AWS Whitepaper

SageMaker Praktik Terbaik Administrasi Studio

SageMaker Praktik Terbaik Administrasi Studio: AWS Whitepaper

Table of Contents

SageMaker Praktik Terbaik Administrasi Studio

Tanggal publikasi: 25 April 2023 () Revisi dokumen

Abstrak

Amazon SageMaker Studio menyediakan antarmuka visual tunggal berbasis web tempat Anda dapat melakukan semua langkah pengembangan pembelajaran mesin (ML), yang meningkatkan produktivitas tim ilmu data. SageMaker Studio memberi Anda akses, kontrol, dan visibilitas lengkap ke setiap langkah yang diperlukan untuk membangun, melatih, dan mengevaluasi model.

Dalam whitepaper ini, kami membahas praktik terbaik untuk subjek termasuk model operasi, manajemen domain, manajemen identitas, manajemen izin, manajemen jaringan, pencatatan, pemantauan, dan penyesuaian. Praktik terbaik yang dibahas di sini ditujukan untuk penerapan SageMaker Studio perusahaan, termasuk penerapan multi-penyewa. Dokumen ini ditujukan untuk administrator platform ML, insinyur ML, dan arsitek ML.

Apakah Anda Well-Architected?

AWSWell-Architected Framework membantu Anda memahami pro dan kontra dari keputusan yang Anda buat saat membangun sistem di cloud. Enam pilar Kerangka memungkinkan Anda mempelajari praktik terbaik arsitektur untuk merancang dan mengoperasikan sistem yang andal, aman, efisien, hemat biaya, dan berkelanjutan. Dengan menggunakan AWS Well-Architected Tool, tersedia tanpa biaya di AWS Management Console, Anda dapat meninjau beban kerja Anda terhadap praktik terbaik ini dengan menjawab serangkaian pertanyaan untuk setiap pilar.

Di Machine Learning Lens, kami fokus pada cara merancang, menyebarkan, dan merancang beban kerja pembelajaran mesin Anda di. AWS Cloud Lensa ini menambah praktik terbaik yang dijelaskan dalam Well-Architected Framework.

Pengantar

Saat Anda mengelola SageMaker Studio sebagai platform ML Anda, Anda memerlukan panduan praktik terbaik untuk membuat keputusan yang tepat guna membantu Anda menskalakan platform ML seiring bertambahnya beban kerja Anda. Untuk menyediakan, mengoperasionalkan, dan menskalakan platform ML Anda, pertimbangkan hal berikut:

- Pilih model operasi yang tepat dan atur lingkungan ML Anda untuk memenuhi tujuan bisnis Anda.
- Pilih cara mengatur otentikasi domain SageMaker Studio untuk identitas pengguna, dan pertimbangkan batasan tingkat domain.
- Putuskan cara menggabungkan identitas dan otorisasi pengguna Anda ke platform ML untuk kontrol akses dan audit yang berbutir halus.
- Pertimbangkan untuk menyiapkan izin dan pagar pembatas untuk berbagai peran persona ML Anda.
- Rencanakan topologi jaringan virtual private cloud (VPC) Anda, dengan mempertimbangkan sensitivitas beban kerja, jumlah pengguna, jenis instans, aplikasi, dan pekerjaan yang diluncurkan.
- Klasifikasi dan lindungi data Anda saat istirahat dan dalam perjalanan dengan enkripsi.
- Pertimbangkan cara mencatat dan memantau berbagai antarmuka pemrograman aplikasi (API) dan aktivitas pengguna untuk kepatuhan.
- Sesuaikan pengalaman notebook SageMaker Studio dengan gambar dan skrip konfigurasi siklus hidup Anda sendiri.

Model operasi

Model operasi adalah kerangka kerja yang menyatukan orang, proses, dan teknologi untuk membantu organisasi memberikan nilai bisnis dengan cara yang terukur, konsisten, dan efisien. Model operasi ML menyediakan proses pengembangan produk standar untuk tim di seluruh organisasi. Ada tiga model untuk menerapkan model operasi, tergantung pada ukuran, kompleksitas, dan driver bisnis:

- Tim ilmu data terpusat — Dalam model ini, semua kegiatan ilmu data terpusat dalam satu tim atau organisasi. Ini mirip dengan model Center of Excellence (COE), di mana semua unit bisnis masuk ke tim ini untuk proyek ilmu data.
- Tim ilmu data terdesentralisasi — Dalam model ini, kegiatan ilmu data didistribusikan di berbagai fungsi atau divisi bisnis, atau berdasarkan lini produk yang berbeda.
- Tim ilmu data federasi — Dalam model ini, fungsi layanan bersama seperti repositori kode, integrasi berkelanjutan dan pipa pengiriman berkelanjutan (CI/CD), dan sebagainya dikelola oleh tim terpusat, dan setiap unit bisnis atau fungsi tingkat produk dikelola oleh tim terdesentralisasi. Ini mirip dengan model hub dan spoke, di mana setiap unit bisnis memiliki tim ilmu data mereka sendiri; Namun, tim unit bisnis ini mengoordinasikan kegiatan mereka dengan tim terpusat.

Sebelum memutuskan untuk meluncurkan domain studio pertama Anda untuk kasus penggunaan produksi, pertimbangkan model operasi dan praktik AWS terbaik untuk mengatur lingkungan Anda. Untuk informasi selengkapnya, lihat Mengatur AWS Lingkungan Anda Menggunakan Beberapa Akun.

Bagian selanjutnya memberikan panduan tentang mengatur struktur akun Anda untuk masing-masing model operasi.

Struktur akun yang direkomendasikan

Pada bagian ini, kami secara singkat memperkenalkan struktur akun model operasi yang dapat Anda mulai dan modifikasi sesuai dengan persyaratan operasi organisasi Anda. Terlepas dari model operasi yang Anda pilih, kami sarankan untuk menerapkan praktik terbaik umum berikut:

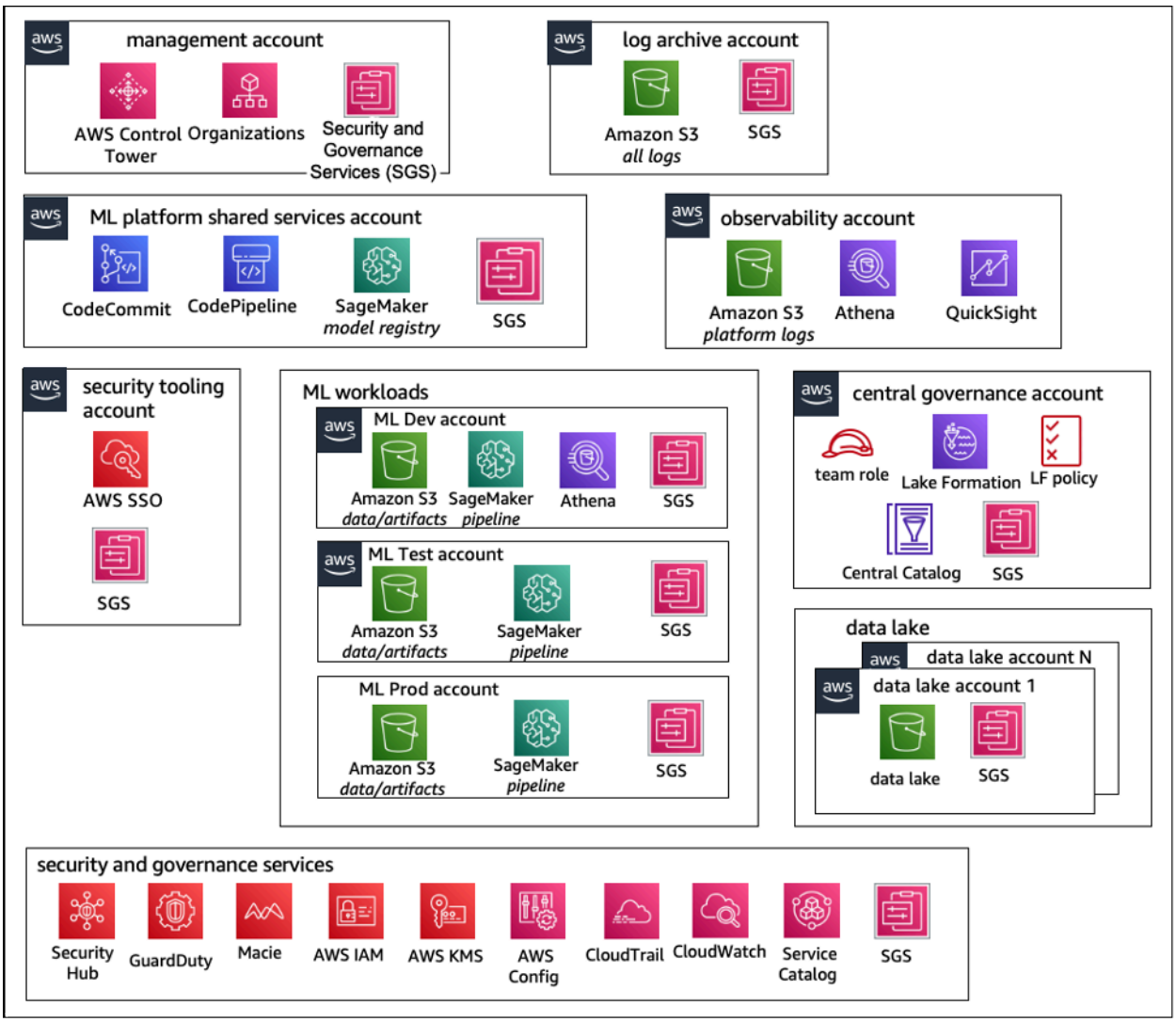
- Gunakan AWS Control Toweruntuk penyiapan, pengelolaan, dan tata kelola akun Anda.
- Pusatkan identitas Anda dengan Penyedia Identitas (IDP), dan Pusat Identitas AWS IAM dengan akun Securitiy Tooling administrator yang didelegasikan dan aktifkan akses aman ke beban kerja.

- Jalankan beban kerja ML dengan isolasi tingkat akun di seluruh beban kerja pengembangan, pengujian, dan produksi.
- Streaming log beban kerja ML ke akun arsip log, lalu filter dan terapkan analisis log di akun observabilitas.
- Jalankan akun tata kelola terpusat untuk penyediaan, pengendalian, dan audit akses data.
- Sematkan layanan keamanan dan tata kelola (SGS) dengan pagar pembatas preventif dan detektif yang sesuai ke dalam setiap akun untuk memastikan keamanan dan kepatuhan, sesuai dengan persyaratan organisasi dan beban kerja Anda.

Struktur akun model terpusat

Dalam model ini, tim platform ML bertanggung jawab untuk menyediakan:

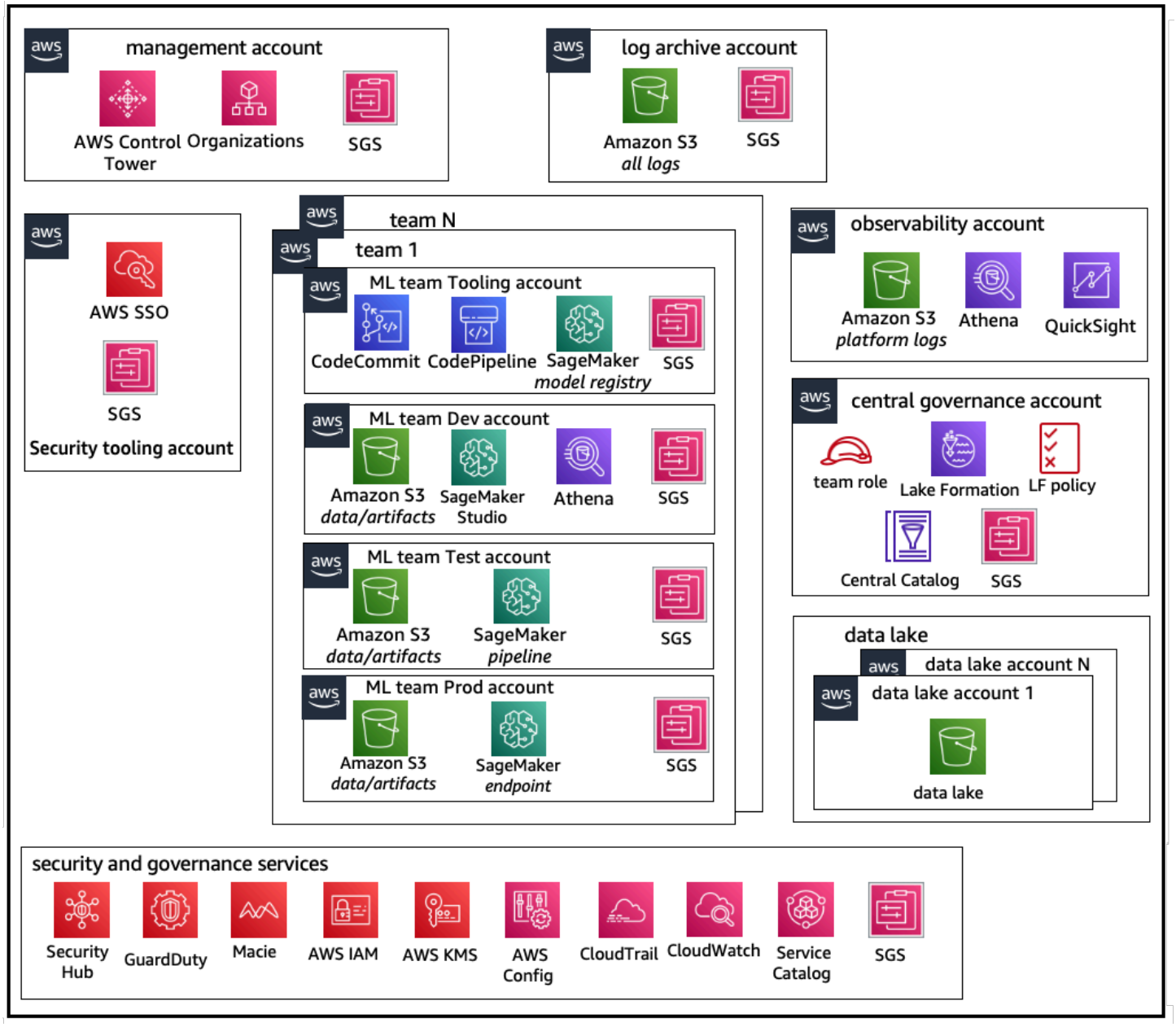
- Akun perkakas layanan bersama yang membahas persyaratan Machine Learning Operations (MLOPs) di seluruh tim ilmu data.
- Akun pengembangan, pengujian, dan produksi beban kerja ML yang dibagikan di seluruh tim ilmu data.
- Kebijakan tata kelola untuk memastikan setiap beban kerja tim ilmu data berjalan secara terpisah.
- Praktik terbaik yang umum.



Struktur akun model operasi terpusat

Struktur akun model terdesentralisasi

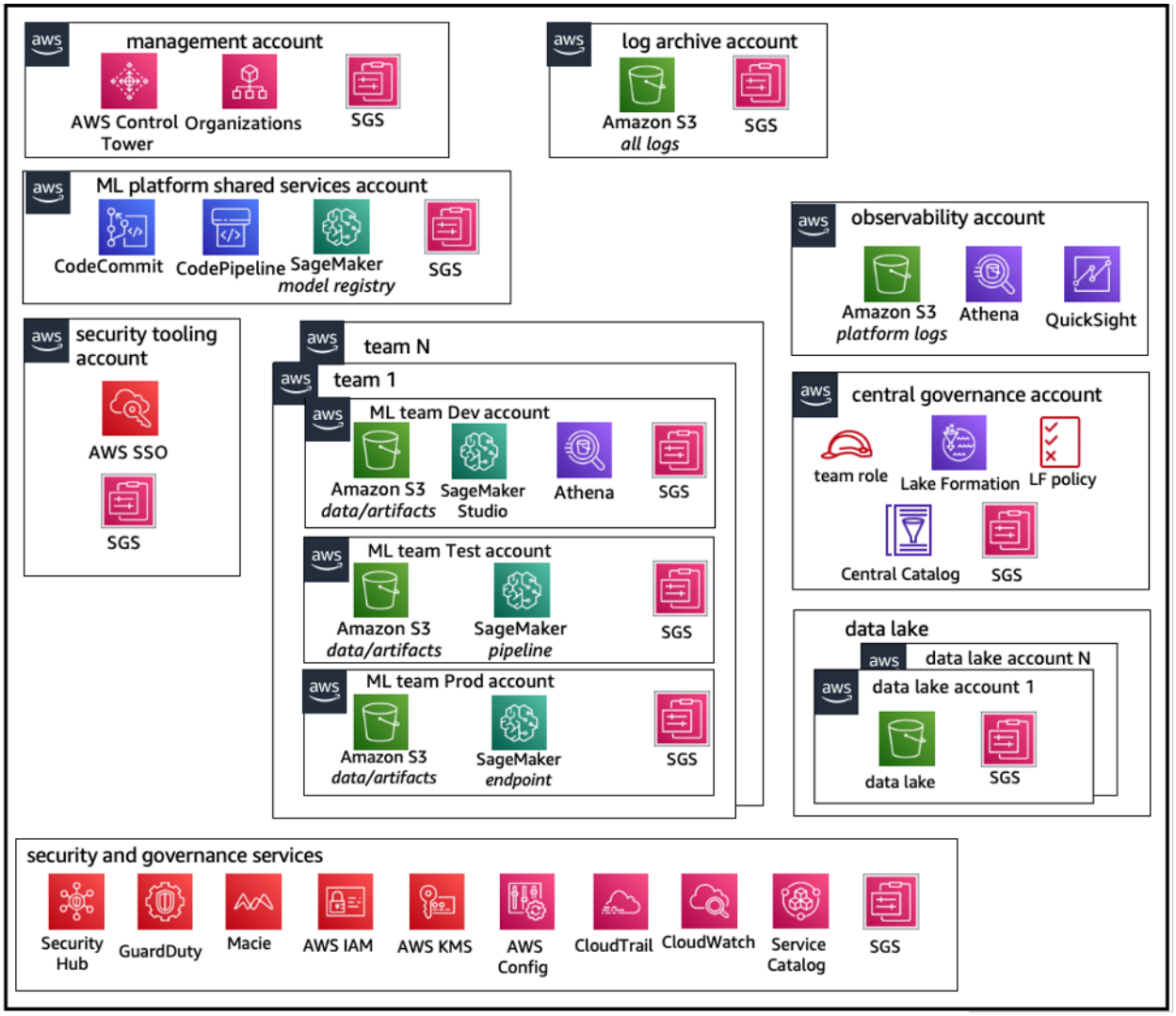
Dalam model ini, setiap tim ML beroperasi secara independen untuk menyediakan, mengelola, dan mengatur akun dan sumber daya ML. Namun, kami merekomendasikan tim ML menggunakan observabilitas terpusat dan pendekatan model tata kelola data untuk menyederhanakan tata kelola data dan manajemen audit.



