.

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

Chiffrez les données stockées dans S3 qui sont utilisées pour partager des blocs-notes

Voici la politique qui permet de chiffrer toutes les données stockées dans le compartiment utilisé pour partager des blocs-notes entre les utilisateurs d'un domaine SageMaker Studio :

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

Limites

- Une fois qu'un domaine est créé, vous ne pouvez pas mettre à jour le volume de stockage EFS joint avec une AWS KMS clé personnalisée.
- Vous ne pouvez pas mettre à jour les tâches de formation/de traitement ou les configurations des terminaux avec des clés KMS une fois qu'elles ont été créées.

Journalisation et surveillance

Pour vous aider à déboguer vos tâches de compilation, vos tâches de traitement, vos tâches de formation, vos points de terminaison, vos tâches de transformation, vos instances de bloc-notes et vos configurations du cycle de vie des instances de bloc-notes, tout ce qu'un conteneur d'algorithmes, un conteneur de modèles ou une configuration du cycle de vie d'une instance de bloc-notes envoie à stdout ou stderr est également envoyé à Amazon Logs. CloudWatch Vous pouvez surveiller SageMaker Studio à l'aide d'Amazon CloudWatch, qui collecte les données brutes et les transforme en indicateurs lisibles en temps quasi réel. Ces statistiques sont conservées pendant 15 mois, afin que vous puissiez accéder à des informations historiques et avoir une meilleure idée des performances de votre application ou service Web.

Se connecter avec CloudWatch

Le processus de science des données étant intrinsèquement expérimental et itératif, il est essentiel de consigner les activités telles que l'utilisation des ordinateurs portables, le temps d'exécution des tâches de formation/de traitement, les indicateurs de formation et les indicateurs de service aux terminaux tels que la latence d'invocation. Par défaut, SageMaker publie les métriques dans CloudWatch Logs, et ces journaux peuvent être chiffrés à l'aide de clés gérées par le client à l'aide de. AWS KMS

Vous pouvez également utiliser des points de terminaison VPC pour envoyer des journaux CloudWatch sans utiliser l'Internet public. Vous pouvez également définir des alarmes qui surveillent certains seuils et envoient des notifications ou prennent des mesures lorsque ces seuils sont atteints. Pour plus d'informations, consultez le guide de CloudWatch l'utilisateur Amazon.

SageMaker crée un seul groupe de journaux pour Studio, sous `/aws/sagemaker/studio`. Chaque profil utilisateur et chaque application ont leur propre flux de journal dans ce groupe de journaux, et les scripts de configuration du cycle de vie ont également leur propre flux de journal. Par exemple, un profil utilisateur nommé « studio-user » associé à une application Jupyter Server associée à un script de cycle de vie, et à une application Data Science Kernel Gateway contient les flux de journaux suivants :

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

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

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

SageMaker Pour envoyer des journaux en votre CloudWatch nom, l'appelant des API de tâches Training/Processing/Transform aura besoin des autorisations suivantes :

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

Pour chiffrer ces journaux avec une AWS KMS clé personnalisée, vous devez d'abord modifier la politique de clé afin de permettre au CloudWatch service de chiffrer et de déchiffrer la clé. Une fois que vous avez créé une AWS KMS clé de chiffrement du journal, modifiez la politique de clé pour inclure les éléments suivants :

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

Notez que vous pouvez toujours utiliser `ArnEquals` et fournir un nom de ressource Amazon (ARN) spécifique pour le CloudWatch journal que vous souhaitez chiffrer. Nous montrons ici que vous pouvez utiliser cette clé pour crypter tous les journaux d'un compte pour plus de simplicité. En outre, les points de terminaison de formation, de traitement et de modélisation publient des métriques concernant l'utilisation du processeur et de la mémoire de l'instance, la latence d'invocation de l'hébergement, etc. Vous pouvez également configurer Amazon SNS pour informer les administrateurs des événements lorsque certains seuils sont dépassés. L'utilisateur des API de formation et de traitement doit disposer des autorisations suivantes :

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

Audit avec AWS CloudTrail

Pour améliorer votre niveau de conformité, auditez toutes vos API avecAWS CloudTrail. Par défaut, toutes les SageMaker API sont enregistrées avec AWS CloudTrail. Vous n'avez pas besoin d'autorisations IAM supplémentaires pour l'activer CloudTrail.

