Opsi konektivitas jaringan AWS untuk penawaran SaaS

AWS Panduan Preskriptif

AWS Panduan Preskriptif: Opsi konektivitas jaringan AWS untuk penawaran SaaS

Table of Contents

Opsi konektivitas jaringan AWS untuk penawaran SaaS

Tomas Sykora dan Luca Schumann, Amazon Web Services

September 2025 (riwayat dokumen)

Panduan ini mengeksplorasi skenario umum untuk menghubungkan aplikasi konsumen ke penyedia perangkat lunak sebagai layanan (SaaS). Ini membahas cara terhubung ke sumber daya yang ada di tempat, di AWS Cloud, di cloud penyedia layanan cloud (CSP) lainnya, atau dalam arsitektur hibrida. Skenario ini meliputi:

- Mengekspos layanan web melalui HTTPS
- Mengekspos layanan berbasis TCP
- Menggunakan AWS AppSyncuntuk mengimplementasikan publish-subscribe (Pub/Sub) dan GraphQL APIs
- Menggunakan AWS sumber daya untuk mengekspos WebSockets aplikasi real-time
- Mengaktifkan akses bi-directional untuk komunikasi layanan interaktif

Dengan menyelaraskan dengan praktik terbaik yang tercakup dalam panduan ini, penyedia SaaS dapat mendorong kepercayaan pelanggan dan mendukung akses yang terukur, aman, dan tangguh ke penawaran SaaS.

Panduan ini juga mencakup kriteria penilaian diri untuk membantu Anda mengevaluasi seberapa sukses Anda memenuhi persyaratan jaringan konsumen untuk penawaran SaaS Anda. Di luar pola konektivitas, Anda akan menemukan perbandingan komprehensif layanan AWS jaringan, diagram arsitektur tingkat tinggi untuk berbagai skenario penyebaran, dan panduan praktis tentang cara memilih pendekatan yang tepat berdasarkan konteks bisnis spesifik Anda. Panduan ini mengeksplorasi pertimbangan keamanan untuk setiap opsi jaringan, membahas perangkap umum yang harus dihindari, dan memberikan rekomendasi implementasi yang menyeimbangkan persyaratan teknis dengan efisiensi operasional. Selain itu, Anda akan menemukan kerangka kerja strategis untuk menyelaraskan keputusan jaringan Anda dengan model bisnis Anda, tujuan pertumbuhan, dan kebutuhan kepatuhan peraturan.

Audiens yang dituju

Panduan ini ditujukan untuk penyedia SaaS. Ini membantu arsitek cloud, manajer produk, dan insinyur jaringan yang merancang, menerapkan, dan mengoptimalkan konektivitas jaringan untuk

penawaran SaaS di. AWS Cloud Untuk memahami konsep dan rekomendasi dalam panduan ini, Anda harus terbiasa dengan AWS dasar-dasar, konsep SaaS inti, dan prinsip-prinsip jaringan tingkat tinggi.

Tujuan

Panduan ini membahas opsi arsitektur jaringan dan praktik terbaik yang telah teruji di lapangan yang membantu konsumen mengoptimalkan akses ke penawaran SaaS. Menerapkan rekomendasi dalam panduan ini mendukung hal-hal berikut:

- Kemudahan integrasi - Berikan perjalanan pelanggan yang tidak rumit dari orientasi ke produksi sehingga Anda dapat mempercepat waktu pelanggan Anda untuk menilai dan mempersingkat siklus pengenalan pendapatan mereka.
- Adaptabilitas — Integrasi secara mulus dengan infrastruktur jaringan pelanggan Anda yang ada dengan beradaptasi dengan kebutuhan mereka yang terus berkembang. Ini meningkatkan proposisi nilai produk Anda.
- Total biaya kepemilikan — Standarisasi akses jaringan untuk mengurangi biaya perubahan dan biaya per penyewa. Dengan meningkatkan konsistensi penerapan, Anda juga dapat mengurangi waktu untuk melakukan analisis atau perbaikan akar penyebab.
- Manajemen ketergantungan — Memahami dependensi, implikasi jangka panjang, dan trade-off dari berbagai opsi akses jaringan. Ini membantu pemimpin produk membuat keputusan produk yang terinformasi dengan baik.
- Komposabilitas dan perluasan — Memisahkan pengembangan fungsionalitas inti dari infrastruktur operasional. Ini membantu tim pengembangan bergerak lebih cepat dan fokus pada menciptakan nilai bagi pelanggan Anda.
- Dorong kepercayaan - Dengan memberikan akses yang tangguh, toleran terhadap kesalahan, aman, dan terukur ke penawaran SaaS, Anda dapat mengurangi risiko peraturan dan mendapatkan kepercayaan pada kemampuan Anda untuk mendukung pertumbuhan pelanggan Anda.