Struktur akun model operasi terdesentralisasi

Struktur akun model federasi

Model ini mirip dengan model terpusat; namun, perbedaan utamanya adalah bahwa setiap tim ilmu data/ML mendapatkan kumpulan akun pengembangan/pengujian/beban kerja produksi mereka sendiri yang memungkinkan isolasi fisik yang kuat dari sumber daya ML mereka, dan juga memungkinkan setiap tim untuk menskalakan secara independen tanpa memengaruhi tim lain.



Struktur akun model operasi federasi

Multitenansi platform ML

Multitenancy adalah arsitektur perangkat lunak di mana satu instance perangkat lunak dapat melayani beberapa kelompok pengguna yang berbeda. Penyewa adalah sekelompok pengguna yang berbagi akses umum dengan hak istimewa khusus untuk instance perangkat lunak. Misalnya, jika Anda membangun beberapa produk ML, maka setiap tim produk dengan persyaratan akses serupa dapat dianggap sebagai penyewa atau tim.

Meskipun memungkinkan untuk mengimplementasikan beberapa tim dalam instance SageMaker Studio (seperti SageMakerDomain), pertimbangkan keuntungan tersebut terhadap trade-off seperti radius ledakan, atribusi biaya, dan batas level akun saat Anda membawa beberapa tim ke dalam satu

domain Studio. SageMaker Pelajari lebih lanjut tentang trade-off dan praktik terbaik tersebut di bagian berikut.

Jika Anda memerlukan isolasi sumber daya absolut, pertimbangkan untuk menerapkan domain SageMaker Studio untuk setiap penyewa di akun yang berbeda. Bergantung pada persyaratan isolasi Anda, Anda dapat menerapkan beberapa lini bisnis (LOB) sebagai beberapa domain dalam satu akun dan Wilayah. Gunakan ruang bersama untuk kolaborasi mendekati waktu nyata antara anggota tim/LOB yang sama. Dengan beberapa domain, Anda masih akan menggunakan kebijakan dan izin manajemen akses identitas (IAM) untuk memastikan isolasi sumber daya.

SageMaker sumber daya yang dibuat dari domain diberi tag otomatis dengan domain Amazon Resource Name (ARN) dan profil pengguna atau ruang ARN untuk isolasi sumber daya yang mudah. Untuk kebijakan sampel, lihat Dokumentasi isolasi sumber daya Domain. Di sana Anda dapat melihat referensi terperinci kapan harus menggunakan strategi multi-akun atau multi-domain, bersama dengan perbandingan fitur dalam dokumentasi, dan Anda dapat melihat skrip contoh untuk mengisi ulang tag untuk domain yang ada di repositori. GitHub

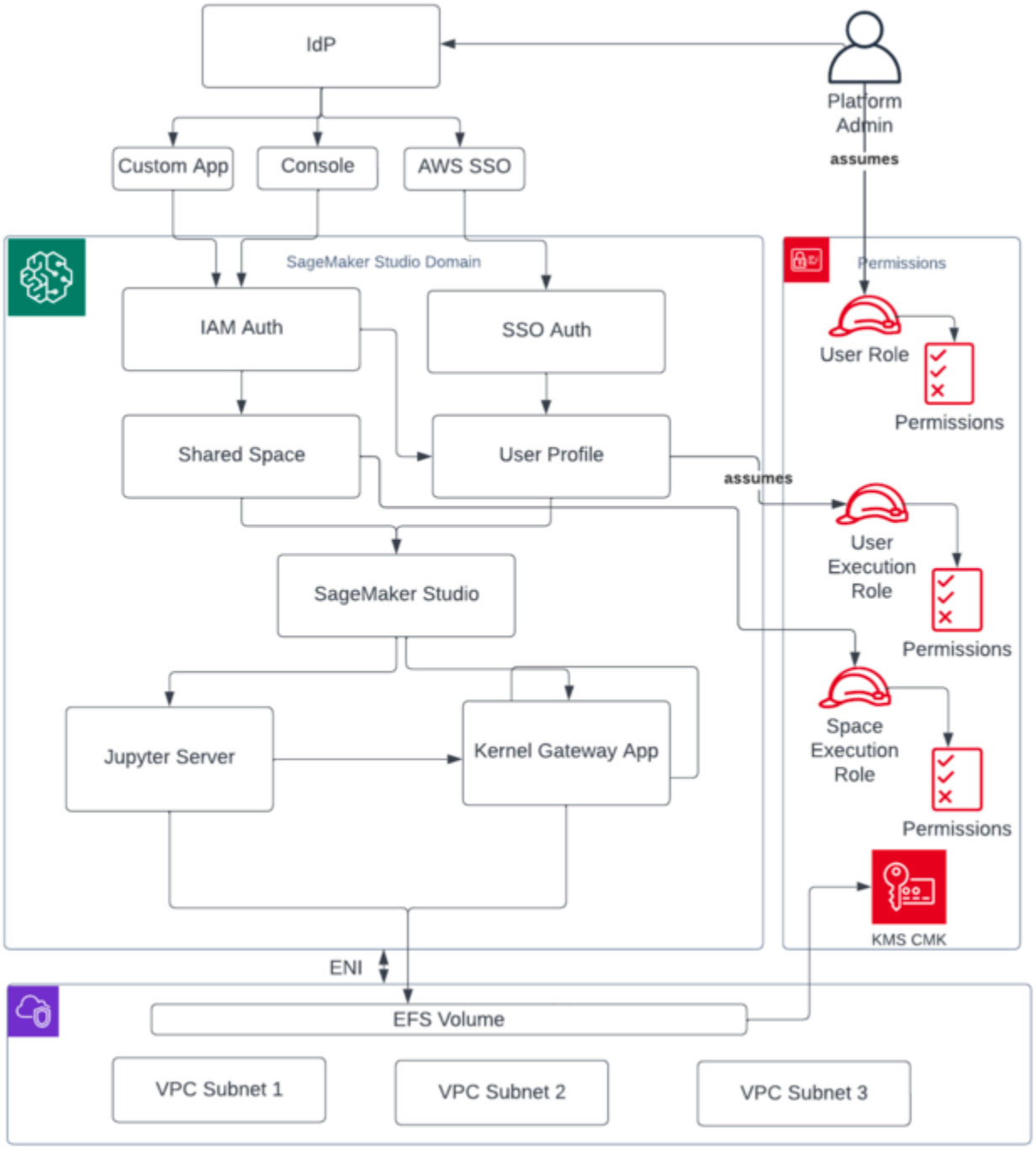
Terakhir, Anda dapat menerapkan penerapan layanan mandiri sumber daya SageMaker Studio ke beberapa akun menggunakan. AWS Service Catalog Untuk informasi selengkapnya, lihat Mengelola AWS Service Catalog produk dalam beberapa Akun AWS dan Wilayah AWS.

Pengelolaan domain

SageMaker Domain Amazon terdiri dari:

- Volume Amazon Elastic File System (Amazon EFS) terkait
- Daftar pengguna yang berwenang
- Berbagai konfigurasi keamanan, aplikasi, kebijakan, dan Amazon Virtual Private Cloud (Amazon VPC)

Diagram berikut memberikan tampilan tingkat tinggi dari berbagai komponen yang merupakan SageMakerStudio domain:

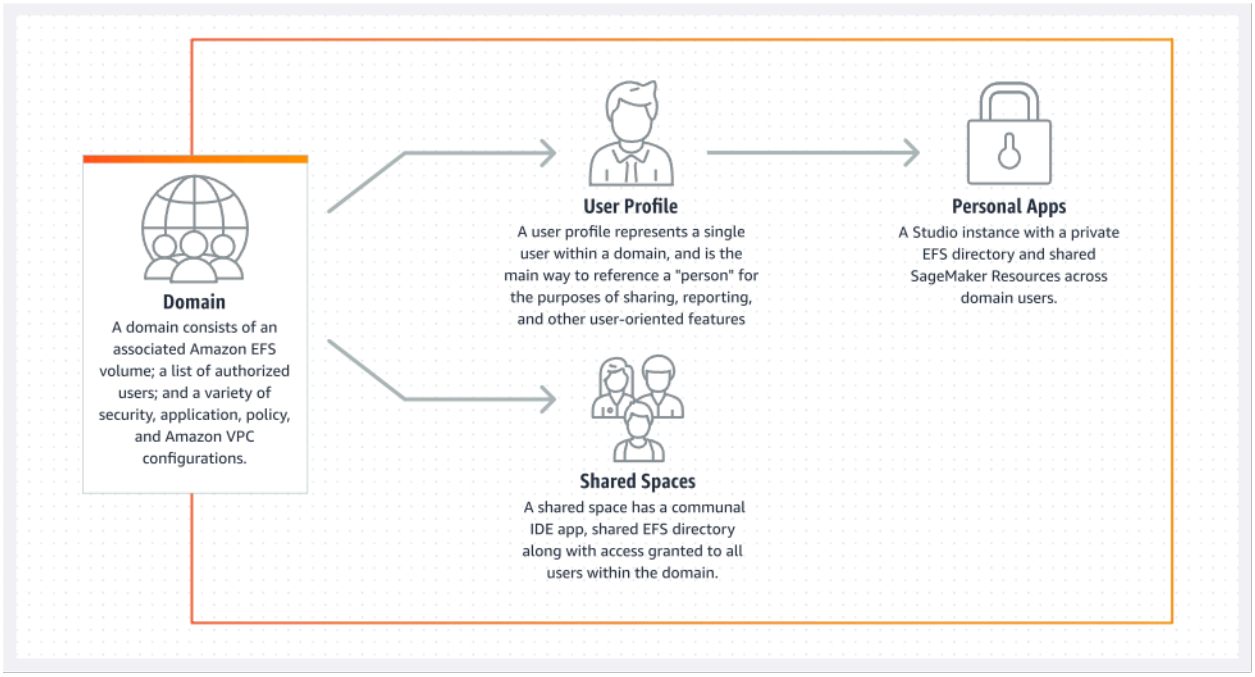


Tampilan tingkat tinggi dari berbagai komponen yang merupakan domain Studio SageMaker

Beberapa domain dan ruang bersama

Amazon SageMaker sekarang mendukung pembuatan beberapa SageMaker domain dalam satu Wilayah AWS untuk setiap akun. Setiap domain dapat memiliki pengaturan domain sendiri, seperti mode otentikasi, dan pengaturan jaringan, seperti VPC dan subnet. Profil pengguna tidak dapat dibagikan di seluruh domain. Jika pengguna manusia adalah bagian dari beberapa tim yang dipisahkan oleh domain, buat profil pengguna untuk pengguna di setiap domain. Lihat Ikhtisar Beberapa Domain untuk mempelajari tentang pengisian ulang tag untuk domain yang ada.

Setiap domain yang diatur dalam mode autentikasi IAM dapat memanfaatkan ruang bersama untuk kolaborasi mendekati waktu nyata antar pengguna. Dengan ruang bersama, pengguna mendapatkan akses ke direktori Amazon EFS bersama, dan JupyterServeraplikasi bersama untuk antarmuka pengguna, dan dapat mengedit bersama dalam waktu dekat. Penandaan otomatis sumber daya yang dibuat oleh ruang bersama memungkinkan administrator melacak biaya pada tingkat proyek. JupyterServer UI bersama juga memfilter sumber daya seperti eksperimen dan entri registri model sehingga hanya item yang relevan dengan upaya HTML bersama yang akan ditampilkan. Diagram berikut memberikan ikhtisar aplikasi pribadi dan ruang bersama dalam setiap domain.



Ikhtisar aplikasi pribadi dan ruang bersama dalam satu domain

Siapkan spasi bersama di domain Anda

Spasi bersama biasanya dibuat untuk usaha atau proyek ML tertentu di mana anggota dari satu domain memerlukan akses hampir real-time ke penyimpanan file dasar dan IDE yang sama. Pengguna dapat mengakses, membaca, mengedit, dan berbagi notebook mereka dalam waktu dekat, yang memberi mereka jalur tercepat untuk mulai iterasi dengan rekan-rekan mereka.

Untuk membuat ruang bersama, Anda harus terlebih dahulu menetapkan peran eksekusi default spasi yang akan mengatur izin untuk setiap pengguna yang menggunakan ruang tersebut. Pada saat penulisan ini, semua pengguna dalam domain akan memiliki akses ke semua ruang bersama di domain mereka. Lihat Membuat ruang bersama untuk dokumentasi terbaru tentang menambahkan spasi bersama ke domain yang ada.

Siapkan domain Anda untuk federasi IAM

Sebelum menyiapkan federasi AWS Identity and Access Management (IAM) untuk domain SageMaker Studio Anda, Anda perlu menyiapkan peran pengguna federasi IAM (seperti administrator platform) di iDP Anda, seperti yang dibahas di bagian Manajemen identitas.

Untuk petunjuk terperinci untuk menyiapkan SageMaker Studio dengan opsi IAM, lihat Onboard ke SageMaker Domain Amazon Menggunakan Pusat Identitas IAM.

Siapkan domain Anda untuk federasi single sign-on (SSO)

Untuk menggunakan federasi sistem masuk tunggal (SSO), Anda harus mengaktifkan akun AWS Organizationsmanajemen Anda AWS IAM Identity Center di Wilayah yang sama di mana Anda perlu menjalankan Studio. SageMaker Langkah-langkah pengaturan domain mirip dengan langkah federasi IAM, kecuali Anda memilih AWS IAM Identity Center(IDC) di bagian Autentikasi.

Untuk petunjuk terperinci, lihat Onboard ke SageMaker Domain Amazon Menggunakan Pusat Identitas IAM.

SageMaker Profil pengguna studio

Profil pengguna mewakili satu pengguna dalam domain, dan merupakan cara utama untuk mereferensikan “orang” untuk tujuan berbagi, melaporkan, dan fitur berorientasi pengguna lainnya. Entitas ini dibuat saat pengguna melakukan onboard toSageMaker Studio. Jika administrator mengundang seseorang melalui email atau mengimpornya dari IDC, profil pengguna akan dibuat

secara otomatis. Profil pengguna adalah pemegang utama pengaturan untuk pengguna individu, dan memiliki referensi ke direktori home Amazon Elastic File System (Amazon EFS) pribadi pengguna. Sebaiknya buat profil pengguna untuk setiap pengguna fisik aplikasi SageMaker Studio. Setiap pengguna memiliki direktori khusus mereka sendiri di Amazon EFS, dan profil pengguna tidak dapat dibagikan di seluruh domain di akun yang sama.

Setiap profil pengguna yang berbagi domain SageMaker Studio mendapatkan sumber daya komputasi khusus (seperti instans SageMaker Amazon Elastic Compute Cloud (Amazon EC2)) untuk menjalankan notebook. Instans komputasi yang dialokasikan untuk pengguna satu sepenuhnya terisolasi dari yang dialokasikan untuk pengguna dua. Demikian pula, sumber daya komputasi yang dialokasikan untuk pengguna dalam satu AWS akun benar-benar terpisah dari yang dialokasikan untuk pengguna di akun lain. Setiap pengguna dapat menjalankan hingga empat aplikasi (aplikasi) dalam wadah Docker yang terisolasi, atau gambar pada jenis instance yang sama.

Aplikasi Jupyter Server

Saat Anda meluncurkan buku catatan Amazon SageMaker Studio untuk pengguna dengan mengakses URL yang telah ditandatangani sebelumnya atau dengan masuk menggunakan AWS IAM iDC, aplikasi Jupyter Server diluncurkan di instance VPC yang dikelola layanan. SageMaker Setiap pengguna mendapatkan aplikasi Jupyter Server khusus mereka sendiri di aplikasi pribadi. Secara default, aplikasi Jupyter Server untuk notebook SageMaker Studio dijalankan pada `ml.t3.medium` instance khusus (dicadangkan sebagai jenis instance sistem). Komputasi untuk contoh ini tidak ditagih kepada pelanggan.