Toutes les SageMaker actions, à l'exception de `InvokeEndpoint` et`InvokeEndpointAsync`, sont enregistrées CloudTrail et documentées dans les opérations. Par exemple, les appels aux `CreateTrainingJob``CreateEndpoint`, et `CreateNotebookInstance` les actions génèrent des entrées dans les fichiers CloudTrail journaux.

Chaque entrée d' CloudTrail événement contient des informations sur la personne qui a généré la demande. Les informations relatives à l'identité permettent de déterminer les éléments suivants :

- Si la demande a été effectuée avec les autorisations utilisateur root ou IAM AWS.
- Si la demande a été effectuée avec des autorisations de sécurité temporaires pour un rôle ou un utilisateur fédéré.
- Si la requête a été effectuée par un autre service AWS. Pour un exemple d'événement, reportez-vous à la section Log SageMaker API Calls avec CloudTrail documentation.

Par défaut, CloudTrail enregistre le nom du rôle d'exécution Studio du profil utilisateur comme identifiant pour chaque événement. Cela fonctionne si chaque utilisateur a son propre rôle d'exécution. Si plusieurs utilisateurs partagent le même rôle d'exécution, vous pouvez utiliser la `sourceIdentity` configuration pour propager le nom du profil utilisateur Studio à CloudTrail. Reportez-vous à la section Surveillance de l'accès aux ressources utilisateur depuis Amazon SageMaker Studio pour activer `sourceIdentity` cette fonctionnalité. Dans un espace partagé, toutes les actions font référence à l'ARN de l'espace en tant que source, et vous ne pouvez pas effectuer d'audit par le biais de celui-ci`sourceIdentity`.

Attribution des coûts

SageMaker Studio intègre des fonctionnalités pour aider les administrateurs à suivre les dépenses de leurs domaines individuels, de leurs espaces partagés et de leurs utilisateurs.

Marquage automatique

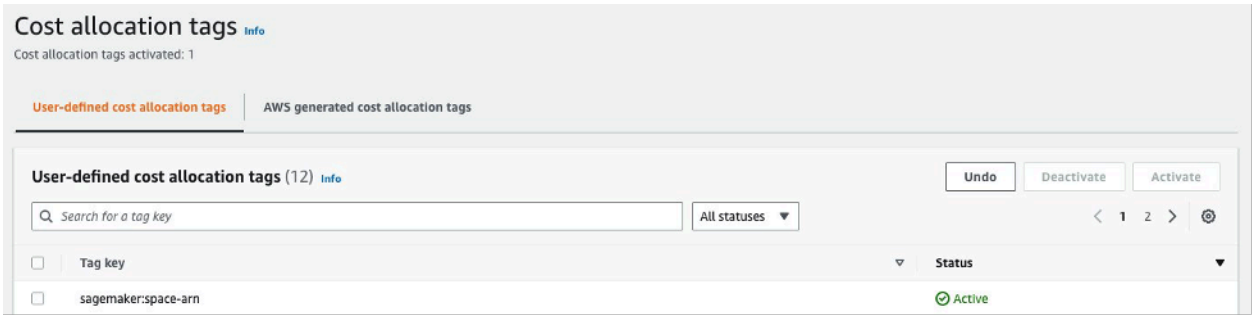
SageMaker Studio étiquette désormais automatiquement les nouvelles SageMaker ressources telles que les tâches de formation, les tâches de traitement et les applications du noyau avec leurs balises respectives`sagemaker:domain-arn`. À un niveau plus détaillé, étiquette SageMaker également la ressource avec le `sagemaker:user-profile-arn` ou `sagemaker:space-arn` pour désigner le créateur principal de la ressource.

SageMaker les volumes EFS de domaine sont étiquetés avec une clé nommée `ManagedByAmazonSageMakerResource` avec la valeur de l'ARN du domaine. Ils ne disposent pas de balises granulaires permettant de comprendre l'utilisation de l'espace au niveau de chaque utilisateur. Les administrateurs peuvent toutefois associer le volume EFS à une instance EC2 pour une surveillance personnalisée.