Menilai keputusan akses jaringan untuk penawaran SaaS

Memahami pasar Anda

Keputusan yang Anda buat sekarang tentang jaringan menentukan apakah proposisi nilai produk SaaS Anda dapat dikirimkan ke pelanggan Anda. Terlepas dari kepentingan strategis dari keputusan ini, menyediakan akses ke penawaran SaaS Anda sering dianggap sebagai topik teknologi murni. Risiko yang dibawa persepsi ini mencakup siklus pengakuan pendapatan yang berkepanjangan, inefisiensi operasional, dan ketidaksejajaran dengan strategi bisnis. Misalnya, jika ekspansi cepat adalah tujuan bisnis strategis, maka cahaya panduan proses pengambilan keputusan Anda harus apakah solusi yang Anda pertimbangkan dapat diskalakan dan cukup fleksibel untuk mendukung ekspansi. Bahkan jika Anda berhasil mengembangkan bisnis Anda, overhead operasional tidak boleh menjadi penghalang bagi pertumbuhan masa depan, dan struktur biaya yang tidak selaras dapat menghabiskan semua keuntungan Anda.

Misalnya, pertimbangkan bagaimana pertimbangan pasar berikut mempengaruhi aspek teknis produk, seperti jaringan:

- Jika model bisnis Anda berbasis langganan, pelanggan Anda cenderung lebih memilih solusi dengan biaya berulang yang dapat diprediksi daripada investasi di muka yang besar.
- Jika strategi bisnis Anda menargetkan pelanggan tingkat perusahaan bernilai tinggi, maka kriteria keamanan, tata kelola, dan kepatuhan terhadap peraturan menentukan apakah penawaran SaaS Anda bahkan akan dipertimbangkan.
- Jika target pasar Anda sebagian besar adalah startup, kemudahan integrasi, waktu untuk menilai, dan kemampuan beradaptasi mungkin merupakan faktor penting. Startup biasanya memprioritaskan kecepatan dan kelincahan. Karena mereka perlu membangun merek dan perlu menghasilkan keuntungan dengan cepat, mereka cenderung lebih memilih solusi yang cepat dan mudah diintegrasikan, dapat meningkatkan skala biaya, mengurangi ketergantungan pada para ahli, dan tidak mengikat siklus berharga.
- Beberapa bisnis memerlukan akses yang stabil, throughput tinggi, dan latensi rendah. Ini termasuk industri hiburan dan media, manufaktur, dan pemrosesan transaksi keuangan. Jika ini adalah target pelanggan Anda, keandalan adalah perhatian utama mereka.

Dalam semua kasus ini, pelanggan mungkin merasakan penawaran SaaS yang sehat jika akses jaringan tidak mulus. Jika jaringan menjadi kendala, ini tidak mendukung kasus bisnis Anda. Jika

pelanggan Anda tidak dapat mengakses layanan yang Anda tawarkan dengan andal, proposisi nilai penawaran SaaS Anda adalah nol.

Memahami peran Anda

Peran Anda dalam mendukung tujuan bisnis tergantung pada siapa Anda, apa tujuan individu dan tim spesifik Anda, dan siapa pelanggan Anda, dan apa yang penting bagi mereka. Bahkan jika Anda bukan bagian dari tim yang biasanya berinteraksi dengan pelanggan, Anda harus peduli dengan siapa mereka dan apa yang mereka butuhkan. Tim teknik dan pengembangan juga harus peduli dengan pelanggan internal mereka, terutama mereka yang berinteraksi dengan mereka secara teratur. Biasanya, ini adalah operasi dan tim sukses pelanggan.

Jika Anda adalah bagian dari organisasi penjualan, penting bagi Anda untuk berkomunikasi dengan tim produk dan teknik tentang jaringan, meskipun itu adalah topik teknologi yang tampaknya murni. Bagikan wawasan tentang struktur target pasar. Komunikasikan poin rasa sakit dan kebutuhan pelanggan dan mitra Anda yang ada dan potensial. Bagikan data dan anekdot tentang peluang yang terlewatkan, prediksi pertumbuhan per segmen, dan peristiwa. Ajukan pertanyaan yang menantang kemampuan organisasi Anda untuk mendukung pertumbuhan bisnis. Ini meningkatkan jumlah peluang dan meningkatkan profitabilitas jangka panjang bisnis Anda. Pada akhirnya, ini membantu organisasi Anda mendanai ekspansi dan pengembangan masa depan.