Aplikasi Jupyter Kernel Gateway

Aplikasi Kernel Gateway dapat dibuat melalui API atau antarmuka SageMaker Studio, dan berjalan pada jenis instance yang dipilih. Aplikasi ini dapat dijalankan menggunakan salah satu gambar SageMaker Studio bawaan yang telah dikonfigurasi sebelumnya dengan ilmu data populer, dan paket pembelajaran mendalam seperti, Apache MXNet TensorFlow, dan. PyTorch

Pengguna dapat memulai dan menjalankan beberapa kernel notebook Jupyter, sesi terminal, dan konsol interaktif dalam aplikasi SageMaker Studio Image/Kernel Gateway yang sama. Pengguna juga dapat menjalankan hingga empat aplikasi Kernel Gateway atau gambar pada instance fisik yang sama—masing-masing diisolasi oleh wadah/gambarnya.

Untuk membuat aplikasi tambahan, Anda perlu menggunakan jenis instance yang berbeda. Profil pengguna hanya dapat menjalankan satu instance, dari jenis instance apa pun. Misalnya, pengguna

dapat menjalankan notebook sederhana menggunakan gambar sains data bawaan SageMaker Studio, dan notebook lain menggunakan TensorFlow gambar bawaan, pada contoh yang sama. Pengguna ditagih untuk waktu instance berjalan. Untuk menghindari biaya saat pengguna tidak aktif menjalankan SageMaker Studio, pengguna perlu mematikan instance. Untuk informasi selengkapnya, lihat Matikan dan perbarui Aplikasi Studio.

Setiap kali Anda mematikan dan membuka kembali aplikasi Kernel Gateway dari antarmuka SageMaker Studio, aplikasi tersebut dimulai pada instance baru. Ini berarti bahwa instalasi paket tidak bertahan melalui restart aplikasi yang sama. Demikian pula, jika pengguna mengubah jenis instance pada notebook, paket yang diinstal dan variabel sesi mereka hilang. Namun, Anda dapat menggunakan fitur seperti membawa gambar dan skrip siklus hidup Anda sendiri untuk membawa paket pengguna sendiri ke SageMaker Studio dan mempertahankannya melalui sakelar instance dan peluncuran instance baru.

Volume Amazon Elastic File System

Saat domain dibuat, satu volume Amazon Elastic File System (Amazon EFS) dibuat untuk digunakan oleh semua pengguna dalam domain. Setiap profil pengguna menerima direktori home pribadi dalam volume Amazon EFS untuk menyimpan notebook, GitHub repositori, dan file data pengguna. Setiap ruang dalam domain menerima direktori pribadi dalam volume Amazon EFS yang dapat diakses oleh beberapa profil pengguna. Akses ke folder dipisahkan oleh pengguna, melalui izin sistem file. SageMaker Studio membuat ID pengguna unik global untuk setiap profil atau ruang pengguna, dan menerapkannya sebagai ID pengguna/grup Antarmuka Sistem Operasi Portabel (POSIX) untuk direktori home pengguna di EFS, yang mencegah pengguna/spasi lain mengakses datanya.

Pencadangan dan pemulihan

Volume EFS yang ada tidak dapat dilampirkan ke SageMaker domain baru. Dalam pengaturan produksi, pastikan volume Amazon EFS dicadangkan (ke volume EFS lain, atau ke Amazon Simple Storage Service (Amazon S3)). Jika volume EFS terhapus secara tidak sengaja, administrator harus meruntuhkan dan membuat ulang domain SageMaker Studio. Prosesnya adalah sebagai berikut:

Cadangkan daftar profil pengguna, spasi, dan ID pengguna EFS (UID) terkait melalui panggilan`ListUserProfiles`,`DescribeUserProfile`,`List Spaces`, dan `DescribeSpace` API.

1. Buat domain SageMaker Studio baru.

2. Buat profil dan spasi pengguna.
3. Untuk setiap profil pengguna, salin file dari cadangan di EFS/Amazon S3.
4. Secara opsional, hapus semua aplikasi dan profil pengguna, di domain SageMaker Studio lama.

Untuk petunjuk terperinci, lihat bagian lampiran Pencadangan dan SageMaker pemulihan domain Studio.

> **Note**
>
> Ini juga dapat dicapai melalui `LifecycleConfigurations` pencadangan data ke dan dari S3 setiap kali pengguna memulai aplikasi mereka.

Volume Amazon EBS

Volume penyimpanan Amazon Elastic Block Store (Amazon EBS) juga dilampirkan ke setiap instans SageMaker Studio Notebook. Ini digunakan sebagai volume root wadah atau gambar yang berjalan pada instance. Meskipun penyimpanan Amazon EFS tetap ada, volume Amazon EBS yang melekat pada kontainer bersifat sementara. Data yang disimpan secara lokal di Amazon EBS volume tidak akan bertahan jika pelanggan menghapus aplikasi.

Mengamankan akses ke URL yang telah ditandatangani sebelumnya

Saat pengguna SageMaker Studio membuka tautan buku catatan, SageMaker Studio memvalidasi kebijakan IAM pengguna federasi untuk mengotorisasi akses, serta membuat serta menyelesaikan URL yang telah ditandatangani sebelumnya untuk pengguna tersebut. Karena SageMaker konsol berjalan pada domain internet, URL yang dihasilkan dan ditandatangani sebelumnya ini terlihat di sesi browser. Ini menyajikan vektor ancaman yang tidak diinginkan untuk pencurian data dan mendapatkan akses ke data pelanggan ketika kontrol akses yang tepat tidak diberlakukan.

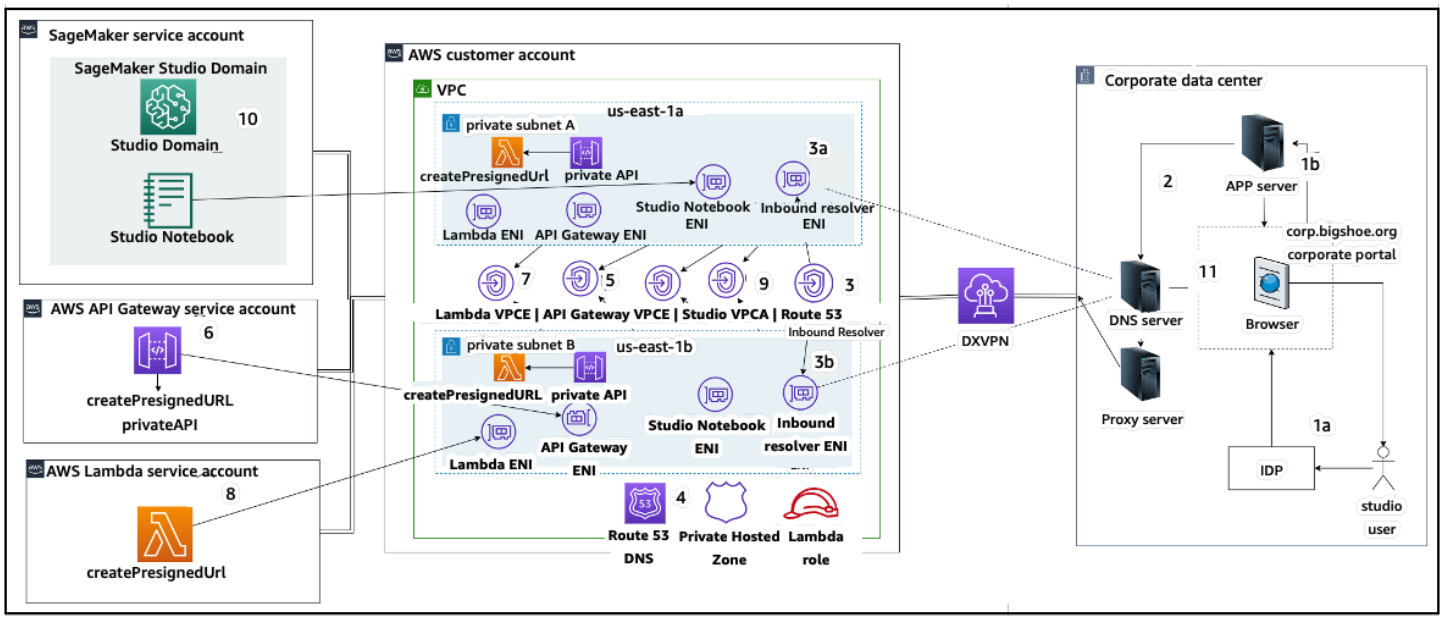
Studio mendukung beberapa metode untuk menegakkan kontrol akses terhadap pencurian data URL yang telah ditandatangani sebelumnya:

- Validasi IP klien menggunakan kondisi kebijakan IAM `aws:sourceIp`
- Validasi VPC klien menggunakan kondisi IAM `aws:sourceVpc`

- Validasi titik akhir VPC klien menggunakan kondisi kebijakan IAM `aws:sourceVpce`

Saat Anda mengakses notebook SageMaker Studio dari SageMaker konsol, satu-satunya opsi yang tersedia adalah menggunakan validasi IP klien dengan kondisi kebijakan IAM. `aws:sourceIp` Namun, Anda dapat menggunakan produk perutean lalu lintas browser seperti Zscaler untuk memastikan skala dan kepatuhan untuk akses internet tenaga kerja Anda. Produk perutean lalu lintas ini menghasilkan IP sumber mereka sendiri, yang rentang IP-nya tidak dikendalikan oleh pelanggan perusahaan. Hal ini membuat tidak mungkin bagi pelanggan perusahaan ini untuk menggunakan `aws:sourceIp` kondisi tersebut.

Untuk menggunakan validasi titik akhir VPC klien menggunakan kondisi kebijakan IAM`aws:sourceVpce`, pembuatan URL yang telah ditandatangani sebelumnya harus berasal dari VPC pelanggan yang sama SageMaker tempat Studio digunakan, dan resolusi URL yang telah ditandatangani sebelumnya perlu dilakukan melalui titik akhir VPC Studio di VPC pelanggan. SageMaker Resolusi URL yang telah ditandatangani sebelumnya selama waktu akses untuk pengguna jaringan perusahaan dapat diselesaikan dengan menggunakan aturan penerusan DNS (baik di Zscaler dan DNS perusahaan), dan kemudian ke titik akhir VPC pelanggan menggunakan resolver masuk Amazon Route 53 seperti yang ditunjukkan dalam arsitektur berikut:



Mengakses URL Studio yang telah ditandatangani sebelumnya dengan titik akhir VPC melalui jaringan perusahaan

Untuk step-by-step panduan menyiapkan arsitektur sebelumnya, lihat URL presigned Amazon SageMaker Studio Aman Bagian 1: Infrastruktur dasar.

SageMaker kuota dan batas domain

- SageMaker Federasi SSO domain studio hanya didukung di Wilayah, di seluruh akun anggota AWS organisasi tempat Pusat AWS Identitas disediakan.
- Spasi bersama saat ini tidak didukung dengan domain yang disiapkan dengan Pusat AWS Identitas.
- Konfigurasi VPC dan subnet tidak dapat diubah setelah membuat domain. Namun, Anda dapat membuat domain baru dengan konfigurasi VPC dan subnet yang berbeda.
- Akses domain tidak dapat beralih antara mode IAM dan SSO setelah membuat domain. Anda dapat membuat domain baru dengan mode otentikasi yang berbeda.
- Ada batas empat aplikasi gateway kernel per jenis instans yang diluncurkan untuk setiap pengguna.
- Setiap pengguna hanya dapat meluncurkan satu instance dari setiap jenis instance.
- Ada batasan sumber daya yang dikonsumsi dalam domain, seperti jumlah instance yang diluncurkan oleh jenis instans, dan jumlah profil pengguna yang dapat dibuat. Lihat halaman kuota layanan untuk daftar lengkap batas layanan.
- Pelanggan dapat mengirimkan kasus dukungan perusahaan dengan justifikasi bisnis untuk meningkatkan batas sumber daya default seperti jumlah domain atau profil pengguna, yang dikenakan pagar pembatas tingkat akun.
- Batas keras pada jumlah aplikasi bersamaan per akun adalah 2.500 aplikasi. Domain dan batas profil pengguna bergantung pada batas keras ini. Misalnya, sebuah akun dapat memiliki satu domain dengan 1.000 profil pengguna, atau 20 domain dengan 50 profil pengguna masing-masing.

Manajemen identitas

Bagian ini membahas bagaimana pengguna tenaga kerja di direktori perusahaan bergabung ke dalam Akun AWS dan mengakses Studio. SageMaker Pertama, kami akan menjelaskan secara singkat bagaimana pengguna, grup, dan peran dipetakan, dan cara kerja federasi pengguna.

Pengguna, grup, dan peran

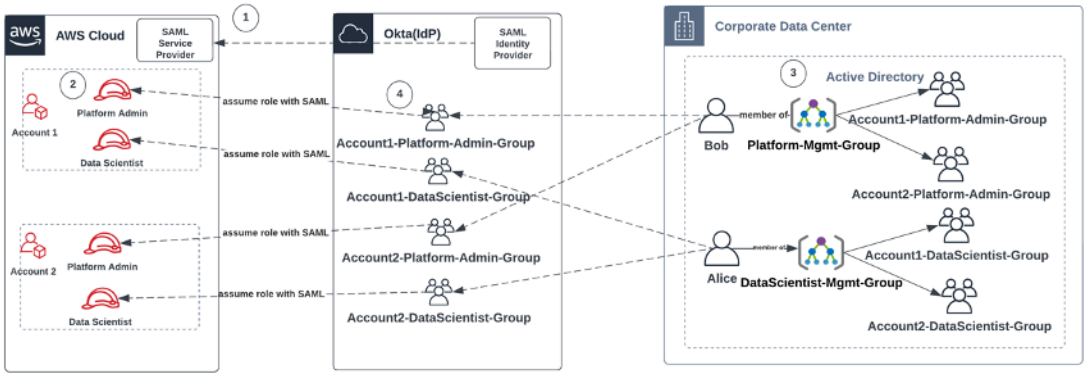
Di AWS, izin sumber daya dikelola menggunakan pengguna, grup, dan peran. Pelanggan dapat mengelola pengguna dan grup mereka baik melalui IAM, atau di direktori perusahaan seperti Active Directory (AD), yang diaktifkan melalui iDP eksternal seperti Okta, yang memungkinkan mereka untuk mengautentikasi pengguna ke berbagai aplikasi yang berjalan di cloud dan lokal.

Seperti yang dibahas di bagian Manajemen Identitas Pilar AWS Keamanan, ini adalah praktik terbaik untuk mengelola identitas pengguna Anda di IDP pusat, karena ini membantu mengintegrasikan dengan mudah dengan proses SDM back-end Anda, dan membantu mengelola akses ke pengguna tenaga kerja Anda.

IdPs seperti Okta memungkinkan pengguna akhir untuk mengautentikasi ke satu atau lebih Akun AWS dan mendapatkan akses ke peran tertentu menggunakan SSO dengan security assertation markup language (SAMP). Admin iDP memiliki kemampuan untuk mengunduh peran dari ke Akun AWS iDP, dan menetapkannya ke pengguna. Saat masuk ke AWS, pengguna akhir disajikan dengan AWS layar yang menampilkan AWS peran daftar yang ditetapkan untuk mereka dalam satu atau lebih Akun AWS. Mereka dapat memilih peran yang akan diambil untuk login, yang menentukan izin mereka selama sesi yang diautentikasi itu.

Grup harus ada di IDP untuk setiap akun tertentu dan kombinasi peran yang ingin Anda berikan aksesnya. Anda dapat menganggap kelompok-kelompok ini sebagai kelompok AWS khusus peran. Setiap pengguna yang merupakan anggota grup khusus peran ini diberikan hak tunggal: akses ke satu peran tertentu dalam satu peran tertentu Akun AWS. Namun, proses hak tunggal ini tidak menskalakan untuk mengelola akses pengguna dengan menetapkan setiap pengguna ke grup AWS peran tertentu. Untuk menyederhanakan administrasi, kami sarankan Anda juga membuat sejumlah grup untuk semua set pengguna yang berbeda di organisasi Anda yang memerlukan serangkaian hak yang berbeda. AWS

Untuk mengilustrasikan penyiapan iDP pusat, pertimbangkan perusahaan dengan penyiapan AD, tempat pengguna dan grup disinkronkan ke direktori iDP. Pada tahun AWS, grup AD ini dipetakan ke peran IAM. Langkah-langkah utama alur kerja berikut:



Alur kerja untuk pengguna AD orientasi, grup AD, dan peran IAM

1. Di AWS, Setup integrasi SAMP untuk masing-masing Anda Akun AWS dengan IDP Anda.
2. Di AWS, atur peran di masing-masing Akun AWS dan sinkronkan ke iDP.
3. Dalam sistem AD perusahaan:
   a. Buat Grup AD untuk setiap peran akun dan sinkronkan ke IDP (misalnya, `Account1-Platform-Admin-Group` (alias Grup AWS Peran)).
   b. Buat grup manajemen di setiap tingkat persona (misalnya,`Platform-Mgmt-Group`) dan tetapkan grup AWS peran sebagai anggota.
   c. Tetapkan pengguna ke grup manajemen tersebut untuk mengizinkan akses ke Akun AWS peran.
4. Di IDP, petakan grup AWS peran (sepertiAccount1-Platform-Admin-Group) ke Akun AWS peran (seperti Admin Platform di Akun1).
5. Ketika Ilmuwan Data Alice masuk ke Idp, mereka disajikan dengan UI Aplikasi AWS Federasi dengan dua opsi untuk dipilih: 'Ilmuwan Data Akun 1' dan 'Ilmuwan Data Akun 2'.
6. Alice memilih opsi 'Ilmuwan Data Akun 1', dan mereka terhubung ke aplikasi resmi mereka di AWS Akun 1 (Konsol). SageMaker

Untuk petunjuk terperinci tentang pengaturan federasi akun SAMP lihat Cara Mengkonfigurasi SAMP 2.0 Okta untuk AWS Federasi Akun.