Suivi des coûts

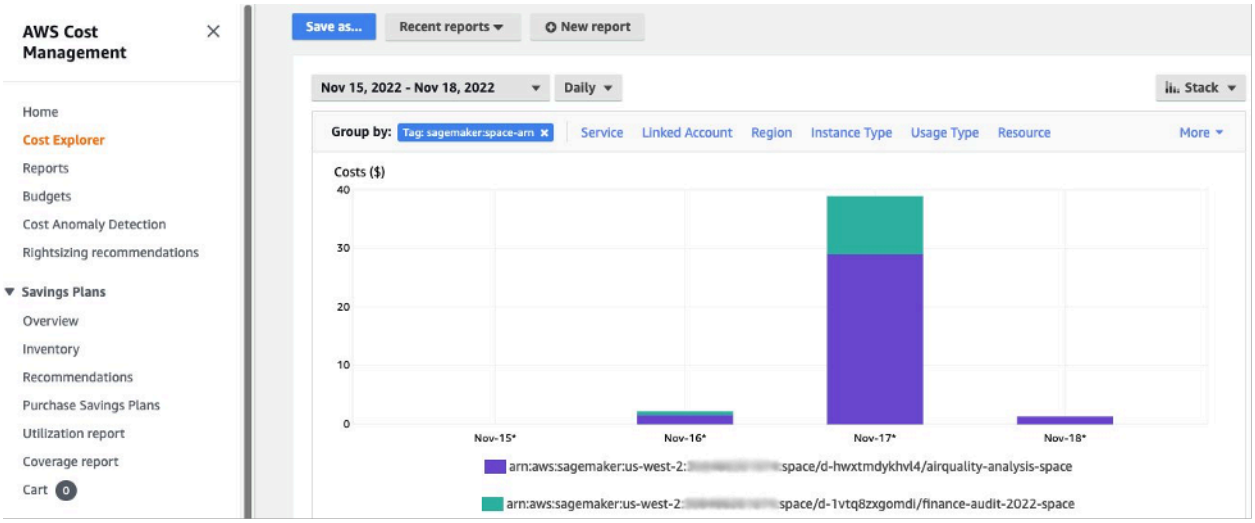
Les balises automatisées permettent aux administrateurs de suivre, de signaler et de surveiller vos dépenses en ML grâce à out-of-the-box des solutions telles que AWS Cost Exploreret AWS Budgets, ainsi qu'à des solutions personnalisées basées sur les données des rapports sur les AWS coûts et l'utilisation (CUR).

Pour utiliser les balises jointes à des fins d'analyse des coûts, elles doivent d'abord être activées dans la section Balises de répartition des coûts de la AWS Billing console. L'affichage des balises dans le panneau des balises de répartition des coûts peut prendre jusqu'à 24 heures. Vous devez donc créer une SageMaker ressource avant de les activer.



Space ARN activé en tant que balises de répartition des coûts sur Cost Explorer

Une fois que vous avez activé une balise de répartition des coûts, vous AWS commencerez à suivre vos ressources étiquetées, et après 24 à 48 heures, les balises apparaîtront sous forme de filtres sélectionnables dans l'explorateur de coûts.



Coûts regroupés par espace partagé pour un exemple de domaine

Contrôle des coûts

Lorsque le premier utilisateur de SageMaker Studio est intégré, il SageMaker crée un volume EFS pour le domaine. Des frais de stockage sont engagés pour ce volume EFS car les blocs-notes et les fichiers de données sont stockés dans le répertoire personnel de l'utilisateur. Lorsque l'utilisateur lance des blocs-notes Studio, ceux-ci sont lancés pour les instances de calcul qui exécutent les blocs-notes. Consultez les SageMaker tarifs d'Amazon pour une ventilation détaillée des coûts.

Les administrateurs peuvent contrôler les coûts de calcul en spécifiant la liste des instances qu'un utilisateur peut créer, en utilisant les politiques IAM mentionnées dans la section Garde-fous communs. En outre, nous recommandons aux clients d'utiliser l'extension d'arrêt automatique de SageMaker Studio pour réduire les coûts en fermant automatiquement les applications inactives. Cette extension de serveur interroge régulièrement les applications en cours d'exécution par profil utilisateur et arrête les applications inactives en fonction d'un délai défini par l'administrateur.

Pour définir cette extension pour tous les utilisateurs de votre domaine, vous pouvez utiliser une configuration de cycle de vie telle que décrite dans la section Personnalisation. En outre, vous pouvez également utiliser le vérificateur d'extension pour vous assurer que l'extension est installée sur tous les utilisateurs de votre domaine.