Jika Anda adalah bagian dari organisasi teknik, pahami strategi bisnis organisasi Anda sebelum mencoba menyusun solusi. Penyelarasan dengan strategi bisnis membantu Anda memilih metrik yang tepat untuk mengevaluasi berbagai opsi akses jaringan. Ini juga dapat mencegah desain ulang jaringan skala besar yang mahal seiring pertumbuhan organisasi Anda. Penyelarasan bisnis membantu tim Anda mengamankan dan mempertahankan sumber daya yang diperlukan untuk tantangan masa depan. Jumlah karyawan tim Anda, anggaran untuk pengembangan profesional, atau akses ke teknologi mutakhir akan bergantung pada kemampuan Anda untuk menunjukkan keselarasan bisnis. Idealnya, Anda dapat menunjukkan bagaimana keputusan Anda berkontribusi pada keberhasilan bisnis organisasi. Oleh karena itu, kami menyarankan Anda menangkap proses pengambilan keputusan, termasuk kriteria pemilihan metrik. Tinjau metrik Anda secara berkala untuk mengonfirmasi bahwa metrik tersebut selaras dengan tujuan bisnis. Ini dapat membantu tim Anda mendapatkan kredit yang layak mereka dapatkan. Tinjauan berkala juga membantu memvalidasi bahwa tim Anda tidak membuat keputusan berdasarkan asumsi atau alasan historis yang usang.

Daftar metrik di bagian berikut ini relevan dengan akses jaringan:

- Metrik produk dan komersial

- Metrik rekayasa yang memengaruhi keputusan jaringan

Panduan ini menggunakan subset dari metrik ini untuk membantu Anda mengidentifikasi pendekatan akses jaringan yang optimal untuk penawaran SaaS Anda. Pilih metrik yang paling penting dan relevan dengan bisnis Anda, lalu evaluasi pendekatan berdasarkan metrik tersebut.

Metrik produk dan komersial yang memengaruhi keputusan jaringan

Tim produk dan komersial menggunakan kriteria keberhasilan untuk mengevaluasi apakah mereka memenuhi tujuan bisnis. Bagian ini menjelaskan metrik produk atau komersial yang dapat dipengaruhi secara positif atau negatif oleh keputusan akses jaringan yang dibuat organisasi Anda.

Gunakan metrik dan pertanyaan penilaian diri ini untuk mengevaluasi bagaimana pendekatan akses jaringan Anda selaras dengan posisi bisnis dan strategi pasar Anda. Penilaian ini membantu Anda menentukan apakah keputusan jaringan Anda saat ini mendukung diferensiasi pasar perusahaan Anda, keunggulan kompetitif, dan kebutuhan audiens target.

Bagian ini berisi metrik dan pertanyaan evaluasi diri untuk topik berikut:

- Model bisnis dan penentuan posisi pasar
- Total pasar yang dapat dialamatkan, tingkat akuisisi klien baru, pertumbuhan, dan skalabilitas
- Pengalaman dan retensi pelanggan
- Efisiensi dan kinerja keuangan
- Kepatuhan terhadap peraturan dan manajemen risiko
- Strategi mitra

Model bisnis dan penentuan posisi pasar

Metrik ini berhubungan dengan posisi perusahaan Anda di pasar, termasuk diferensiasi kompetitif, jangkauan pasar, dan persepsi merek. Sangat penting bahwa Anda menilai keselarasan antara pendekatan akses jaringan dan model bisnis. Lakukan penilaian terlepas dari apakah itu berbasis langganan, berbasis penggunaan, freemium, berjenjang, pasar, API pertama, atau berlabel putih. Pastikan bahwa model tersebut mendukung tujuan organisasi dan sasaran pelanggan.

Kriteria skor tinggi

Pendekatan akses jaringan secara mulus selaras dengan model bisnis. Ini memudahkan adopsi dan pengiriman layanan. Ini mendukung kelangsungan finansial jangka panjang dari model bisnis, dan struktur biaya kompatibel dengan pertumbuhan yang diharapkan. Ini meminimalkan gesekan bagi pelanggan atau mitra saat mengadopsi penawaran. Ini meningkatkan pengalaman pengguna dan mendorong penyerapan layanan yang lebih luas.