Federasi pengguna

Otentikasi untuk SageMaker Studio dapat dilakukan dengan menggunakan IAM atau IAM iDC. Jika pengguna dikelola melalui IAM, mereka dapat memilih mode IAM. Jika perusahaan menggunakan

iDP eksternal, mereka dapat melakukan federasi melalui IAM atau IAM iDC. Perhatikan bahwa mode autentikasi tidak dapat diperbarui untuk domain SageMaker Studio yang ada, jadi sangat penting untuk membuat keputusan sebelum membuat domain SageMaker Studio produksi.

Jika SageMaker Studio diatur dalam mode IAM, pengguna SageMaker Studio mengakses aplikasi melalui URL yang telah ditandatangani sebelumnya yang secara otomatis menandatangani pengguna ke aplikasi SageMaker Studio saat diakses melalui browser.

Pengguna IAM

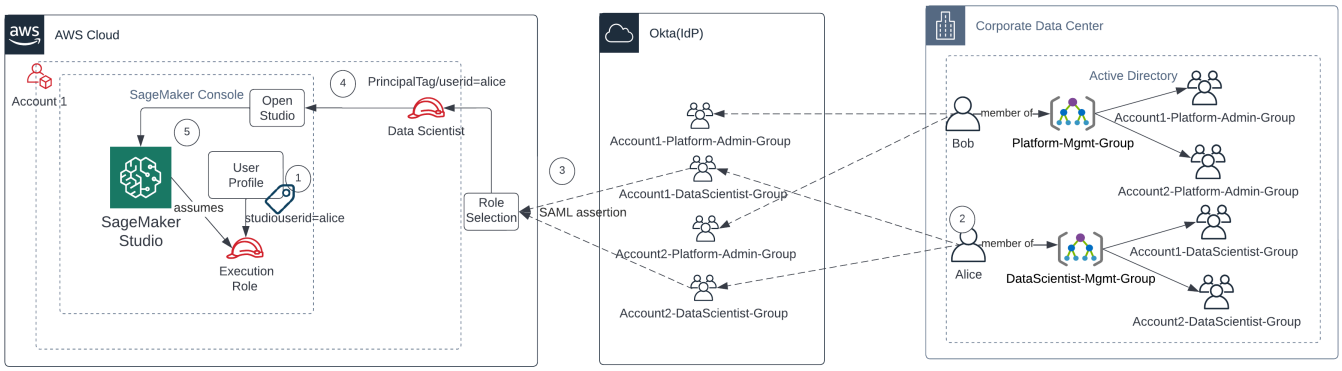
Untuk pengguna IAM, administrator membuat profil pengguna SageMaker Studio untuk setiap pengguna, dan mengaitkan profil pengguna dengan peran IAM yang memungkinkan tindakan yang diperlukan yang perlu dilakukan pengguna dari dalam Studio. Untuk membatasi AWS pengguna hanya mengakses profil pengguna SageMaker Studio mereka, administrator harus menandai profil pengguna SageMaker Studio dan melampirkan kebijakan IAM ke pengguna yang memungkinkan mereka mengakses hanya jika nilai tag sama dengan nama pengguna. AWS Pernyataan kebijakan terlihat seperti ini:

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

AWS IAM atau federasi akun

Metode Akun AWS federasi memungkinkan pelanggan untuk berfederasi ke SageMaker Konsol dari IDP SAMP mereka, seperti Okta. Untuk membatasi pengguna mengakses hanya profil

pengguna mereka, administrator harus menandai profil pengguna SageMaker Studio, menambahkan `PrincipalTags` pada iDP, dan mengaturnya sebagai tag transitif. Diagram berikut menggambarkan bagaimana pengguna federasi (Data Scientist Alice) berwenang untuk mengakses profil pengguna SageMaker Studio mereka sendiri.



Mengakses SageMaker Studio dalam mode federasi IAM

1. Profil pengguna Alice SageMaker Studio ditandai dengan ID pengguna mereka, dan terkait dengan peran eksekusi.
2. Alice mengautentikasi ke iDP (Okta).
3. IDP mengotentikasi Alice dan memposting pernyataan SAMP dengan dua peran (Data Scientist untuk akun 1 dan 2) Alice adalah anggota. Alice memilih peran Data Scientist untuk akun 1.
4. Alice masuk ke SageMaker Konsol Akun 1, dengan peran yang diasumsikan sebagai Data Scientist. Alice membuka instance aplikasi Studio mereka dari daftar instance aplikasi studio.
5. Tag utama Alice dalam sesi peran yang diasumsikan divalidasi terhadap tag profil pengguna instance aplikasi SageMaker Studio yang dipilih. Jika tag profil valid, instance aplikasi SageMaker Studio akan diluncurkan, dengan asumsi peran eksekusi.

Jika Anda ingin mengotomatiskan pembuatan peran dan kebijakan SageMaker Eksekusi sebagai bagian dari orientasi pengguna, berikut ini adalah salah satu cara untuk melakukannya:

1. Siapkan grup AD seperti `SageMaker-Account1-Group` di setiap akun dan tingkat Domain Studio.
2. Tambahkan SageMaker -Account1-Group ke keanggotaan grup pengguna saat Anda perlu melakukan onboard pengguna ke Studio. SageMaker

Siapkan proses otomatisasi yang mendengarkan acara `SageMaker-Account1-Group` keanggotaan, dan gunakan AWS API untuk membuat peran, kebijakan, tag, dan profil pengguna SageMaker Studio berdasarkan keanggotaan grup AD mereka. Lampirkan peran ke profil pengguna. Untuk kebijakan sampel, lihatMencegah pengguna SageMaker Studio mengakses profil pengguna lain.

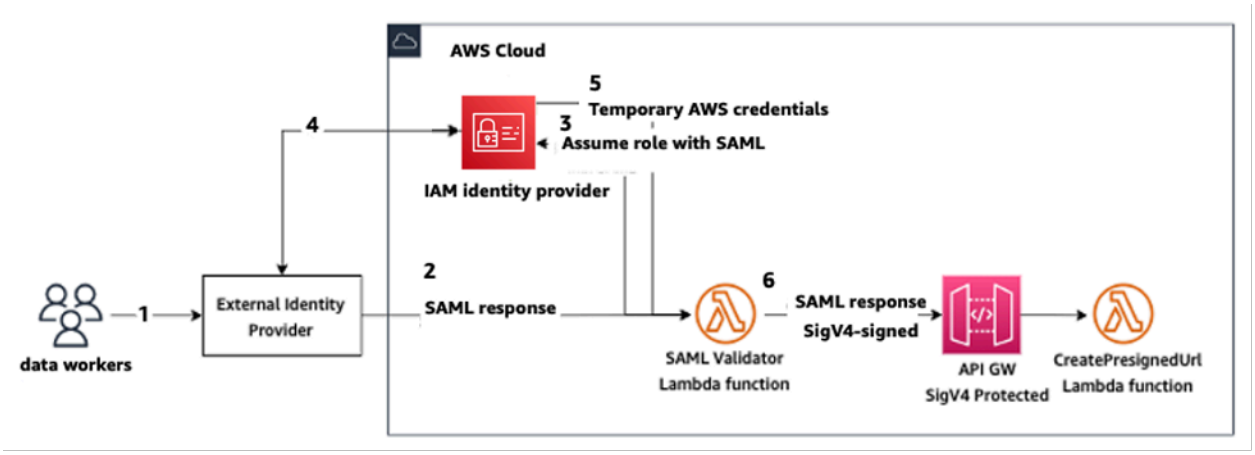
Otentikasi SAMP menggunakan AWS Lambda

Dalam mode IAM, pengguna juga dapat diautentikasi ke SageMaker Studio menggunakan pernyataan SAMP. Dalam arsitektur ini, pelanggan memiliki IDP yang sudah ada, di mana mereka dapat membuat aplikasi SAMP bagi pengguna untuk mengakses Studio (bukan aplikasi Federasi AWS Identitas). IDP pelanggan ditambahkan ke IAM. AWS Lambda Fungsi membantu memvalidasi pernyataan SAMP menggunakan IAM dan STS, dan kemudian memanggil gateway API atau fungsi Lambda secara langsung, untuk membuat URL domain yang telah ditandatangani sebelumnya.

Keuntungan dari solusi ini adalah bahwa fungsi Lambda dapat menyesuaikan logika untuk akses ke SageMaker Studio. Sebagai contoh:

- Secara otomatis membuat profil pengguna jika tidak ada.
- Lampirkan atau hapus peran atau dokumen kebijakan ke peran eksekusi SageMaker Studio dengan mengurai atribut SAMP.
- Sesuaikan profil pengguna dengan menambahkan Life Cycle Configuration (LCC) dan menambahkan tag.

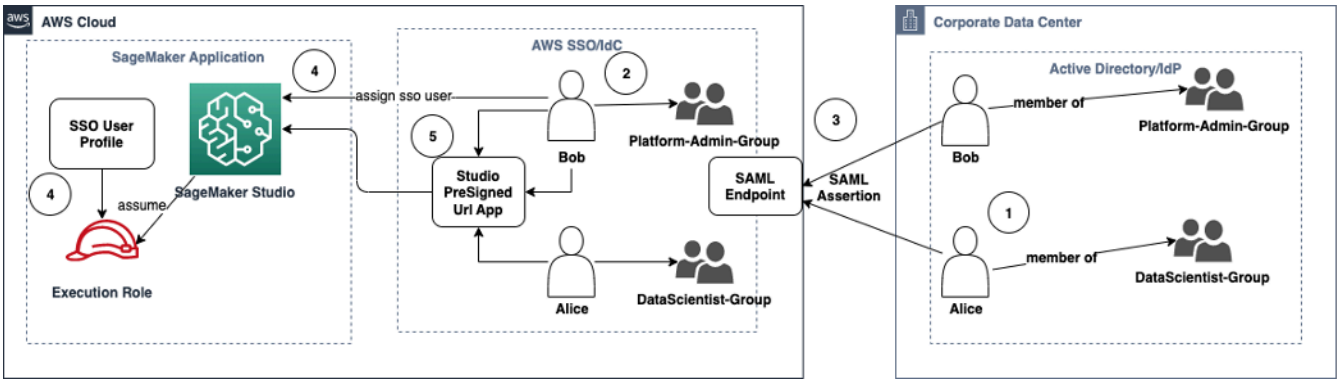
Singkatnya, solusi ini akan mengekspos SageMaker Studio sebagai aplikasi SAML2.0 dengan logika khusus untuk otentikasi dan otorisasi. Lihat bagian lampiran Akses SageMaker studio menggunakan pernyataan SAMP untuk detail implementasi.



Mengakses SageMaker Studio menggunakan aplikasi SAMP khusus

Federasi AWS IAM iDC

Metode federasi IDC memungkinkan pelanggan untuk federasi langsung ke aplikasi SageMaker Studio dari SAMP iDP mereka (seperti Okta). Diagram berikut menggambarkan bagaimana pengguna federasi diberi wewenang untuk mengakses instance SageMaker Studio mereka sendiri.



Mengakses SageMaker Studio dalam mode IAM iDC

1. Dalam iklan perusahaan, pengguna adalah anggota grup AD seperti grup Admin Platform dan grup Data Scientist.
2. Pengguna AD dan grup AD dari Identity Provider (iDP) disinkronkan ke AWS IAM Identity Center dan tersedia sebagai pengguna dan grup masuk tunggal untuk penetapan masing-masing.
3. IdP memposting pernyataan SAMP ke titik akhir IDC SAMP. AWS
4. Di SageMaker Studio, pengguna IDC ditugaskan ke aplikasi SageMaker Studio. Penugasan ini dapat dilakukan dengan menggunakan IDC Group dan SageMaker Studio akan berlaku di setiap tingkat pengguna IDC. Saat tugas ini dibuat, SageMaker Studio membuat profil pengguna IDC dan melampirkan peran eksekusi domain.
5. Pengguna mengakses Aplikasi SageMaker Studio menggunakan URL presigned aman yang dihosting sebagai aplikasi cloud dari IDC. SageMaker Studio mengasumsikan peran eksekusi yang dilampirkan ke profil pengguna IDC mereka.

Panduan otentikasi domain

Berikut adalah beberapa pertimbangan saat memilih mode otentikasi domain:

1. Jika Anda ingin pengguna tidak mengakses AWS Management Console dan melihat UI SageMaker Studio secara langsung, gunakan mode masuk tunggal dengan AWS IAM iDC.
2. Jika Anda ingin pengguna tidak mengakses AWS Management Console dan melihat UI SageMaker Studio secara langsung dalam mode IAM, Anda dapat melakukannya dengan menggunakan fungsi Lambda di backend untuk menghasilkan URL yang telah ditetapkan sebelumnya untuk profil pengguna dan mengarahkannya ke UI Studio. SageMaker
3. Dalam mode IDC, setiap pengguna dipetakan ke satu profil pengguna.
4. Semua profil pengguna secara otomatis diberi peran eksekusi default dalam mode IDC. Jika Anda ingin pengguna Anda diberi peran eksekusi yang berbeda, Anda perlu memperbarui profil pengguna menggunakan UpdateUserProfileAPI.
5. Jika Anda ingin membatasi akses UI SageMaker Studio dalam mode IAM (menggunakan URL presigned yang dihasilkan) ke titik akhir VPC, tanpa melintasi internet, Anda dapat menggunakan resolver DNS khusus. Lihat URL presigned Amazon SageMaker Studio Aman Bagian 1: Posting blog infrastruktur dasar.

Manajemen izin

Bagian ini membahas praktik terbaik untuk menyiapkan peran, kebijakan, dan pagar pembatas IAM yang umum digunakan untuk menyediakan dan mengoperasikan domain Studio. SageMaker

Peran dan kebijakan IAM

Sebagai praktik terbaik, Anda mungkin ingin terlebih dahulu mengidentifikasi orang dan aplikasi yang relevan, yang dikenal sebagai prinsipal yang terlibat dalam siklus hidup ML, dan AWS izin apa yang perlu Anda berikan kepada mereka. Seperti SageMaker layanan terkelola, Anda juga perlu mempertimbangkan prinsip layanan yang merupakan AWS layanan yang dapat melakukan panggilan API atas nama pengguna. Diagram berikut menggambarkan berbagai peran IAM yang mungkin ingin Anda buat, sesuai dengan persona yang berbeda dalam organisasi.

SageMaker Peran IAM

Peran ini dijelaskan secara rinci, bersama dengan beberapa contoh IAMPermissions spesifik yang akan mereka butuhkan.

- Peran pengguna Admin ML — Ini adalah prinsipal yang menyediakan lingkungan bagi ilmuwan data dengan membuat domain studio dan profil pengguna (`sagemaker:CreateDomain`,`sagemaker:CreateUserProfile`), membuat AWS Key Management Service (AWS KMS) kunci untuk pengguna, membuat bucket S3 untuk ilmuwan data, dan membuat repositori Amazon ECR untuk menampung wadah. Mereka juga dapat mengatur konfigurasi default dan skrip siklus hidup untuk pengguna, membangun dan melampirkan gambar kustom ke domain SageMaker Studio, dan menyediakan produk Service Catalog seperti proyek kustom, template EMR Amazon.

  Karena kepala sekolah ini tidak akan menjalankan pekerjaan pelatihan, misalnya, mereka tidak memerlukan izin untuk meluncurkan SageMaker pelatihan atau pekerjaan pemrosesan. Jika mereka menggunakan infrastruktur sebagai templat kode, seperti CloudFormation atau Terraform,

untuk menyediakan domain dan pengguna, peran ini akan diasumsikan oleh layanan penyediaan untuk membuat sumber daya atas nama admin. Peran ini mungkin memiliki akses hanya-baca untuk SageMaker menggunakan. AWS Management Console

Peran pengguna ini juga akan memerlukan izin EC2 tertentu untuk meluncurkan domain di dalam VPC pribadi, izin KMS untuk mengenkripsi volume EFS, serta izin untuk membuat peran terkait layanan untuk Studio (). `iam:CreateServiceLinkedRole` Kami akan menjelaskan izin granular tersebut nanti dalam dokumen.

- Peran pengguna Data Scientist - Prinsip ini adalah pengguna yang masuk ke SageMaker Studio, menjelajahi data, membuat pekerjaan dan saluran pipa pemrosesan dan pelatihan, dan sebagainya. Izin utama yang dibutuhkan pengguna adalah izin untuk meluncurkan SageMaker Studio, dan kebijakan lainnya dapat dikelola oleh peran layanan SageMaker eksekusi.
- SageMaker peran layanan eksekusi — Karena SageMaker merupakan layanan terkelola, ia meluncurkan pekerjaan atas nama pengguna. Peran ini sering kali paling luas dalam hal izin yang diizinkan, karena banyak pelanggan memilih untuk menggunakan peran eksekusi tunggal untuk menjalankan pekerjaan pelatihan, pekerjaan pemrosesan, atau model pekerjaan hosting. Meskipun ini adalah cara mudah untuk memulai, karena pelanggan matang dalam perjalanan mereka, mereka sering membagi peran eksekusi notebook menjadi peran terpisah untuk tindakan API yang berbeda, terutama saat menjalankan pekerjaan tersebut di lingkungan yang diterapkan.