Personnalisation

Configuration du cycle de vie

Les configurations du cycle de vie sont des scripts shell initiés par des événements du cycle de vie de SageMaker Studio, tels que le démarrage d'un nouveau bloc-notes SageMaker Studio. Vous pouvez utiliser ces scripts shell pour automatiser la personnalisation de vos environnements SageMaker Studio, comme l'installation de packages personnalisés, l'extension Jupyter pour l'arrêt automatique des applications de bloc-notes inactives et la configuration de Git. Pour obtenir des instructions détaillées sur la façon de créer des configurations de cycle de vie, consultez ce blog : Personnaliser Amazon SageMaker Studio à l'aide des configurations de cycle de vie.

Images personnalisées pour les blocs-notes SageMaker Studio

Les blocs-notes Studio sont fournis avec un ensemble d'images prédéfinies, qui comprennent le SDK Amazon SageMaker Python et la dernière version du runtime ou du noyau IPython. Grâce à cette fonctionnalité, vous pouvez ajouter vos propres images personnalisées aux SageMaker carnets Amazon. Ces images sont ensuite accessibles à tous les utilisateurs authentifiés dans le domaine.

Les développeurs et les data scientists peuvent avoir besoin d'images personnalisées pour différents cas d'utilisation :

- Accès à des versions spécifiques ou récentes de frameworks ML courants tels que TensorFlow MXnet ou PyTorch autres.
- Importez du code personnalisé ou des algorithmes développés localement dans les blocs-notes SageMaker Studio pour accélérer l'itération et l'apprentissage des modèles.
- Accès aux lacs de données ou aux magasins de données sur site via des API. Les administrateurs doivent inclure les pilotes correspondants dans l'image.
- Accès à un environnement d'exécution principal (également appelé noyau), autre que IPython (tel que R, Julia ou autres). Vous pouvez également utiliser l'approche décrite pour installer un noyau personnalisé.

Pour obtenir des instructions détaillées sur la création d'une image personnalisée, reportez-vous à la section Création d'une SageMaker image personnalisée.

JupyterLab extensions

Avec SageMaker Studio JuypterLab 3 Notebook, vous pouvez tirer parti de la communauté toujours croissante d'extensions open source JupyterLab. Cette section met en évidence quelques-unes qui s'intègrent naturellement dans le flux de travail des SageMaker développeurs, mais nous vous encourageons à parcourir les extensions disponibles ou même à créer les vôtres.

JupyterLab 3 facilite désormais considérablement le processus d'empaquetage et d'installation des extensions. Vous pouvez installer les extensions susmentionnées par le biais de scripts bash. Par exemple, dans SageMaker Studio, ouvrez le terminal système à partir du lanceur Studio et exécutez les commandes suivantes. En outre, vous pouvez automatiser l'installation de ces extensions à l'aide de configurations de cycle de vie afin qu'elles soient conservées entre les redémarrages de Studio. Vous pouvez le configurer pour tous les utilisateurs du domaine ou au niveau d'un utilisateur individuel.

Par exemple, pour installer une extension pour un navigateur de fichiers Amazon S3, exécutez les commandes suivantes dans le terminal système et assurez-vous d'actualiser votre navigateur :

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

Pour plus d'informations sur la gestion des extensions, notamment sur la façon d'écrire des configurations de cycle de vie qui fonctionnent à la fois pour les versions 1 et 3 des JupyterLab ordinateurs portables à des fins de rétrocompatibilité, reportez-vous à la section Installation JupyterLab et extensions Jupyter Server.

Référentiels Git

SageMaker Studio est préinstallé avec une extension Jupyter Git permettant aux utilisateurs de saisir l'URL personnalisée d'un dépôt Git, de le cloner dans votre répertoire EFS, d'effectuer des modifications et de consulter l'historique des validations. Les administrateurs peuvent configurer les dépôts git suggérés au niveau du domaine afin qu'ils apparaissent sous forme de listes déroulantes pour les utilisateurs finaux. Reportez-vous à la section Attacher des dépôts Git suggérés à Studio pour up-to-date obtenir des instructions.