Indikator skor rendah

Pendekatan akses jaringan yang dipilih tidak selaras dengan model bisnis yang harus didukungnya. Struktur biaya dan waktu tunggu untuk penyebaran mewakili pemblokir untuk diadopsi di pasar sasaran. Infrastruktur yang sedang berlangsung dan biaya operasional menghambat potensi keuntungan. Ini mencegah pertumbuhan bisnis dan membuatnya sulit untuk beroperasi pada skala yang diinginkan. Atau, properti pendekatan akses jaringan dapat mencegah pelanggan mempertimbangkan layanan karena alasan peraturan.

Pertanyaan penilaian diri

- Apa implikasi biaya dari pendekatan akses jaringan yang dipilih untuk penyebaran awal dan pengiriman berkelanjutan? Berapa biaya tetap dan variabel dari pendekatan ini?
- Dapatkah pendekatan akses jaringan skala efektif dan efisien untuk memenuhi tuntutan pertumbuhan model bisnis? Pertimbangkan ukuran penyewa individu dan jumlah penyewa onboard.
- Apakah pendekatan akses jaringan memaksakan batasan teknis atau operasional yang dapat membatasi fleksibilitas atau kemampuan beradaptasi model bisnis?
- Untuk pendekatan akses jaringan, bagaimana lead time penyebaran selaras dengan kecepatan ke pasar yang dibutuhkan model bisnis?

Total pasar yang dapat dialamatkan, tingkat akuisisi klien baru, pertumbuhan, dan skalabilitas

Sangat penting bahwa Anda menilai dampak dari keputusan jaringan pada kapasitas organisasi untuk memperluas ke pasar baru, memperoleh pelanggan secara efektif, dan mempertahankan skalabilitas operasional. Faktor-faktor ini mempengaruhi tingkat konversi. Mereka juga mempengaruhi apakah pendekatan akses jaringan mendukung ekspansi ke segmen pasar yang signifikan atau membatasi Anda untuk hanya melayani jenis pelanggan tertentu.

Kriteria skor tinggi

Pendekatan akses jaringan membantu organisasi untuk mencapai sebagian besar target pasar, atau dapat secara efektif dikombinasikan dengan pendekatan jaringan lain untuk memperluas jangkauan pasar. Pendekatan ini harus membutuhkan upaya integrasi tambahan minimal. Pendekatan ini mendukung waktu tunggu yang singkat untuk penyebaran, entri pasar yang cepat, dan ekspansi. Hal ini memungkinkan untuk sejumlah besar penyebaran paralel. Integrasi sangat mudah bagi pelanggan, yang menurunkan hambatan adopsi dan meningkatkan pengalaman pelanggan. Pendekatan ini meminimalkan overhead operasional, menjaga kapasitas operasional, dan mendukung proyeksi pertumbuhan.

Indikator skor rendah

Pendekatan akses jaringan hanya mendukung sebagian kecil dari target pasar atau cocok terutama untuk segmen niche yang tidak diprioritaskan dalam strategi bisnis. Ini tidak secara efektif melengkapi pendekatan akses jaringan lain yang sudah didukung. Waktu tunggu untuk permintaan pasar lag penerapan, yang membatasi ekspansi pasar dan akuisisi klien baru. Model penyebaran bersifat berurutan, yang meningkatkan risiko kemacetan layanan seiring dengan meningkatnya permintaan. Proses integrasi yang kompleks menghalangi klien potensial, yang berdampak negatif pada tingkat akuisisi dan tingkat konversi. Overhead operasional yang signifikan mengurangi kapasitas operasional organisasi. Ini menjadi penghambat pertumbuhan yang diproyeksikan.

Untuk indikator ini, evaluasi apakah memperkenalkan pendekatan akses jaringan baru dapat membantu organisasi mencapai tujuan bisnis strategisnya. Pertimbangkan apakah pendekatan akses jaringan baru dapat menciptakan dependensi produk baru atau mengkonsumsi sumber daya operasional tanpa memberikan hasil yang diinginkan.