  Anda mengaitkan peran dengan domain SageMaker Studio saat pembuatan. Namun, karena pelanggan mungkin memerlukan fleksibilitas untuk memiliki peran berbeda yang terkait dengan profil pengguna yang berbeda di domain (misalnya, berdasarkan fungsi pekerjaan mereka), Anda juga dapat mengaitkan peran IAM terpisah dengan setiap profil pengguna. Kami menyarankan Anda memetakan satu pengguna fisik ke satu profil pengguna. Jika Anda tidak melampirkan peran ke profil pengguna saat pembuatan, perilaku default adalah mengaitkan peran eksekusi SageMakerStudio domain dengan profil pengguna juga.

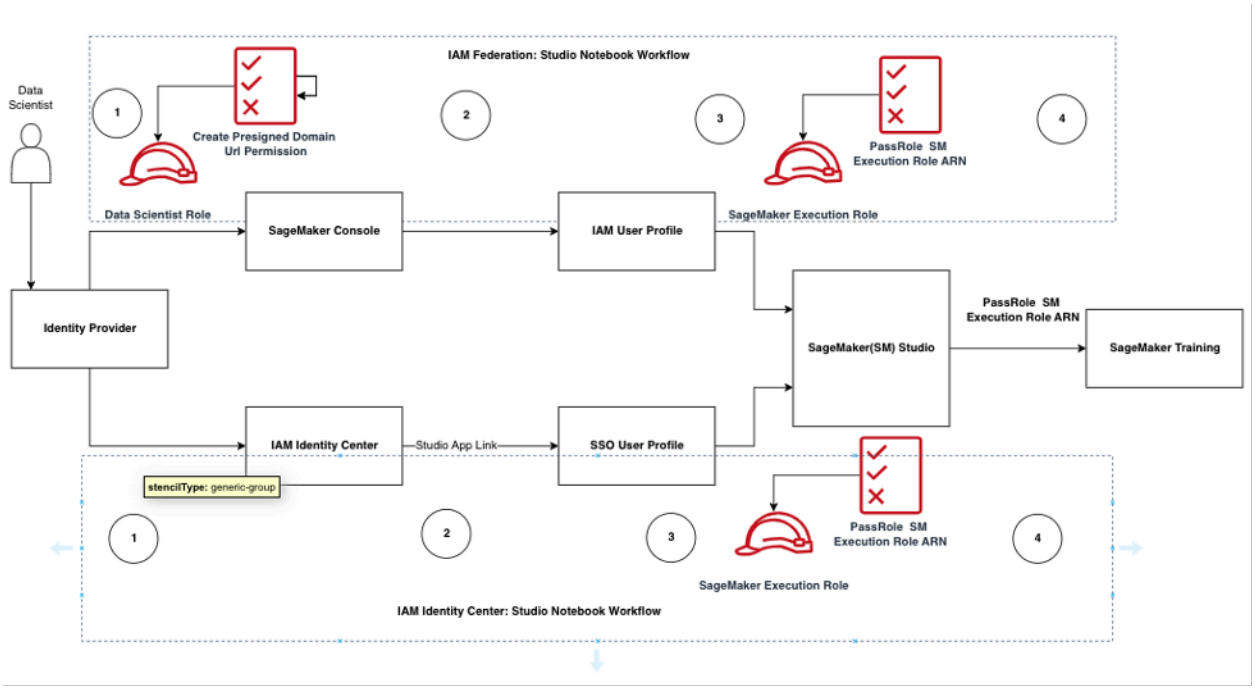
  Dalam kasus di mana beberapa ilmuwan data dan teknisi ML bekerja sama dalam sebuah proyek dan memerlukan model izin bersama untuk mengakses sumber daya, kami sarankan Anda membuat peran eksekusi SageMaker layanan tingkat tim untuk berbagi izin IAM di seluruh anggota tim Anda. Dalam kasus di mana Anda perlu mengunci izin di setiap tingkat pengguna, Anda dapat membuat peran eksekusi SageMaker layanan tingkat pengguna individu; Namun, Anda harus memperhatikan batas layanan Anda.

SageMaker Alur kerja otorisasi Notebook Studio

Bagian ini, membahas cara kerja otorisasi SageMaker Studio Notebook untuk berbagai aktivitas yang perlu dilakukan Data Scientist untuk membangun dan melatih model langsung dari SageMaker Studio Notebook. SageMaker Domain mendukung dua mode otorisasi:

- Federasi IAM
- Pusat Identitas IAM

Selanjutnya, paper ini memandu Anda melalui alur kerja otorisasi Data Scientist untuk masing-masing mode tersebut.



Alur kerja otentikasi dan otorisasi untuk pengguna Studio

Federasi IAM: Alur kerja Notebook SageMaker Studio

1. Seorang Ilmuwan Data mengautentikasi ke penyedia identitas perusahaan mereka dan mengasumsikan peran pengguna Data Scientist (peran federasi pengguna) di SageMaker konsol. Peran federasi ini memiliki izin `iam:PassRole` API pada peran SageMaker eksekusi untuk meneruskan peran Amazon Resource Name (ARN) ke SageMaker Studio.
2. Data Scientist memilih link Open Studio dari profil pengguna Studio IAM mereka yang terkait dengan peran eksekusi SageMaker

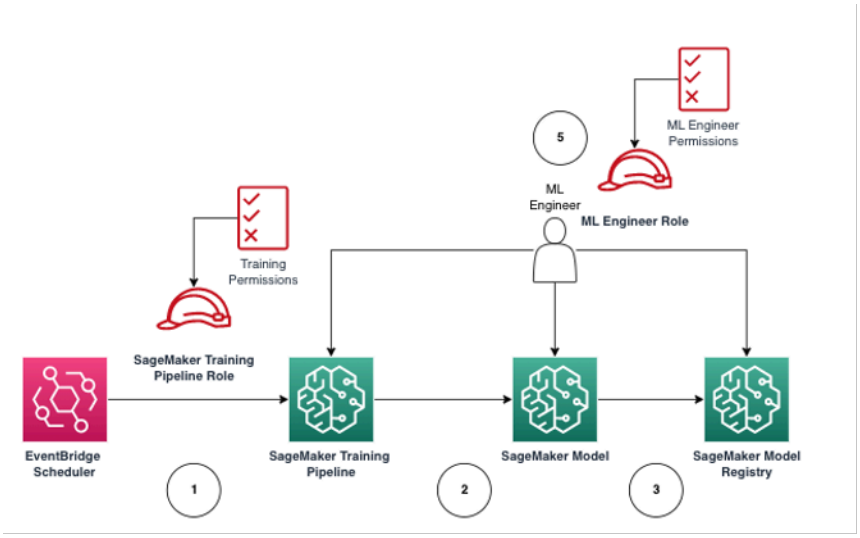
3. Layanan SageMaker Studio IDE diluncurkan, dengan asumsi izin peran SageMaker eksekusi profil pengguna. Peran ini memiliki izin `iam:PassRole` API pada peran SageMaker eksekusi untuk meneruskan peran ARN ke layanan SageMaker pelatihan.
4. Saat Data Scientist meluncurkan pekerjaan pelatihan di node komputasi jarak jauh, peran SageMaker eksekusi ARN diteruskan ke layanan pelatihan. SageMaker Ini menciptakan sesi peran baru dengan ARN ini dan menjalankan pekerjaan pelatihan. Jika Anda perlu mencatat izin lebih lanjut untuk pekerjaan pelatihan, Anda dapat membuat peran khusus pelatihan dan meneruskan peran tersebut ARN saat memanggil API pelatihan.

Pusat Identitas IAM: Alur kerja Notebook SageMaker Studio

1. Ilmuwan Data mengautentikasi ke penyedia identitas perusahaan mereka dan mengklik Pusat Identitas AWS IAM. Ilmuwan Data disajikan dengan Portal Pusat Identitas untuk pengguna.
2. Data Scientist mengeklik tautan SageMaker Studio App yang dibuat dari profil pengguna IDC mereka, yang terkait dengan peran SageMaker eksekusi.
3. Layanan SageMaker Studio IDE diluncurkan, dengan asumsi izin peran SageMaker eksekusi profil pengguna. Peran ini memiliki izin `iam:PassRole` API pada peran SageMaker eksekusi untuk meneruskan peran ARN ke layanan SageMaker pelatihan.
4. Saat Data Scientist meluncurkan pekerjaan pelatihan di node komputasi jarak jauh, peran SageMaker eksekusi ARN diteruskan ke layanan pelatihan. SageMaker Peran eksekusi ARN menciptakan sesi peran baru dengan ARN ini, dan menjalankan pekerjaan pelatihan. Jika Anda perlu mengurangi izin lebih lanjut untuk pekerjaan pelatihan, Anda dapat membuat peran khusus pelatihan dan meneruskan peran tersebut ARN saat memanggil API pelatihan.

Lingkungan yang digunakan: alur kerja SageMaker pelatihan

Di lingkungan yang digunakan seperti pengujian dan produksi sistem, pekerjaan dijalankan melalui penjadwal otomatis dan pemicu peristiwa, dan akses manusia ke lingkungan tersebut dibatasi dari SageMaker Studio Notebook. Bagian ini membahas bagaimana peran IAM bekerja dengan jalur SageMaker pelatihan di lingkungan yang digunakan.



SageMaker alur kerja pelatihan dalam lingkungan produksi yang dikelola

1. EventBridgePenjadwal Amazon memicu pekerjaan pipa SageMaker pelatihan.
2. Pekerjaan pipa SageMaker pelatihan mengasumsikan peran pipa SageMaker pelatihan untuk melatih model.
3. SageMaker Model terlatih terdaftar ke dalam SageMaker Model Registry.
4. Seorang insinyur ML mengasumsikan peran pengguna insinyur ML untuk mengelola pipa dan SageMaker model pelatihan.

Izin data

Kemampuan bagi pengguna SageMaker Studio untuk mengakses sumber data apa pun diatur oleh izin yang terkait dengan peran eksekusi SageMaker IAM mereka. Kebijakan yang dilampirkan dapat mengizinkan mereka untuk membaca, menulis, atau menghapus dari bucket atau awalan Amazon S3 tertentu, dan terhubung ke database Amazon RDS.

Mengakses data AWS Lake Formation

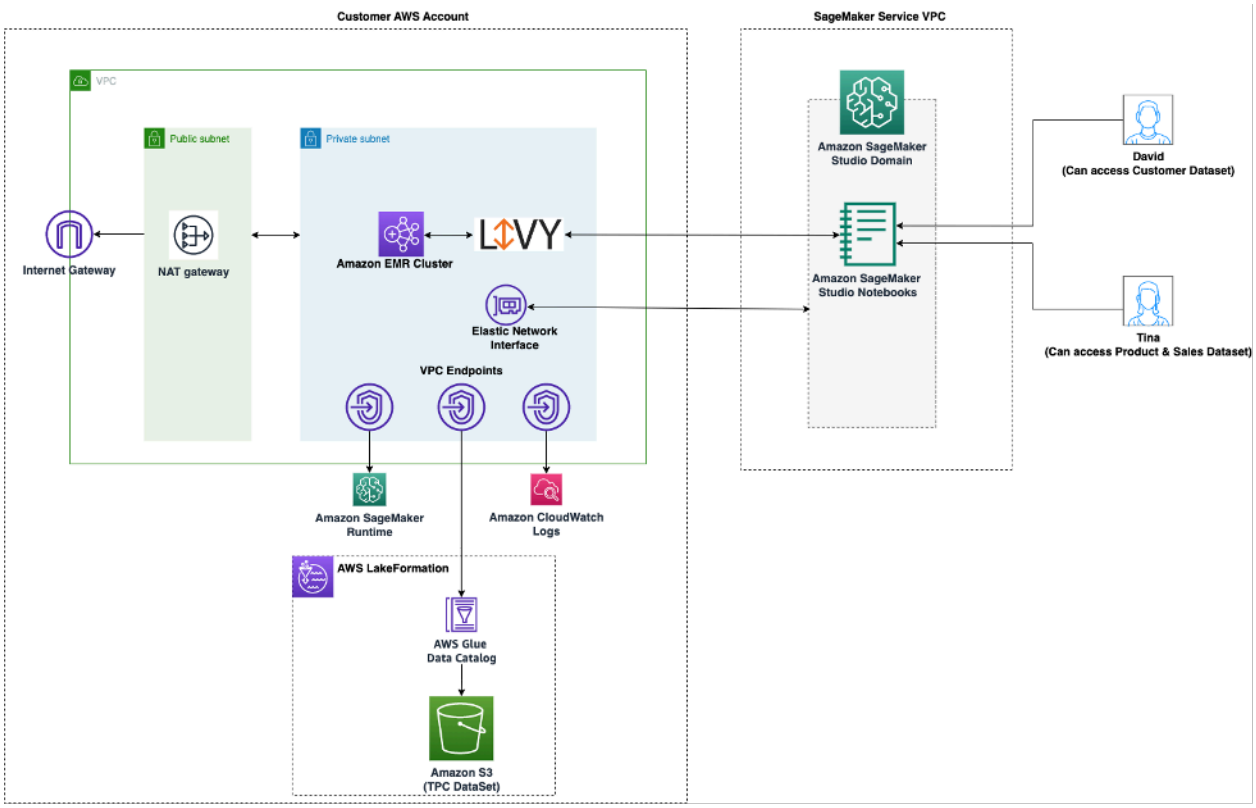
Banyak perusahaan telah mulai menggunakan data lake yang diatur oleh AWS Lake Formationuntuk memungkinkan akses data berbutir halus bagi pengguna mereka. Sebagai contoh data yang diatur tersebut, administrator dapat menutupi kolom sensitif untuk beberapa pengguna sambil tetap mengaktifkan kueri dari tabel dasar yang sama.

Untuk memanfaatkan Lake Formation dari SageMaker Studio, administrator dapat mendaftarkan peran eksekusi SageMaker IAM sebagai. `DataLakePrincipals` Untuk informasi selengkapnya, lihat Referensi Izin Lake Formation. Setelah diotorisasi, ada tiga metode utama untuk mengakses dan menulis data yang diatur dari SageMaker Studio:

1. Dari SageMaker Studio Notebook, pengguna dapat menggunakan mesin kueri seperti Amazon Athena atau pustaka yang dibangun di atas boto3 untuk menarik data langsung ke notebook. AWS SDK for Pandas (sebelumnya dikenal sebagai awswrangler) adalah perpustakaan yang populer. Berikut ini adalah contoh kode untuk menunjukkan betapa mulusnya hal ini:

```
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. Gunakan konektivitas asli SageMaker Studio ke Amazon EMR untuk membaca dan menulis data dalam skala besar. Melalui penggunaan peran runtime Apache Livy dan Amazon EMR, SageMaker Studio telah membangun konektivitas asli yang memungkinkan Anda meneruskan peran IAM SageMaker eksekusi (atau peran resmi lainnya) ke klaster EMR Amazon untuk akses dan pemrosesan data. Lihat Connect ke Amazon EMR Cluster dari Studio untuk up-to-date petunjuk.



Arsitektur untuk mengakses data yang dikelola oleh Lake Formation dari Studio SageMaker

3. Gunakan konektivitas asli SageMaker Studio ke sesi AWS Glue interaktif untuk membaca dan menulis data dalam skala besar. SageMaker Studio Notebook memiliki kernel bawaan yang memungkinkan pengguna menjalankan perintah secara interaktif. AWS Glue Ini memungkinkan penggunaan backend Python, Spark, atau Ray yang dapat diskalakan yang dapat membaca dan menulis data dengan mulus dalam skala besar dari sumber data yang diatur. Kernel memungkinkan pengguna untuk lulus SageMaker eksekusi mereka atau peran IAM resmi lainnya. Lihat Siapkan Data menggunakan Sesi AWS Glue Interaktif untuk informasi lebih lanjut.

Pagar pembatas umum

Bagian ini membahas pagar pembatas yang paling umum digunakan untuk menerapkan tata kelola pada sumber daya ML Anda menggunakan kebijakan IAM, kebijakan sumber daya, kebijakan titik akhir VPC, dan kebijakan kontrol layanan (SCP).

Batasi akses notebook ke instance tertentu

Kebijakan kontrol layanan ini dapat digunakan untuk membatasi tipe instans yang dapat diakses oleh ilmuwan data, saat membuat notebook Studio. Perhatikan bahwa setiap pengguna akan memerlukan

instance “sistem” yang diizinkan untuk membuat aplikasi Jupyter Server default yang meng-host 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"
                     ]
                 }
             }
         }
     ]
 }
```

Batasi domain Studio yang tidak sesuai SageMaker

Untuk domain SageMaker Studio, kebijakan kontrol layanan berikut dapat digunakan untuk menegakkan lalu lintas untuk mengakses sumber daya pelanggan sehingga mereka tidak melalui internet publik, melainkan melalui VPC pelanggan:

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

Batasi peluncuran gambar yang tidak sah SageMaker

Kebijakan berikut mencegah pengguna meluncurkan SageMaker gambar yang tidak sah di dalam domain mereka:f