Si un dépôt est privé, l'extension demandera à l'utilisateur de saisir ses informations d'identification dans le terminal à l'aide de l'installation git standard. L'utilisateur peut également stocker les informations d'identification SSH dans son répertoire EFS individuel pour en faciliter la gestion.

Environnement Conda

SageMaker Les blocs-notes Studio utilisent Amazon EFS comme couche de stockage persistante. Les data scientists peuvent utiliser le stockage persistant pour créer des environnements conda personnalisés et utiliser ces environnements pour créer des noyaux. Ces noyaux sont soutenus par EFS et sont persistants entre les redémarrages du noyau, de l'application ou de Studio. Studio sélectionne automatiquement tous les environnements valides sous forme de KernelGateway noyaux.

Le processus de création d'un environnement conda est simple pour un data scientist, mais les noyaux mettent environ une minute à être renseignés dans le sélecteur de noyau. Pour créer un environnement, exécutez ce qui suit dans un terminal système :

```
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
```

Pour obtenir des instructions détaillées, reportez-vous à la section relative aux environnements Persist Conda du volume Studio EFS dans Quatre approches pour gérer les packages Python dans les blocs-notes Amazon SageMaker Studio.

Conclusion

Dans ce livre blanc, nous avons passé en revue plusieurs bonnes pratiques dans des domaines tels que le modèle d'exploitation, la gestion des domaines, la gestion des identités, la gestion des autorisations, la gestion du réseau, la journalisation, la surveillance et la personnalisation afin de permettre aux administrateurs de la plateforme de configurer et de gérer SageMaker Studio Platform.

Annexe

Comparaison entre plusieurs locataires

Tableau 2 — Comparaison entre plusieurs locataires

Multi-domaines	Comptes multiples	Contrôle d'accès basé sur les attributs (ABAC) au sein d'un seul domaine
L'isolation des ressources est réalisée à l'aide de balises. SageMaker Studio étiquette automatiquement toutes les ressources avec l'ARN du domaine et l'ARN du profil/espace utilisateur.	Chaque locataire a son propre compte, il y a donc une isolation absolue des ressources.	L'isolation des ressources est réalisée à l'aide de balises. Les utilisateurs doivent gérer le balisage des ressources créées pour ABAC.
Les API de liste ne peuvent pas être limitées par des balises. Le filtrage des ressources par l'interface utilisateur est effectué sur les espaces partagés, mais les appels d'API List effectués via le SDK Boto3 AWS CLI ou via le SDK Boto3 listeront les ressources de la région.	L'isolation des API de liste est également possible, car les locataires se trouvent dans leurs comptes dédiés.	Les API de liste ne peuvent pas être limitées par des balises. Les appels d'API de liste effectués via le SDK Boto3 AWS CLI ou le SDK Boto3 répertorieront les ressources de la région.
SageMaker Les coûts de calcul et de stockage en studio par locataire peuvent être facilement surveillés en utilisant l'ARN du domaine comme balise de répartition des coûts.	SageMaker Les coûts de calcul et de stockage en studio par locataire sont faciles à contrôler avec un compte dédié.	SageMaker Les coûts de calcul du studio par locataire doivent être calculés à l'aide de balises personnalisées. SageMaker Les coûts de stockage du studio ne peuvent

Multi-domaines	Comptes multiples	Contrôle d'accès basé sur les attributs (ABAC) au sein d'un seul domaine
		pas être surveillés par domaine puisque tous les locataires partagent le même volume EFS.
Les quotas de service sont définis au niveau du compte, de sorte qu'un seul locataire peut toujours utiliser toutes les ressources.	Les quotas de service peuvent être définis au niveau du compte pour chaque locataire.	Les quotas de service sont définis au niveau du compte, de sorte qu'un seul locataire peut toujours utiliser toutes les ressources.
La mise à l'échelle vers plusieurs locataires peut être réalisée par le biais de l'infrast ructure sous forme de code (IaC) ou du Service Catalog.	L'extension à plusieurs locataires implique des Organisations et la vente de plusieurs comptes.	Le dimensionnement nécessite un rôle spécifiqu e au locataire pour chaque nouveau locataire, et les profils utilisateur doivent être étiquetés manuellement avec les noms des locataires.
La collaboration entre les utilisateurs au sein d'un locataire est possible grâce à des espaces partagés.	La collaboration entre utilisate urs au sein d'un locataire est possible grâce à des espaces partagés.	Tous les locataires auront accès au même espace partagé pour la collaboration.