Pertanyaan penilaian diri

- Apakah ada kesenjangan dalam pendekatan saat ini yang mencegah Anda menjangkau segmen yang lebih besar dari target pasar?
- Berapa set minimum pendekatan akses jaringan yang tidak tumpang tindih, standar, yang harus Anda dukung untuk mencakup 70-90% dari target pasar?
- Jangkauan apa yang dimungkinkan oleh setiap pendekatan akses jaringan, dan apa peningkatan terkait dalam metrik penting, seperti biaya infrastruktur, siklus operasional, dan ketergantungan pada para ahli?
- Bagaimana kemampuan penyebaran dan batas layanan infrastruktur jaringan selaras dengan ekspektasi pertumbuhan di pasar target Anda?

- Apakah integrasi jaringan menciptakan hambatan untuk masuk bagi pelanggan baru? Bagaimana ini dapat diatasi untuk meningkatkan tingkat konversi?
- Bagaimana overhead operasional pengelolaan jaringan memengaruhi kapasitas Anda untuk pertumbuhan dan skalabilitas?
- Strategi apa yang dapat Anda terapkan untuk mengurangi waktu tunggu untuk penyebaran jaringan dan meningkatkan ekspansi pasar dan akuisisi pelanggan?
- Apakah ada ketergantungan pada sumber daya ahli yang akan menunda penyebaran atau integrasi dengan ekosistem pelanggan?

Pengalaman dan retensi pelanggan

Metrik di bagian ini membantu Anda memahami kemampuan organisasi Anda untuk memperoleh dan, yang paling penting, mempertahankan pelanggan. Memahami hubungan antara pendekatan akses jaringan dan kepuasan pelanggan dapat membantu tim produk dan teknik membuat keputusan yang diinformasikan oleh data.

Kriteria skor tinggi

Pendekatan akses jaringan dapat diandalkan dan mudah dikelola. Ini berkontribusi pada kepuasan pelanggan yang tinggi (CSAT) dan hasil skor promotor bersih (NPS). Skor ini menunjukkan reputasi merek yang kuat dan loyalitas pelanggan. Berkat integrasi yang mulus dengan ekosistem pelanggan Anda yang ada, gesekan adopsi rendah, dan ada ketergantungan yang rendah pada para ahli. Organisasi Anda secara konsisten memenuhi perjanjian tingkat layanan (SLAs), yang memperkuat kepercayaan pelanggan dan kewajiban kontrak. Karena pelanggan menikmati layanan yang stabil dan dapat diandalkan, Anda memiliki retensi pelanggan yang tinggi.

Indikator skor rendah

Integrasi yang sulit dan akses yang tidak konsisten ke layanan biasanya menyebabkan frustrasi pelanggan dan umpan balik negatif. Ini merusak reputasi merek. Pelanggan baru gagal mengonversi dari paket gratis atau uji coba ke layanan berbayar karena ketergantungan pada para ahli atau karena waktu orientasi dan integrasi yang berkepanjangan. Kegagalan yang sering terjadi untuk memenuhi SLAs mengakibatkan hukuman finansial dan hilangnya kredibilitas, berpotensi mengurangi tingkat retensi pelanggan.

Pertanyaan penilaian diri

- Bagaimana kinerja jaringan (seperti kecepatan, uptime, dan latensi) secara langsung mempengaruhi hasil CSAT dan NPS? Peningkatan jaringan spesifik apa yang dapat mendorong skor ini lebih tinggi?
- Bagaimana latensi jaringan dan metrik uptime saat ini memengaruhi pengalaman pengguna awal dan tingkat adopsi? Peningkatan kinerja jaringan spesifik apa yang diperlukan untuk mengoptimalkan metrik ini?
- Apakah ada masalah berulang dalam konfigurasi jaringan atau pengaturan keamanan yang mempersulit integrasi untuk pelanggan baru? Bagaimana Anda bisa merampingkan proses ini?
- Bagaimana kemudahan mengonfigurasi akses jaringan memengaruhi pengalaman orientasi bagi pengguna baru? Apakah ada titik akses jaringan tertentu atau waktu tunggu yang dapat dioptimalkan untuk meningkatkan tayangan pengguna awal?
- Apa tantangan untuk mengotomatisasi penyediaan layanan jaringan untuk klien baru. Bagaimana Anda bisa menyesuaikan proses ini untuk meningkatkan skalabilitas dan keandalan?
- Analisis akar penyebab pelanggaran SLA baru-baru ini. Apakah mereka terkait dengan konfigurasi jaringan, perencanaan kapasitas, atau masalah vendor eksternal?
- Seberapa sering masalah jaringan menyebabkan Anda melewatkan komitmen SLA? Apa kegagalan terkait jaringan yang paling sering terjadi?
- Peningkatan kinerja jaringan mana yang menunjukkan dampak positif paling signifikan pada kepuasan pelanggan di masa lalu?