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

Luncurkan notebook hanya melalui titik akhir SageMaker VPC

Selain titik akhir VPC untuk bidang SageMaker kontrol, mendukung titik akhir SageMaker VPC bagi pengguna untuk terhubung ke SageMaker notebook Studio atau instance notebook. SageMaker Jika Anda telah menyiapkan titik akhir VPC untuk instance SageMaker Studio/Notebook, kunci kondisi IAM berikut hanya akan mengizinkan koneksi ke notebook SageMaker Studio jika dibuat melalui titik akhir SageMaker VPC Studio atau melalui titik akhir 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"
                     ]
                 }
             }
         }
     ]
 }
```

Batasi akses notebook SageMaker Studio ke rentang IP terbatas

Perusahaan akan sering membatasi akses SageMaker Studio ke rentang IP perusahaan tertentu yang diizinkan. Kebijakan IAM berikut dengan kunci `SourceIP` kondisi dapat membatasi ini.

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

Mencegah pengguna SageMaker Studio mengakses profil pengguna lain

Sebagai administrator, saat Anda membuat profil pengguna, pastikan profil tersebut ditandai dengan nama pengguna SageMaker Studio dengan kunci `studiouserid` tag. Prinsipal (pengguna atau peran yang dilampirkan ke pengguna) juga harus memiliki tag dengan kunci `studiouserid` (tag ini dapat diberi nama apa saja, dan tidak terbatas pada`studiouserid`).

Selanjutnya, lampirkan kebijakan berikut ke peran yang akan diasumsikan pengguna saat meluncurkan SageMaker Studio.

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

```
            }
        }
    ]
}
```

Menegakkan penandaan

Ilmuwan data perlu menggunakan notebook SageMaker Studio untuk mengeksplorasi data, dan membangun serta melatih model. Menerapkan tag ke notebook membantu memantau penggunaan dan pengendalian biaya, serta memastikan kepemilikan dan auditabilitas.

Untuk aplikasi SageMaker Studio, pastikan profil pengguna diberi tag. Tag secara otomatis disebarkan ke aplikasi dari profil pengguna. Untuk menerapkan pembuatan profil pengguna dengan tag (didukung melalui CLI dan SDK), pertimbangkan untuk menambahkan kebijakan ini ke peran admin:

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

Untuk sumber daya lain, seperti pekerjaan pelatihan dan pekerjaan pemrosesan, Anda dapat membuat tag wajib menggunakan kebijakan berikut:

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

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

Akses root di SageMaker Studio

Di SageMaker Studio, notebook berjalan dalam wadah Docker yang, secara default, tidak memiliki akses root ke instance host. Demikian pula, selain pengguna run-as default, semua rentang ID pengguna lain di dalam wadah dipetakan ulang sebagai User-ID yang tidak memiliki hak istimewa pada instance host itu sendiri. Akibatnya, ancaman eskalasi hak istimewa terbatas pada wadah notebook itu sendiri.

Saat membuat gambar kustom, Anda mungkin ingin memberi pengguna izin non-root untuk kontrol yang lebih ketat; misalnya, menghindari menjalankan proses yang tidak diinginkan sebagai root, atau menginstal paket yang tersedia untuk umum. Dalam kasus seperti itu, Anda dapat membuat gambar untuk dijalankan sebagai pengguna non-root dalam Dockerfile. Apakah Anda membuat pengguna sebagai root atau non-root, Anda perlu memastikan bahwa UID/GID pengguna identik dengan UID/GID di untuk aplikasi kustom, yang membuat konfigurasi AppImageConfig SageMaker untuk menjalankan aplikasi menggunakan gambar kustom. Misalnya, jika Dockerfile Anda dibuat untuk pengguna non-root seperti berikut ini:

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

`AppImageConfig`File perlu menyebutkan UID dan GID yang sama dalam: `KernelGatewayConfig`

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

Nilai UID/GID yang dapat diterima untuk gambar khusus adalah 0/0 dan 1000/100 untuk gambar Studio. Untuk contoh pembuatan gambar kustom dan `AppImageConfig` pengaturan terkait, lihat repositori Github ini.

Untuk menghindari pengguna merusak ini, jangan berikan, `CreateAppImageConfigUpdateAppImageConfig`, atau `DeleteAppImageConfig` izin kepada pengguna notebook SageMaker Studio.

Manajemen jaringan

Untuk mengatur domain SageMaker Studio, Anda perlu menentukan jaringan VPC, subnet, dan grup keamanan. Saat menentukan VPC dan subnet, pastikan Anda mengalokasikan IP dengan mempertimbangkan volume penggunaan dan pertumbuhan yang diharapkan yang dibahas di bagian berikut.

Perencanaan jaringan VPC

Subnet VPC pelanggan yang terkait dengan domain SageMaker Studio harus dibuat dengan rentang Classless Inter-domain Routing (CIDR) yang sesuai, tergantung pada faktor-faktor berikut:

- Jumlah pengguna.
- Jumlah aplikasi per pengguna.
- Jumlah jenis instance unik per pengguna.
- Rata-rata jumlah instans pelatihan per pengguna.
- Persentase pertumbuhan yang diharapkan.

SageMaker dan AWS layanan yang berpartisipasi menyuntikkan antarmuka jaringan elastis (ENI) ke subnet VPC pelanggan untuk kasus penggunaan berikut:

- Amazon EFS menyuntikkan ENI untuk target pemasangan EFS untuk SageMaker domain (satu IP per subnet/Availability Zone yang dilampirkan ke domain). SageMaker
- SageMaker Studio menyuntikkan ENI untuk setiap instance unik yang digunakan oleh profil pengguna atau ruang bersama. Sebagai contoh:
  - Jika profil pengguna menjalankan aplikasi server Jupyter default (satu instance 'sistem'), aplikasi Ilmu Data, dan aplikasi Python Dasar (keduanya berjalan pada `ml.t3.medium` instance), Studio menyuntikkan dua alamat IP.
  - Jika profil pengguna menjalankan aplikasi server Jupyter default (satu instance 'sistem'), aplikasi GPU Tensorflow (pada instance), dan aplikasi data wrangler (pada `ml.g4dn.xlarge` instance), Studio menyuntikkan tiga alamat `ml.m5.4xlarge` IP.
- ENI untuk setiap titik akhir VPC di seluruh subnet/Availability Zone VPC domain disuntikkan (empat IP untuk titik akhir VPC; ~ enam IP untuk layanan yang berpartisipasi SageMaker titik akhir VPC seperti S3, ECR, dan.) CloudWatch

- Jika pekerjaan SageMaker pelatihan dan pemrosesan diluncurkan dengan konfigurasi VPC yang sama, setiap pekerjaan membutuhkan dua alamat IP per instance.

> **Note**
>
> Pengaturan VPC untuk SageMaker Studio, seperti subnet dan lalu lintas khusus VPC, tidak secara otomatis diteruskan ke pekerjaan pelatihan/pemrosesan yang dibuat dari Studio. SageMaker Pengguna perlu mengatur pengaturan VPC dan isolasi jaringan seperlunya saat memanggil Create*Job API. Lihat Jalankan Pelatihan dan Kontainer Inferensi dalam Mode Bebas Internet untuk informasi lebih lanjut.

Skenario: Ilmuwan data menjalankan eksperimen pada dua jenis instance yang berbeda

Dalam skenario ini, asumsikan SageMaker domain diatur dalam mode lalu lintas khusus VPC. Ada titik akhir VPC yang disiapkan, seperti SageMaker API, SageMaker runtime, Amazon S3, dan Amazon ECR.

Seorang ilmuwan data menjalankan eksperimen pada notebook Studio, berjalan pada dua jenis instance yang berbeda (misalnya, `ml.t3.medium` dan`ml.m5.large`), dan meluncurkan dua aplikasi di setiap jenis instance.

Asumsikan ilmuwan data juga secara bersamaan menjalankan pekerjaan pelatihan dengan konfigurasi VPC yang sama pada sebuah `ml.m5.4xlarge` instance.

Untuk skenario ini, layanan SageMaker Studio akan menyuntikkan ENI sebagai berikut:

Tabel 1 — ENI disuntikkan ke VPC pelanggan untuk skenario eksperimen

Entitas	Target	ENI disuntikkan	Catatan	Tingkat
Target pemasangan EFS	Subnet VPC	Tiga	Tiga AZS/Subnet	Domain
Titik akhir VPC	Subnet VPC	30	Tiga AZS/Subnet dengan masing-masing 10 VPCE	Domain

Entitas	Target	ENI disuntikkan	Catatan	Tingkat
Server Jupyter	Subnet VPC	One	Satu IP per instance	Pengguna
KernelGateway aplikasi	Subnet VPC	Dua	Satu IP per jenis instans	Pengguna
Pelatihan	Subnet VPC	Dua	Dua IP per contoh pelatihan Lima IP per instance pelatihan jika EFA digunakan	Pengguna

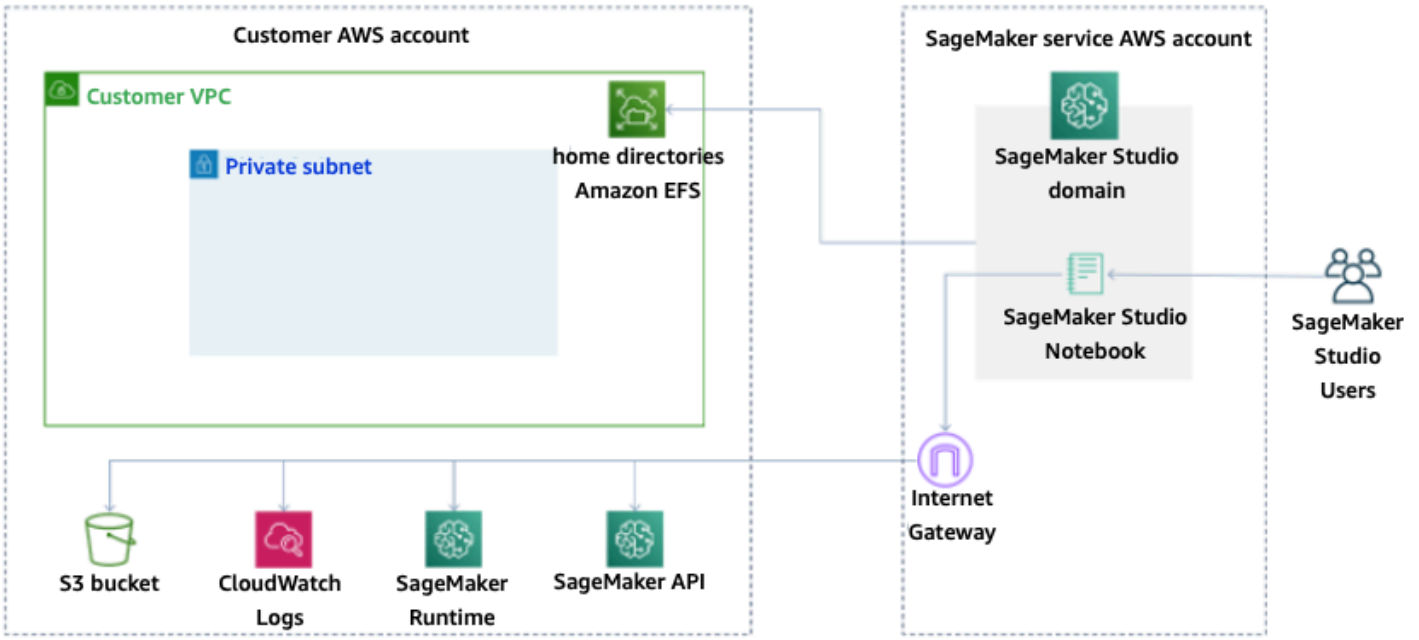
Untuk skenario ini, ada total 38 IP yang dikonsumsi dalam VPC pelanggan di mana 33 IP dibagikan di seluruh pengguna di tingkat domain, dan lima IP dikonsumsi di tingkat pengguna. Jika Anda memiliki 100 pengguna dengan profil pengguna serupa di domain ini yang melakukan aktivitas ini secara bersamaan, maka Anda akan mengkonsumsi lima x 100 = 500 IP di tingkat pengguna, di atas konsumsi IP tingkat domain, yaitu 11 IP per subnet, dengan total 511 IP. Untuk skenario ini, Anda perlu membuat subnet VPC CIDR dengan/22 yang akan mengalokasikan 1024 alamat IP, dengan ruang untuk tumbuh.

Opsi jaringan VPC

Domain SageMaker Studio mendukung konfigurasi jaringan VPC dengan salah satu opsi berikut:

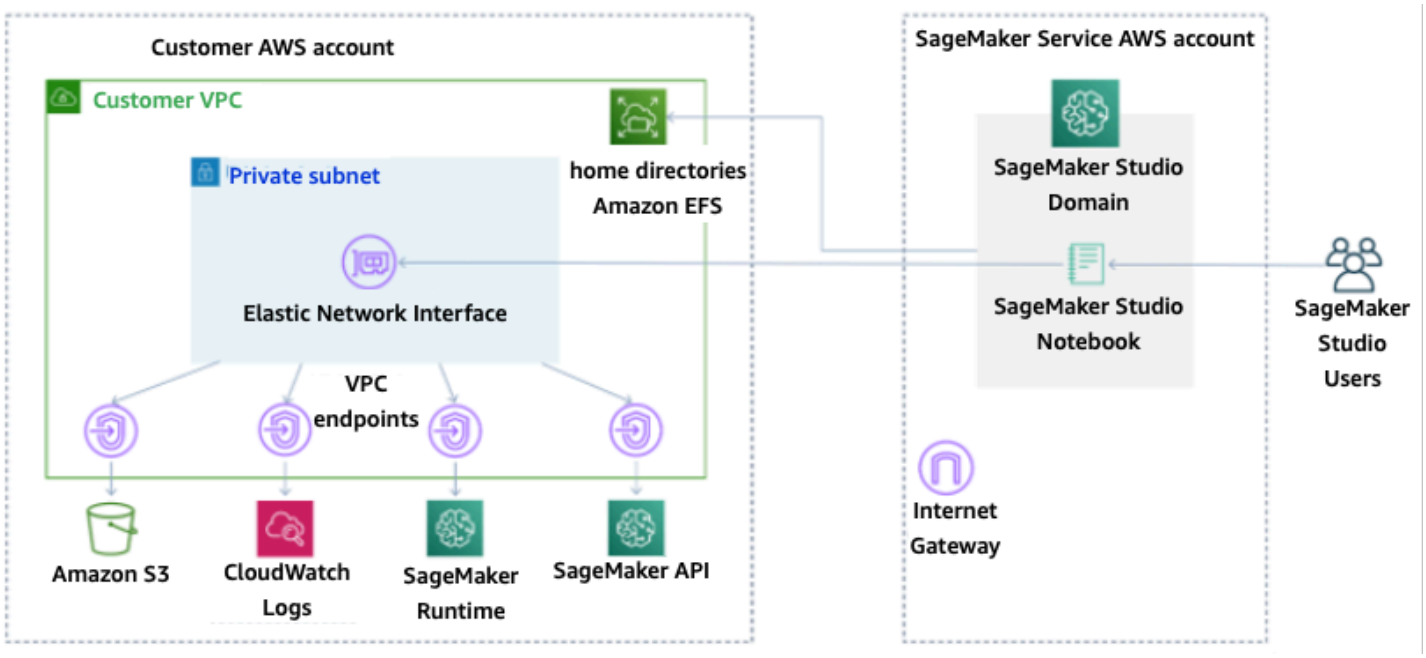
- Hanya internet publik
- Hanya VPC

Opsi khusus internet publik memungkinkan layanan SageMaker API untuk menggunakan internet publik melalui gateway internet yang disediakan di VPC, yang dikelola oleh akun SageMaker layanan, seperti yang terlihat pada diagram berikut:



Mode default: Akses Internet melalui akun SageMaker layanan

Opsi hanya VPC menonaktifkan perutean internet dari VPC yang dikelola oleh akun SageMaker layanan, dan memungkinkan pelanggan untuk mengonfigurasi lalu lintas yang akan dirutekan melalui titik akhir VPC, seperti yang terlihat pada diagram berikut:



Mode khusus VPC: Tidak ada akses internet melalui akun layanan SageMaker

Untuk pengaturan domain dalam mode khusus VPC, siapkan grup keamanan per profil pengguna untuk memastikan isolasi lengkap instance yang mendasarinya. Setiap domain dalam AWS akun dapat memiliki konfigurasi VPC dan mode internet sendiri. Untuk detail selengkapnya mengenai pengaturan konfigurasi jaringan VPC, lihat Connect SageMaker Studio Notebooks dalam VPC ke Sumber Daya Eksternal.

Batasan

- Setelah domain SageMaker Studio dibuat, Anda tidak dapat mengaitkan subnet baru ke domain tersebut.
- Jenis jaringan VPC (hanya internet publik atau hanya VPC) tidak dapat diubah.

Perlindungan data

Sebelum merancang beban kerja ML, praktik dasar yang memengaruhi keamanan harus ada. Misalnya, klasifikasi data menyediakan cara untuk mengkategorikan data berdasarkan tingkat sensitivitas, dan enkripsi melindungi data dengan membuatnya tidak dapat dipahami oleh akses yang tidak sah. Metode-metode ini penting, karena mendukung tujuan seperti mencegah kesalahan penanganan atau mematuhi kewajiban peraturan.

SageMaker Studio menyediakan beberapa fitur untuk melindungi data saat istirahat dan dalam perjalanan. Namun, seperti yang dijelaskan dalam model Tanggung Jawab AWS Bersama, pelanggan bertanggung jawab untuk menjaga kontrol atas konten yang di-host di infrastruktur AWS Global. Di bagian ini, kami menjelaskan bagaimana pelanggan dapat menggunakan fitur tersebut untuk melindungi data mereka.

Lindungi data saat istirahat

Untuk melindungi notebook SageMaker Studio Anda bersama dengan data pembuatan model dan artefak model Anda, SageMaker enkripsi notebook, serta output dari pelatihan dan pekerjaan transformasi batch. SageMaker mengenkripsi ini secara default, menggunakan Kunci AWS Terkelola untuk Amazon S3. Kunci AWS Terkelola untuk Amazon S3 ini tidak dapat dibagikan untuk akses lintas akun. Untuk akses lintas akun, tentukan kunci yang dikelola pelanggan Anda sambil membuat SageMaker sumber daya sehingga dapat dibagikan untuk akses lintas akun.

Dengan SageMaker Studio, data dapat disimpan di lokasi berikut:

- Bucket S3 — Saat notebook yang dapat dibagikan diaktifkan, SageMaker Studio membagikan snapshot dan metadata notebook dalam bucket S3.
- Volume EFS — SageMaker Studio melampirkan volume EFS ke domain Anda untuk menyimpan notebook dan file data. Volume EFS ini tetap ada bahkan setelah domain dihapus.
- Volume EBS — EBS dilampirkan ke instance tempat notebook berjalan. Volume ini bertahan selama durasi instance.

Enkripsi saat istirahat dengan AWS KMS

- Anda dapat meneruskan AWS KMSkunci Anda untuk mengenkripsi volume EBS yang dilampirkan ke notebook, pelatihan, penyetelan, pekerjaan transformasi batch, dan titik akhir.

- Jika Anda tidak menentukan kunci KMS, SageMaker mengenkripsi volume sistem operasi (OS) dan volume data ML dengan kunci KMS yang dikelola sistem.
- Data sensitif yang perlu dienkripsi dengan kunci KMS untuk alasan kepatuhan harus disimpan dalam volume penyimpanan ML atau di Amazon S3, yang keduanya dapat dienkripsi menggunakan kunci KMS yang Anda tentukan.

Melindungi data saat transit

SageMaker Studio memastikan bahwa artefak model ML dan artefak sistem lainnya dienkripsi saat transit dan saat istirahat. Permintaan ke SageMaker API dan konsol dibuat melalui koneksi aman (SSL). Beberapa data intra-jaringan dalam transit (di dalam platform layanan) tidak dienkripsi. Hal ini mencakup:

- Komunikasi perintah dan kontrol antara pesawat kontrol layanan dan instance pekerjaan pelatihan (bukan data pelanggan).
- Komunikasi antar node dalam pemrosesan terdistribusi dan pekerjaan pelatihan (intra-jaringan).

Namun, Anda dapat memilih untuk mengenkripsi komunikasi antar node dalam cluster pelatihan. Mengaktifkan enkripsi lalu lintas antar kontainer dapat meningkatkan waktu pelatihan, terutama jika Anda menggunakan algoritme pembelajaran mendalam terdistribusi.

Secara default, Amazon SageMaker menjalankan tugas pelatihan di Amazon VPC untuk membantu menjaga keamanan data Anda. Anda dapat menambahkan tingkat keamanan lain untuk melindungi wadah pelatihan dan data Anda dengan mengonfigurasi VPC pribadi. Selanjutnya, Anda dapat mengonfigurasi domain SageMaker Studio agar berjalan dalam mode VPC saja, dan mengatur titik akhir VPC untuk merutekan lalu lintas melalui jaringan pribadi tanpa mengurangi lalu lintas melalui internet.

Pagar perlindungan data

Enkripsi volume SageMaker hosting saat istirahat

Gunakan kebijakan berikut untuk menerapkan enkripsi selama menghosting SageMaker titik akhir untuk inferensi online:

```
{
```

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

Enkripsi bucket S3 yang digunakan selama Pemantauan Model

Model Monitoring menangkap data yang dikirim ke SageMaker titik akhir Anda dan menyimpannya dalam bucket S3. Saat menyiapkan Data Capture Config, Anda perlu mengenkripsi bucket S3. Saat ini tidak ada kontrol kompensasi untuk ini.

Selain menangkap output titik akhir, layanan Pemantauan Model memeriksa penyimpangan terhadap garis dasar yang telah ditentukan sebelumnya. Anda perlu mengenkripsi output dan volume penyimpanan menengah yang digunakan untuk memantau penyimpangan.

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

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

Mengenkripsi volume penyimpanan domain SageMaker Studio

Menerapkan enkripsi ke volume penyimpanan yang dilampirkan ke domain Studio. Kebijakan ini mengharuskan pengguna untuk menyediakan CMK untuk mengenkripsi volume penyimpanan yang dilampirkan ke domain studio.

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

Enkripsi data yang disimpan di S3 yang digunakan untuk berbagi notebook

Ini adalah kebijakan untuk mengenkripsi data apa pun yang disimpan dalam bucket yang digunakan untuk berbagi buku catatan antar pengguna dalam domain SageMaker Studio:

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

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

Keterbatasan:

- Setelah domain dibuat, Anda tidak dapat memperbarui penyimpanan volume EFS terlampir dengan AWS KMS kunci khusus.
- Anda tidak dapat memperbarui pekerjaan pelatihan/pemrosesan atau konfigurasi titik akhir dengan kunci KMS setelah dibuat.

Pencatatan dan pemantauan

Untuk membantu Anda men-debug pekerjaan kompilasi, pekerjaan pemrosesan, pekerjaan pelatihan, titik akhir, tugas transformasi, instance buku catatan, dan konfigurasi siklus hidup instance notebook, apa pun yang dikirim oleh wadah algoritme, wadah model, atau konfigurasi siklus hidup instance notebook ke stdout atau stderr juga dikirim ke Amazon Logs. CloudWatch Anda dapat memantau SageMaker Studio menggunakan Amazon CloudWatch, yang mengumpulkan data mentah dan memprosesnya menjadi metrik yang dapat dibaca, mendekati waktu nyata. Statistik ini disimpan selama 15 bulan, sehingga Anda dapat mengakses informasi historis dan mendapatkan perspektif yang lebih baik tentang kinerja aplikasi atau layanan web Anda.

Logging dengan CloudWatch

Karena proses ilmu data secara inheren bersifat eksperimental dan berulang, penting untuk mencatat aktivitas seperti penggunaan notebook, waktu kerja pelatihan/pemrosesan, metrik pelatihan, dan metrik penyajian titik akhir seperti latensi pemanggilan. Secara default, SageMaker menerbitkan metrik ke CloudWatch Log, dan log ini dapat dienkripsi dengan kunci yang dikelola pelanggan menggunakan. AWS KMS

Anda juga dapat menggunakan titik akhir VPC untuk mengirim log CloudWatch tanpa menggunakan internet publik. Anda juga dapat mengatur alarm yang memperhatikan ambang batas tertentu dan mengirim notifikasi atau mengambil tindakan saat ambang batas tersebut terpenuhi. Untuk informasi selengkapnya, lihat Panduan CloudWatch Pengguna Amazon.

SageMaker membuat grup log tunggal untuk Studio, di bawah`/aws/sagemaker/studio`. Setiap profil pengguna dan aplikasi memiliki aliran log mereka sendiri di bawah grup log ini, dan skrip konfigurasi siklus hidup memiliki aliran log mereka sendiri juga. Misalnya, profil pengguna bernama 'studio-user' dengan aplikasi Jupyter Server dan dengan skrip siklus hidup terlampir, dan aplikasi Data Science Kernel Gateway memiliki aliran log berikut:

`/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 Untuk mengirim log atas nama Anda, pemanggil API pekerjaan Training/Processing/Transform akan memerlukan izin berikut: CloudWatch

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

Untuk mengenkripsi log tersebut dengan AWS KMS kunci khusus, Anda harus terlebih dahulu memodifikasi kebijakan kunci untuk memungkinkan CloudWatch layanan mengenkripsi dan mendekripsi kunci. Setelah Anda membuat AWS KMS kunci enkripsi log, ubah kebijakan kunci untuk menyertakan yang berikut:

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

Perhatikan bahwa Anda selalu dapat menggunakan `ArnEquals` dan memberikan Amazon Resource Name (ARN) tertentu untuk CloudWatch log yang ingin Anda enkripsi. Di sini kami menunjukkan bahwa Anda dapat menggunakan kunci ini untuk mengenkripsi semua log dalam akun untuk kesederhanaan. Selain itu, pelatihan, pemrosesan, dan titik akhir model menerbitkan metrik tentang CPU instance dan pemanfaatan memori, latensi pemanggilan hosting, dan sebagainya. Anda dapat mengonfigurasi Amazon SNS lebih lanjut untuk memberi tahu administrator tentang peristiwa ketika ambang batas tertentu dilewati. Konsumen API pelatihan dan pemrosesan harus memiliki izin berikut:

```
{
     "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 dengan AWS CloudTrail

Untuk meningkatkan postur kepatuhan Anda, audit semua API Anda denganAWS CloudTrail. Secara default, semua SageMaker API dicatat dengan AWS CloudTrail. Anda tidak memerlukan izin IAM tambahan untuk mengaktifkan. CloudTrail

Semua SageMaker tindakan, dengan pengecualian `InvokeEndpoint` dan`InvokeEndpointAsync`, dicatat oleh CloudTrail dan didokumentasikan dalam operasi. Misalnya, panggilan ke`CreateTrainingJob`,`CreateEndpoint`, dan `CreateNotebookInstance` tindakan menghasilkan entri dalam file CloudTrail log.

Setiap entri CloudTrail acara berisi informasi tentang siapa yang membuat permintaan. Informasi identitas membantu Anda menentukan berikut ini:

- Apakah permintaan dibuat dengan kredenal pengguna root atau AWS IAM.
- Jika permintaan tersebut dibuat dengan kredensial keamanan sementara untuk peran atau pengguna federasi.
- Bahwa permintaan dibuat oleh layanan AWS lain. Untuk contoh peristiwa, lihat Panggilan SageMaker API Log dengan CloudTrail dokumentasi.

Secara default, CloudTrail mencatat nama peran eksekusi Studio dari profil pengguna sebagai pengenal untuk setiap peristiwa. Ini berfungsi jika setiap pengguna memiliki peran eksekusi mereka sendiri. Jika beberapa pengguna berbagi peran eksekusi yang sama, Anda dapat menggunakan

`sourceIdentity` konfigurasi untuk menyebarkan nama profil pengguna Studio ke CloudTrail. Lihat Memantau akses sumber daya pengguna dari Amazon SageMaker Studio untuk mengaktifkan `sourceIdentity` fitur. Dalam ruang bersama, semua tindakan merujuk ke ruang ARN sebagai sumber, dan Anda tidak dapat mengaudit melalui. `sourceIdentity`

Atribusi biaya

SageMaker Studio memiliki kemampuan bawaan untuk membantu administrator melacak pengeluaran masing-masing domain, ruang bersama, dan pengguna.

Penandaan otomatis

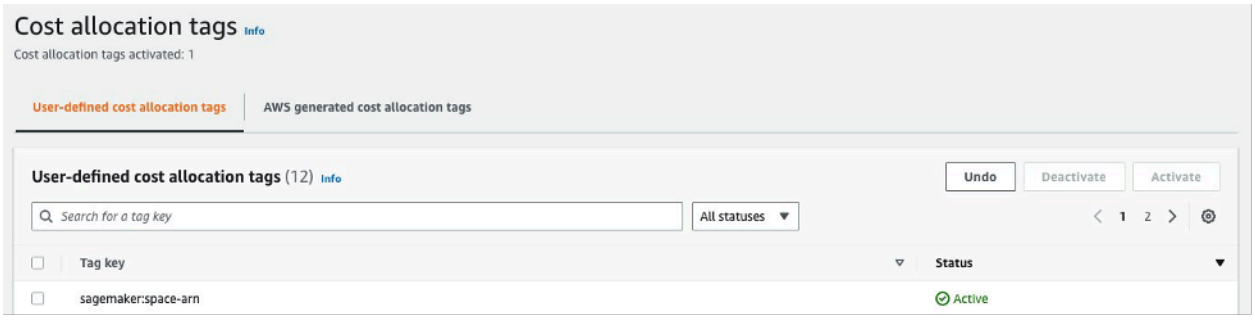
SageMaker Studio sekarang secara otomatis menandai SageMaker sumber daya baru seperti pekerjaan pelatihan, pekerjaan pemrosesan, dan aplikasi kernel dengan masing-masing`sagemaker:domain-arn`. Pada tingkat yang lebih terperinci, SageMaker juga menandai sumber daya dengan `sagemaker:user-profile-arn` atau `sagemaker:space-arn` untuk menunjuk pencipta utama sumber daya.

SageMaker volume EFS domain ditandai dengan kunci bernama `ManagedByAmazonSageMakerResource` dengan nilai domain ARN. Mereka tidak memiliki tag granular untuk memahami penggunaan ruang pada tingkat per pengguna. Administrator dapat melampirkan volume EFS ke instans EC2 untuk pemantauan yang dipesan lebih dahulu.

Pemantauan biaya

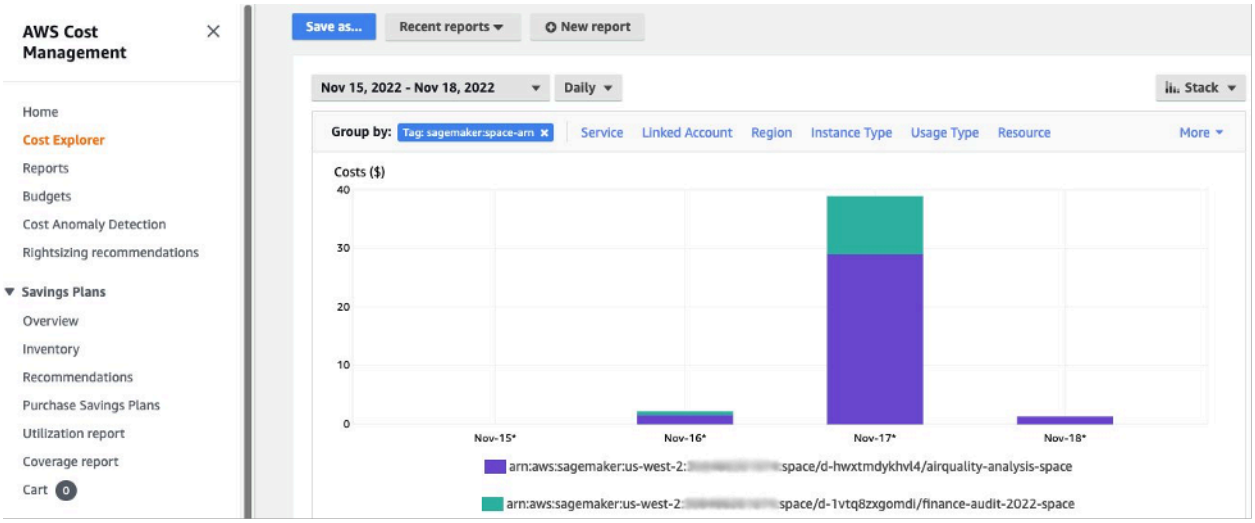
Tag otomatis memungkinkan Administrator melacak, melaporkan, dan memantau pengeluaran ML Anda melalui out-of-the-box solusi seperti AWS Cost Explorerdan AWS Budgets, serta solusi khusus yang dibuat berdasarkan data dari Laporan AWS Biaya dan Penggunaan (CURs).

Untuk menggunakan tag terlampir untuk analisis biaya, tag tersebut harus diaktifkan terlebih dahulu di bagian Tag alokasi biaya di AWS Billing konsol. Diperlukan waktu hingga 24 jam agar tag muncul di panel tag alokasi biaya, jadi Anda harus membuat SageMaker sumber daya sebelum mengaktifkannya.



Space ARN diaktifkan sebagai tag alokasi biaya pada Cost Explorer

Setelah Anda mengaktifkan tag alokasi biaya, AWS akan mulai melacak sumber daya yang ditandai, dan setelah 24-48 jam, tag akan muncul sebagai filter yang dapat dipilih di penjelajah biaya.



Biaya dikelompokkan berdasarkan ruang bersama untuk domain sampel

Kontrol biaya

Saat pengguna SageMaker Studio pertama di-onboard, SageMaker buat volume EFS untuk domain tersebut. Biaya penyimpanan dikeluarkan untuk volume EFS ini karena notebook dan file data disimpan di direktori home pengguna. Saat pengguna meluncurkan notebook Studio, notebook tersebut diluncurkan untuk instance komputasi yang menjalankan notebook. Lihat SageMaker harga Amazon untuk rincian rinci biaya.

Administrator dapat mengontrol biaya komputasi dengan menentukan daftar instance yang dapat diputar pengguna, menggunakan kebijakan IAM seperti yang disebutkan di bagian Common guardrails. Selain itu, kami menyarankan agar pelanggan menggunakan ekstensi auto shutdown SageMaker Studio untuk menghemat biaya dengan mematikan aplikasi idle secara otomatis. Ekstensi server ini secara berkala melakukan polling untuk menjalankan aplikasi per profil pengguna, dan mematikan aplikasi idle berdasarkan batas waktu yang ditetapkan oleh administrator.