SageMaker Sauvegarde et restauration de domaines Studio

En cas de suppression accidentelle d'EFS, ou lorsqu'un domaine doit être recréé en raison de modifications apportées au réseau ou à l'authentification, suivez ces instructions.

Option 1 : sauvegarde à partir d'un EFS existant à l'aide d'EC2

SageMaker Sauvegarde du domaine Studio

1. Répertoriez les profils utilisateur et les espaces dans SageMaker Studio (CLI, SDK).

2. Mappez les profils/espaces utilisateur aux UID sur EFS.
   a. Pour chaque utilisateur figurant dans la liste des utilisateurs/espaces, décrivez le profil/espace utilisateur (CLI, SDK).
   b. Mappez le profil/espace utilisateur à. `HomeEfsFileSystemUid`
   c. Mappez le profil utilisateur selon `UserSettings['ExecutionRole']` si les utilisateurs ont des rôles d'exécution distincts.
   d. Identifiez le rôle d'exécution de Space par défaut.
3. Créez un nouveau domaine et spécifiez le rôle d'exécution par défaut de Space.
4. Créez des profils utilisateur et des espaces.
   - Pour chaque utilisateur de la liste d'utilisateurs, créez un profil utilisateur (CLI, SDK) à l'aide du mappage des rôles d'exécution.
5. Créez un mappage pour les nouveaux EFS et UID.
   a. Pour chaque utilisateur de la liste d'utilisateurs, décrivez le profil utilisateur (CLI, SDK).
   b. Associer le profil utilisateur à`HomeEfsFileSystemUid`.
6. Vous pouvez éventuellement supprimer toutes les applications, tous les profils utilisateur, tous les espaces, puis supprimer le domaine.

Sauvegarde EFS

Pour sauvegarder le fichier EFS, suivez les instructions suivantes :

1. Lancez l'instance EC2 et associez les groupes de sécurité entrants/sortants de l'ancien domaine SageMaker Studio à la nouvelle instance EC2 (autorisez le trafic NFS via TCP sur le port 2049). Reportez-vous à la section Connecter les blocs-notes SageMaker Studio d'un VPC à des ressources externes.
2. Montez le volume SageMaker Studio EFS sur la nouvelle instance EC2. Reportez-vous à la section Montage des systèmes de fichiers EFS.
3. Copiez les fichiers sur le stockage local EBS : `>sudo cp -rp /efs /studio-backup:`
   a. Attachez les nouveaux groupes de sécurité de domaine à l'instance EC2.
   b. Montez le nouveau volume EFS sur l'instance EC2.
   c. Copiez les fichiers sur le nouveau volume EFS.
   d. Pour chaque utilisateur de la collection de l'utilisateur :
      i. Créez le répertoire :`mkdir new_uid`.

ii. Copiez les fichiers de l'ancien répertoire UID vers le nouveau répertoire UID.

iii. Changer de propriétaire pour tous les fichiers : `chown <new_UID>` pour tous les fichiers.

Option 2 : sauvegarde à partir d'un EFS existant à l'aide de S3 et de la configuration du cycle de vie

1. Reportez-vous à la section Migrer votre travail vers une instance de SageMaker bloc-notes Amazon avec Amazon Linux 2.
2. Créez un compartiment S3 pour la sauvegarde (par exemple>`studio-backup`.
3. Répertoriez tous les profils utilisateur dotés de rôles d'exécution.
4. Dans le domaine SageMaker Studio actuel, définissez un script LCC par défaut au niveau du domaine.
   - Dans le LCC, copiez tout dans `/home/sagemaker-user` le préfixe du profil utilisateur dans S3 (par exemple,`s3://studio-backup/studio-user1`).
5. Redémarrez toutes les applications Jupyter Server par défaut (pour que le LCC soit exécuté).
6. Supprimez toutes les applications, tous les profils utilisateur et tous les domaines.
7. Créez un nouveau domaine SageMaker Studio.
8. Créez de nouveaux profils utilisateur à partir de la liste des profils utilisateur et des rôles d'exécution.
9. Configurez un LCC au niveau du domaine :
   - Dans le LCC, copiez tout ce qui se trouve dans le préfixe du profil utilisateur dans S3 vers `/home/sagemaker-user`
10. Créez des applications Jupyter Server par défaut pour tous les utilisateurs avec la configuration LCC (CLI, SDK).