Efisiensi dan kinerja keuangan

Kategori ini menilai aspek kesehatan keuangan dan profitabilitas bisnis Anda, seperti efisiensi biaya, kelayakan jangka panjang, profitabilitas, laba atas investasi (ROI), dan total biaya kepemilikan (TCO). Dengan merampingkan operasi jaringan melalui standardisasi, Anda dapat mengurangi biaya overhead dan pemeliharaan operasional. Ini mendukung tujuan pertumbuhan organisasi Anda.

Kriteria skor tinggi

Struktur biaya dari pendekatan akses jaringan selaras dengan model bisnis. Ini mendukung pertumbuhan yang berkelanjutan, dan penghematan biaya yang signifikan yang Anda capai meningkatkan profitabilitas. Akses jaringan yang efisien memungkinkan orientasi pelanggan yang cepat, yang mempersingkat waktu untuk memberikan nilai dan mempercepat penetrasi pasar. Ini secara langsung mempersingkat siklus pengakuan pendapatan.

Indikator skor rendah

Pelanggan beralih ke pesaing Anda untuk mempercepat pengiriman aplikasi dan layanan mereka. Organisasi Anda telah meningkatkan biaya operasional yang terkait dengan konfigurasi jaringan yang kompleks dan beragam serta waktu tunggu yang diperpanjang. Struktur biaya dan model bisnis tidak selaras, yang dapat menyebabkan biaya dimuka yang tinggi untuk layanan berbasis langganan. Proses orientasi yang rumit mengurangi penetrasi pasar dan menunda pengakuan pendapatan.

Pertanyaan penilaian diri

- Berapa waktu tunggu saat ini untuk penyebaran layanan baru, dan bagaimana pengaruhnya terhadap waktu ke pasar dan pengakuan pendapatan?
- Seberapa efektif operasi jaringan standar mengurangi biaya overhead dan pemeliharaan?
- Apakah sumber daya ahli diperlukan untuk berhasil menyelesaikan integrasi awal, beroperasi setiap hari, memecahkan masalah, atau menerapkan perubahan?
- Seberapa berkelanjutan investasi jaringan saat ini dalam hal kemajuan teknologi? Apakah Anda berinvestasi dalam teknologi masa depan yang selaras dengan perkembangan pasar yang diproyeksikan?
- Seberapa efektif Anda mengalokasikan dan melacak biaya yang terkait dengan lalu lintas jaringan dan penggunaan oleh penyewa individu?

Kepatuhan terhadap peraturan dan manajemen risiko

Sangat penting untuk memvalidasi kepatuhan terhadap peraturan terkait jaringan. Ini menegaskan bahwa Anda beroperasi secara legal dan dapat menjaga kepercayaan pelanggan. Standardisasi di seluruh operasi jaringan menyederhanakan proses kepatuhan dan mempromosikan konsistensi di berbagai yurisdiksi dan geografi. Langkah-langkah ini membantu Anda memperluas layanan Anda.

Kriteria skor tinggi

Operasi jaringan secara konsisten mematuhi standar hukum tanpa komplikasi, yang berkontribusi pada ekspansi pasar, mengurangi gesekan adopsi, dan meningkatkan kepercayaan pelanggan. Kepatuhan yang ditunjukkan dengan kerangka peraturan penting, seperti Digital Operational Resilience Act (DORA) dan National Institute of Standards and Technology (NIST), membantu Anda memenangkan pelanggan yang sensitif terhadap kepatuhan peraturan. Visibilitas berkelanjutan ke status kepatuhan Anda mengurangi waktu yang diperlukan untuk menyelesaikan audit.

Indikator skor rendah

Kesenjangan dalam kepatuhan jaringan menyebabkan gesekan adopsi yang tinggi, penundaan peluncuran layanan, tantangan hukum, dan potensi denda. Tantangan-tantangan ini menyebabkan rencana ekspansi yang tertunda atau dibatalkan ke pasar baru. Sulit untuk mempertahankan praktik kepatuhan standar di berbagai yurisdiksi, dan ini memengaruhi efisiensi operasional dan reputasi pasar.