Untuk mengatur ekstensi ini untuk semua pengguna di domain Anda, Anda dapat menggunakan konfigurasi siklus hidup seperti yang dijelaskan di bagian Kustomisasi. Selain itu, Anda juga dapat menggunakan pemeriksa ekstensi untuk memastikan semua pengguna domain Anda memiliki ekstensi yang diinstal.

Kustomisasi

Konfigurasi siklus hidup

Konfigurasi siklus hidup adalah skrip shell yang diprakarsai oleh peristiwa siklus hidup SageMaker Studio, seperti memulai notebook Studio baru. SageMaker Anda dapat menggunakan skrip shell ini untuk mengotomatiskan penyesuaian untuk lingkungan SageMaker Studio Anda, seperti menginstal paket kustom, ekstensi Jupyter untuk mematikan otomatis aplikasi notebook yang tidak aktif, dan menyiapkan konfigurasi Git. Untuk petunjuk terperinci tentang cara membuat konfigurasi siklus hidup, lihat blog ini: Sesuaikan Amazon SageMaker Studio menggunakan Konfigurasi Siklus Hidup.

Gambar kustom untuk notebook SageMaker Studio

Notebook studio dilengkapi dengan satu set gambar pra-bangun, yang terdiri dari Amazon SageMaker Python SDK dan versi terbaru dari runtime atau kernel IPython. Dengan fitur ini, Anda dapat membawa gambar kustom Anda sendiri ke SageMaker notebook Amazon. Gambar-gambar ini kemudian tersedia untuk semua pengguna yang diautentikasi ke dalam domain.

Pengembang dan ilmuwan data mungkin memerlukan gambar khusus untuk beberapa kasus penggunaan yang berbeda:

- Akses ke versi spesifik atau terbaru dari kerangka kerja HTML populer seperti TensorFlow, PyTorch MXNet, atau lainnya.
- Bawa kode kustom atau algoritme yang dikembangkan secara lokal ke notebook SageMaker Studio untuk iterasi cepat dan pelatihan model.
- Akses ke data lake atau penyimpanan data lokal melalui API. Admin harus menyertakan driver yang sesuai dalam gambar.
- Akses ke runtime backend (juga disebut kernel), selain IPython (seperti R, Julia, atau lainnya). Anda juga dapat menggunakan pendekatan yang diuraikan untuk menginstal kernel khusus.

Untuk petunjuk terperinci tentang cara membuat gambar kustom, lihat Membuat SageMaker gambar kustom.

JupyterLab ekstensi

Dengan SageMaker Studio JuypterLab 3 Notebook, Anda dapat memanfaatkan komunitas ekstensi sumber terbuka JupyterLab yang terus berkembang. Bagian ini menyoroti beberapa yang secara alami sesuai dengan alur kerja SageMaker pengembang, tetapi kami mendorong Anda untuk menelusuri ekstensi yang tersedia atau bahkan membuatnya sendiri.

JupyterLab 3 sekarang membuat proses pengemasan dan pemasangan ekstensi secara signifikan lebih mudah. Anda dapat menginstal ekstensi yang disebutkan di atas melalui skrip bash. Misalnya, di SageMaker Studio, buka terminal sistem dari peluncur Studio dan jalankan perintah berikut. Selain itu, Anda dapat mengotomatiskan penginstalan ekstensi ini menggunakan konfigurasi siklus hidup sehingga tetap ada di antara restart Studio. Anda dapat mengonfigurasi ini untuk semua pengguna di domain atau pada tingkat pengguna individu.

Misalnya, untuk menginstal ekstensi untuk browser file Amazon S3, jalankan perintah berikut di terminal sistem dan pastikan refresh browser Anda:

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

Untuk informasi selengkapnya tentang manajemen ekstensi, termasuk cara menulis konfigurasi siklus hidup yang berfungsi untuk JupyterLab notebook versi 1 dan 3 untuk kompatibilitas mundur, lihat ekstensi JupyterLab Instalasi dan Jupyter Server.

Repositori Git

SageMaker Studio dilengkapi pra-instal dengan ekstensi Jupyter Git bagi pengguna untuk memasukkan URL yang dipesan lebih dahulu dari repositori Git, mengkloningnya ke direktori EFS Anda, mendorong perubahan, dan melihat riwayat komit. Administrator dapat mengonfigurasi repo git yang disarankan di tingkat domain sehingga muncul sebagai pilihan drop-down untuk pengguna akhir. Lihat Lampirkan Repos Git yang Disarankan ke Studio untuk up-to-date instruksi.

Jika repositori bersifat pribadi, ekstensi akan meminta pengguna untuk memasukkan kredensialnya ke terminal menggunakan instalasi git standar. Atau, pengguna dapat menyimpan kredensyal ssh di direktori EFS masing-masing untuk pengelolaan yang lebih mudah.

Lingkungan Conda

SageMaker Notebook studio menggunakan Amazon EFS sebagai lapisan penyimpanan persisten. Ilmuwan data dapat menggunakan penyimpanan persisten untuk membuat lingkungan conda khusus dan menggunakan lingkungan ini untuk membuat kernel. Kernel ini didukung oleh EFS, dan persisten antara kernel, aplikasi, atau Studio restart. Studio secara otomatis mengambil semua lingkungan yang valid sebagai KernelGateway kernel.

Proses untuk membuat lingkungan conda sangat mudah bagi seorang ilmuwan data, tetapi kernel membutuhkan waktu sekitar satu menit untuk mengisi pemilih kernel. Untuk membuat lingkungan, jalankan yang berikut ini di terminal sistem:

```
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
```

Untuk petunjuk terperinci, lihat lingkungan Persist Conda ke bagian volume Studio EFS dalam Empat pendekatan untuk mengelola paket Python di notebook Amazon Studio. SageMaker

Kesimpulan

Dalam whitepaper ini, kami meninjau beberapa praktik terbaik di berbagai bidang seperti model operasi, manajemen domain, manajemen identitas, manajemen izin, manajemen jaringan, pencatatan, pemantauan, dan penyesuaian untuk memungkinkan administrator platform menyiapkan dan mengelola Platform Studio. SageMaker

Lampiran

Perbandingan multi-penyewaan

Tabel 2 - Perbandingan multi-penyewaan

Multi-domain	Multi-akun	Kontrol akses berbasis atribut (ABAC) dalam satu domain
Isolasi sumber daya dicapai dengan menggunakan tag. SageMaker Studio secara otomatis menandai semua sumber daya dengan ARN domain dan profil pengguna/ruang ARN.	Setiap penyewa ada di akun mereka sendiri, jadi ada isolasi sumber daya absolut.	Isolasi sumber daya dicapai dengan menggunakan tag. Pengguna harus mengelola penandaan sumber daya yang dibuat untuk ABAC.
API daftar tidak dapat dibatasi oleh tag. Pemfilteran sumber daya UI dilakukan pada ruang bersama, namun, panggilan List API yang dilakukan melalui AWS CLI atau Boto3 SDK akan mencantumkan sumber daya di seluruh Wilayah.	Isolasi API daftar juga dimungkinkan, karena penyewa ada di akun khusus mereka.	API daftar tidak dapat dibatasi oleh tag. Daftar panggilan API yang dilakukan melalui AWS CLI atau Boto3 SDK akan mencantumkan sumber daya di seluruh Wilayah.
SageMaker Biaya komputasi dan penyimpanan studio per penyewa dapat dengan mudah dipantau dengan menggunakan Domain ARN sebagai tag alokasi biaya.	SageMaker Biaya komputasi dan penyimpanan studio per penyewa mudah dipantau dengan akun khusus.	SageMaker Biaya komputasi studio per penyewa perlu dihitung menggunakan tag khusus. SageMaker Biaya penyimpan an studio tidak dapat dipantau per domain karena semua

Multi-domain	Multi-akun	Kontrol akses berbasis atribut (ABAC) dalam satu domain
		penyewa memiliki volume EFS yang sama.
Kuota layanan ditetapkan pada tingkat akun, sehingga penyewa tunggal masih dapat menggunakan semua sumber daya.	Kuota layanan dapat ditetapkan pada tingkat akun untuk setiap penyewa.	Kuota layanan ditetapkan pada tingkat akun, sehingga penyewa tunggal masih dapat menggunakan semua sumber daya.
Penskalaan ke beberapa penyewa dapat dicapai melalui infrastruktur sebagai kode (IAc) atau Service Catalog.	Penskalaan ke beberapa penyewa melibatkan Organizations dan penjual beberapa akun.	Penskalaan membutuhkan peran khusus penyewa untuk setiap penyewa baru, dan profil pengguna harus ditandai secara manual dengan nama penyewa.
Kolaborasi antara pengguna dalam penyewa dimungkinkan melalui ruang bersama.	Kolaborasi antara pengguna dalam penyewa dimungkinkan melalui ruang bersama.	Semua penyewa akan memiliki akses ke ruang bersama yang sama untuk kolaborasi.

SageMaker Pencadangan dan pemulihan domain studio

Jika terjadi penghapusan EFS yang tidak disengaja, atau ketika domain perlu dibuat ulang karena perubahan jaringan atau otentikasi, ikuti petunjuk ini.

Opsi 1: Cadangkan dari EFS yang ada menggunakan EC2

SageMaker Pencadangan domain studio

1. Buat daftar profil dan spasi pengguna di SageMaker Studio (CLI, SDK).
2. Memetakan profil/spasi pengguna ke UID di EFS.
   a. Untuk setiap pengguna dalam daftar pengguna/spasi, jelaskan profil/spasi pengguna (CLI, SDK).

   b. Petakan profil/spasi pengguna ke. `HomeEfsFileSystemUid`
   c. Petakan profil pengguna ke `UserSettings['ExecutionRole']` jika pengguna memiliki peran eksekusi yang berbeda.
   d. Identifikasi peran eksekusi Space default.
3. Buat domain baru dan tentukan peran eksekusi Space default.
4. Buat profil dan spasi pengguna.
   - Untuk setiap pengguna dalam daftar pengguna, buat profil pengguna (CLI, SDK) menggunakan pemetaan peran eksekusi.
5. Buat pemetaan untuk EFS dan UID baru.
   a. Untuk setiap pengguna dalam daftar pengguna, jelaskan profil pengguna (CLI, SDK).
   b. Peta profil pengguna ke`HomeEfsFileSystemUid`.
6. Secara opsional, hapus semua aplikasi, profil pengguna, spasi, lalu hapus domain.

Cadangan EFS

Untuk membuat cadangan EFS, gunakan instruksi berikut:

1. Luncurkan instans EC2, dan lampirkan grup keamanan masuk/keluar domain SageMaker Studio lama ke instans EC2 baru (izinkan lalu lintas NFS melalui TCP pada port 2049. Lihat Connect SageMaker Studio Notebook dalam VPC ke Sumber Daya Eksternal.
2. Pasang volume SageMaker Studio EFS ke instans EC2 baru. Lihat sistem file Mounting EFS.
3. Salin file ke penyimpanan lokal EBS: `>sudo cp -rp /efs /studio-backup:`
   a. Lampirkan grup keamanan domain baru ke instans EC2.
   b. Pasang volume EFS baru ke instans EC2.
   c. Salin file ke volume EFS baru.
   d. Untuk setiap pengguna dalam koleksi pengguna:
      i. Buat direktori:`mkdir new_uid`.
      ii. Salin file dari direktori UID lama ke direktori UID baru.
      iii. Ubah kepemilikan untuk semua file: `chown <new_UID>` untuk semua file.

Opsi 2: Cadangkan dari EFS yang ada menggunakan konfigurasi S3 dan siklus hidup

1. Lihat Memigrasi pekerjaan Anda ke instans SageMaker notebook Amazon dengan Amazon Linux 2.
2. Buat bucket S3 untuk cadangan (seperti>`studio-backup`.
3. Buat daftar semua profil pengguna dengan peran eksekusi.
4. Di domain SageMaker Studio saat ini, tetapkan skrip LCC default di tingkat domain.
   - Di LCC, salin semuanya `/home/sagemaker-user` ke awalan profil pengguna di S3 (misalnya,). `s3://studio-backup/studio-user1`
5. Mulai ulang semua aplikasi Server Jupyter default (agar LCC dijalankan).
6. Hapus semua aplikasi, profil pengguna, dan domain.
7. Buat domain SageMaker Studio baru.
8. Buat profil pengguna baru dari daftar profil pengguna dan peran eksekusi.
9. Siapkan LCC di tingkat domain:
   - Di LCC, salin semua yang ada di awalan profil pengguna di S3 ke `/home/sagemaker-user`
10. Buat aplikasi Jupyter Server default untuk semua pengguna dengan konfigurasi LCC (CLI, SDK).

SageMaker Akses studio menggunakan pernyataan SAMP

Pengaturan solusi:

1. Buat aplikasi SAMP di IDP eksternal Anda.
2. Siapkan iDP eksternal sebagai Penyedia Identitas di IAM.
3. Buat fungsi `SAMLValidator` Lambda yang dapat diakses oleh IDP (melalui URL fungsi atau API Gateway).
4. Buat fungsi `GeneratePresignedUrl` Lambda dan API Gateway untuk mengakses fungsi tersebut.
5. Buat peran IAM yang dapat diasumsikan pengguna untuk memanggil API Gateway. Peran ini harus diteruskan dalam pernyataan SAMP sebagai atribut dalam format berikut:
   - Nama atribut: https://aws.amazon.com/SAML/Attributes/Role
   - Nilai atribut:`<IdentityProviderARN>`, `<RoleARN>`

6. Perbarui titik akhir SAMP Assertion Consumer Service (ACS) ke URL pemanggilan. `SAMLValidator`

Kode contoh validator SAMP:

```
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
```

Bacaan lebih lanjut

- Menyiapkan lingkungan pembelajaran mesin yang aman dan diatur dengan baik di AWS (blog) AWS
- Mengonfigurasi Amazon SageMaker Studio untuk tim dan grup dengan isolasi sumber daya lengkap (AWSblog)
- Orientasi Amazon SageMaker Studio dengan AWS SSO dan Okta Universal Directory (blog) AWS
- Cara Mengkonfigurasi SAMP 2.0 untuk Federasi AWS Akun (dokumentasi Okta)
- Bangun Platform Machine Learning Perusahaan yang Aman di AWS (panduan AWS teknis)
- Kustomisasi Amazon SageMaker Studio menggunakan Konfigurasi Siklus Hidup (blog) AWS
- Membawa gambar kontainer kustom Anda sendiri ke notebook Amazon SageMaker Studio (AWSblog)
- Membangun Template SageMaker Proyek Kustom - Praktik Terbaik (AWSblog)
- Penerapan model multi-akun dengan Amazon SageMaker Pipelines (blog) AWS
- Bagian 1: Bagaimana NatWest Grup membangun platform MLOP yang terukur, aman, dan berkelanjutan (blog) AWS
- URL presigned Amazon SageMaker Studio yang aman Bagian 1: Infrastruktur dasar (blog) AWS

Kontributor

Kontributor dokumen ini meliputi:

- Ram Vittal, Arsitek Solusi ML, Amazon Web Services
- Sean Morgan, Arsitek Solusi ML, Amazon Web Services
- Durga Sury, Arsitek Solusi ML, Amazon Web Services

Terima kasih khusus kepada berikut ini yang menyumbangkan ide, revisi, dan perspektif:

- Alessandro Cerè, Arsitek Solusi AI/ML, Amazon Web Services
- Sumit Thakur, Pemimpin SageMaker Produk, Amazon Web Services
- Han Zhang, Sr. Insinyur Pengembangan Perangkat Lunak, Amazon Web Services
- Bhadrinath Pani, Insinyur Pengembangan Perangkat Lunak, Amazon Web Services, Amazon Web Services

Revisi dokumen

Untuk diberitahu tentang pembaruan pada whitepaper ini, berlangganan RSS feed.

Perubahan	Deskripsi	Tanggal
Whitepaper diperbarui	Tautan rusak diperbaiki dan banyak perubahan editorial di seluruh.	April 25, 2023
Publikasi awal	Whitepaper diterbitkan.	Oktober 19, 2022

Pemberitahuan

Pelanggan bertanggung jawab untuk membuat penilaian independen mereka sendiri atas informasi dalam dokumen ini. Dokumen ini: (a) hanya untuk tujuan informasi, (b) mewakili penawaran dan praktik AWS produk saat ini, yang dapat berubah tanpa pemberitahuan, dan (c) tidak membuat komitmen atau jaminan apa pun dari AWS dan afiliasinya, pemasok, atau pemberi lisensinya. AWSproduk atau layanan disediakan “sebagaimana adanya” tanpa jaminan, representasi, atau kondisi apa pun, baik tersurat maupun tersirat. Tanggung jawab dan kewajiban AWS kepada pelanggannya dikendalikan oleh AWS perjanjian, dan dokumen ini bukan bagian dari, juga tidak mengubah, perjanjian apa pun antara AWS dan pelanggannya.

AWSGlosarium

Untuk AWS terminologi terbaru, lihat AWSglosarium di Referensi. Glosarium AWS

Terjemahan disediakan oleh mesin penerjemah. Jika konten terjemahan yang diberikan bertentangan dengan versi bahasa Inggris aslinya, utamakan versi bahasa Inggris.