SageMaker Accès au studio à l'aide d'une assertion SAML

Configuration de la solution :

1. Créez une application SAML dans votre IdP externe.
2. Configurez l'IdP externe en tant que fournisseur d'identité dans IAM.
3. Créez une fonction `SAMLValidator` Lambda accessible à l'IdP (via une URL de fonction ou une API Gateway).

4. Créez une fonction `GeneratePresignedUrl` Lambda et une API Gateway pour accéder à la fonction.
5. Créez un rôle IAM que les utilisateurs peuvent assumer pour appeler l'API Gateway. Ce rôle doit être transmis dans une assertion SAML sous forme d'attribut au format suivant :
   - Nom de l'attribut : https://aws.amazon.com/SAML/Attributes/Role
   - Valeur de l'attribut :`<IdentityProviderARN>`, `<RoleARN>`
6. Mettez à jour le point de terminaison SAML Assertion Consumer Service (ACS) vers l'URL `SAMLValidator` d'appel.

Exemple de code de validateur SAML :

```
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
```

Suggestions de lecture

- Configuration d'environnements d'apprentissage automatique sécurisés et bien gérés sur AWS (AWSblog)
- Configuration d'Amazon SageMaker Studio pour les équipes et les groupes avec isolation complète des ressources (AWSblog)
- Intégration d'Amazon SageMaker Studio avec le AWS SSO et Okta Universal Directory (blog) AWS
- Comment configurer SAML 2.0 pour la fédération de AWS comptes (documentation Okta)
- Création d'une plateforme de Machine Learning d'entreprise sécurisée sur AWS (guide AWS technique)
- Personnalisez Amazon SageMaker Studio à l'aide des configurations du cycle de vie (AWSblog)
- Intégrer votre propre image de conteneur personnalisée aux blocs-notes Amazon SageMaker Studio (AWSblog)
- Création de modèles de SageMaker projets personnalisés — Meilleures pratiques (AWSblog)
- Déploiement de modèles multi-comptes avec Amazon SageMaker Pipelines (AWSblog)
- Partie 1 : Comment le NatWest groupe a créé une plateforme mLOPS évolutive, sécurisée et durable (blog) AWS
- URL présignées Amazon SageMaker Studio sécurisées, partie 1 : infrastructure de base (blog) AWS

Collaborateurs

Les contributeurs à ce document incluent :

- Ram Vittal, architecte de solutions ML, Amazon Web Services
- Sean Morgan, architecte de solutions ML, Amazon Web Services
- Durga Sury, architecte de solutions ML, Amazon Web Services

Nous remercions tout particulièrement les personnes suivantes qui ont apporté des idées, des révisions et des points de vue :

- Alessandro Cerè, architecte de solutions d'intelligence artificielle et d'apprentissage automatique, Amazon Web Services
- Sumit Thakur, chef de SageMaker produit, Amazon Web Services
- Han Zhang, ingénieur principal en développement logiciel, Amazon Web Services
- Bhadrinath Pani, ingénieur en développement logiciel, Amazon Web Services, Amazon Web Services

Révisions du document

Pour être informé des mises à jour de ce livre blanc, abonnez-vous au flux RSS.

Modification	Description	Date
Livre blanc mis à jour	Les liens rompus ont été corrigés et de nombreuses modifications éditoriales ont été apportées.	25 avril 2023
Publication initiale	Livre blanc publié.	19 octobre 2022

Avis

Les clients sont tenus de procéder à leur propre évaluation indépendante des informations contenues dans ce document. Ce document : (a) est fourni à titre informatif uniquement, (b) représente les offres de AWS produits et les pratiques actuelles, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ni aucune assurance de la part de AWS ses filiales, fournisseurs ou concédants de licence. AWSles produits ou services sont fournis « tels quels » sans garantie, représentation ou condition d'aucune sorte, expresse ou implicite. Les responsabilités et obligations AWS de ses clients sont régies par AWS des accords, et ce document ne fait partie d'aucun accord conclu entre AWS et ses clients et ne les modifie pas.

Glossaire AWS

Pour connaître la terminologie la plus récente d’AWS, consultez le Glossaire AWS dans la Référence Glossaire AWS.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.