Pertanyaan penilaian diri

- Seberapa baik operasi jaringan Anda selaras dengan peraturan atau pedoman industri yang berlaku? Apa yang diungkapkan audit kepatuhan Anda baru-baru ini?
- Bagaimana Anda menjaga kepatuhan terhadap peraturan yang muncul di bidang keamanan digital dan jaringan?
- Seberapa efektif proses dokumentasi dan pelaporan Anda dalam memenuhi persyaratan badan pengatur yang berbeda?
- Strategi manajemen risiko apa yang Anda miliki untuk mengidentifikasi dan mengatasi potensi risiko kepatuhan sebelum menimbulkan tantangan hukum?
- Tingkat pelatihan kepatuhan dan kesadaran apa yang dibutuhkan tim manajemen jaringan Anda untuk mendukung pendekatan akses jaringan Anda?

Strategi mitra

Menilai seberapa baik pendekatan akses jaringan selaras dengan ekosistem mitra, platform, dan pasar yang diakui. Ini penting, terutama jika strategi pertumbuhan Anda bergantung pada penskalaan melalui mitra.

Kriteria skor tinggi

Pendekatan akses jaringan terintegrasi di seluruh ekosistem mitra Anda. Struktur biayanya selaras dengan model bisnis mitra utama Anda. Mitra memiliki keterampilan jaringan yang diperlukan untuk integrasi yang mulus dari penawaran SaaS Anda, dan mereka dapat memberikan akses dan fungsionalitas yang berkelanjutan.

Indikator skor rendah

Pendekatan akses jaringan yang dipilih menuntut keterampilan khusus, sumber daya, atau peralatan yang langka atau sulit diperoleh. Ini berbeda dari protokol akses jaringan standar yang biasa

digunakan oleh platform dan pasar. Ini menghasilkan struktur biaya yang tidak dapat diprediksi yang menantang untuk didamaikan. Pendekatan akses jaringan tidak selaras dengan model bisnis mitra utama Anda.

Pertanyaan penilaian diri

- Apa implikasi biaya dari pendekatan akses jaringan untuk mitra. Bagaimana biaya-biaya ini selaras dengan model bisnis mereka? Sisi integrasi mana yang menanggung sebagian besar biaya, dan berapa banyak siklus operasional yang harus diinvestasikan?
- Untuk pendekatan akses jaringan, apakah ada hambatan untuk integrasi atau pemeliharaan yang dapat memengaruhi hubungan mitra atau skalabilitas ekosistem?
- Bagaimana pendekatan akses jaringan dapat dioptimalkan untuk meningkatkan kompatibilitas dan kemudahan integrasi di seluruh ekosistem?

Metrik rekayasa yang memengaruhi keputusan jaringan

Seperti tim produk dan komersial, tim teknik juga menggunakan kriteria keberhasilan untuk mengevaluasi apakah mereka memenuhi tujuan bisnis. Namun, metrik ini berbeda dan mereka fokus pada kemampuan tim untuk mengembangkan, mengoperasikan, dan memenuhi persyaratan keamanan dan kepatuhan. Bagian ini menjelaskan metrik teknik yang dapat dipengaruhi secara positif atau negatif oleh keputusan akses jaringan yang dibuat organisasi Anda.

Gunakan metrik dan pertanyaan penilaian diri ini untuk mengevaluasi pendekatan akses jaringan Anda saat ini terhadap persyaratan bisnis dan kemampuan teknis Anda. Penilaian ini membantu Anda mengidentifikasi kesenjangan dalam arsitektur Anda dan memprioritaskan perbaikan yang selaras dengan tujuan strategis Anda. Dengan meninjau kriteria ini secara teratur, Anda dapat memastikan bahwa strategi akses jaringan Anda terus mendukung kebutuhan pelanggan dan rencana pertumbuhan organisasi Anda.

Bagian ini berisi metrik dan pertanyaan evaluasi diri untuk kategori dan topik berikut:

- Metrik pengembangan
  - Frekuensi penyebaran, waktu untuk menyebarkan, dan kecepatan sprint
  - Fleksibilitas dan pengiriman fitur
  - Ubah tingkat kegagalan
  - Kualitas kode dan kinerja tim teknik
  - Pengurangan utang teknis

Metrik pengembangan yang terkait dengan akses jaringan untuk penawaran SaaS

Bagian ini berisi metrik berikut:

- Frekuensi penyebaran, waktu untuk menyebarkan, dan kecepatan sprint
- Fleksibilitas dan pengiriman fitur
- Ubah tingkat kegagalan
- Kualitas kode dan kinerja tim teknik
- Pengurangan utang teknis
- Skalabilitas, kapasitas, dan kinerja

Frekuensi penyebaran, waktu untuk menyebarkan, dan kecepatan sprint

Untuk mengoptimalkan efisiensi siklus pengembangan, penting bagi Anda untuk memahami pengaruh penyediaan tumpukan jaringan pada kecepatan sprint.

Kriteria skor tinggi

Penyediaan tumpukan jaringan disederhanakan dan otomatis, dan memerlukan intervensi manual minimal. Itu tidak secara signifikan mempengaruhi kecepatan sprint. Penyediaan dan pemindahan tumpukan jaringan dapat dilakukan oleh anggota tim mana pun. Ini mengurangi kemacetan dan ketergantungan pada sumber daya khusus.

Indikator skor rendah

Sejumlah besar poin cerita diperlukan untuk penyediaan tumpukan jaringan. Ini menunjukkan proses yang kompleks dan memakan waktu yang mengurangi pengembangan fitur baru. Pemindahan tumpukan jaringan yang sering menimbulkan biaya overhead dan waktu yang cukup besar. Tugas

penyediaan jaringan memerlukan keahlian teknik khusus, yang menciptakan kemacetan dan memperlambat siklus pengembangan.

Pertanyaan penilaian diri

- Langkah manual apa, jika ada, yang terlibat dalam proses penyebaran. Bagaimana pengaruhnya terhadap frekuensi dan waktu penyebaran?
- Bagaimana rollback ditangani jika terjadi kegagalan penerapan. Apa dampaknya terhadap frekuensi penyebaran dan waktu pemulihan?
- Berapa banyak poin cerita yang diperlukan untuk menyediakan tumpukan jaringan saat Anda menyiapkan lingkungan baru?
- Berapa banyak biaya tambahan dan overhead waktu yang terkait dengan seringnya pemindahan tumpukan jaringan selama proses pengembangan?
- Apakah penyediaan tumpukan jaringan bergantung pada keahlian teknik khusus, atau apakah itu tugas yang dapat dikelola oleh anggota tim mana pun?

Fleksibilitas dan pengiriman fitur

Pendekatan akses jaringan dapat memengaruhi kemampuan tim teknik untuk berinovasi dan menerapkan fitur baru secara efisien.

Kriteria skor tinggi

Pendekatan akses jaringan menawarkan fleksibilitas yang dibutuhkan untuk penyebaran fitur yang cepat dan mulus. Ini mendukung berbagai protokol komunikasi, komunikasi searah dan dua arah, dan ukuran pesan. Ini tidak memaksakan kendala yang signifikan pada proses pengembangan atau inovasi.

Indikator skor rendah

Pendekatan akses jaringan membatasi kemampuan tim untuk meluncurkan fitur baru karena kurangnya protokol komunikasi yang didukung, ketidakfleksibelan dalam ukuran pesan, atau ketergantungan pada teknologi tertentu dan sumber daya ahli terkait. Hal ini dapat menyebabkan siklus pengembangan lebih lambat dan menghambat evolusi layanan.

Pertanyaan penilaian diri

- Bagaimana pendekatan akses jaringan memengaruhi kelincahan tim dalam mengembangkan dan menerapkan fitur baru?

- Apakah ada batasan dalam pendekatan akses jaringan yang membatasi dukungan protokol atau teknologi komunikasi tertentu?
- Bagaimana pendekatan memfasilitasi atau membatasi integrasi teknologi dan inovasi baru ke dalam layanan?
- Bagaimana pendekatan akses jaringan mempengaruhi jadwal pengembangan dan peta jalan produk?

Ubah tingkat kegagalan

Pendekatan akses jaringan yang Anda pilih dapat memengaruhi tingkat kegagalan perubahan saat menerapkan layanan atau fitur baru. Kontrol yang lebih besar sering berarti fleksibilitas yang lebih besar, tetapi juga meningkatkan potensi kesalahan konfigurasi, seperti ketika mengelola pengaturan perutean yang kompleks.

Kriteria skor tinggi

Anda dapat menerapkan perubahan pada tumpukan jaringan dengan risiko kegagalan minimal. Ada mekanisme pengujian yang memadai, mekanisme rollback yang efisien ada, dan pemantauan yang efektif membantu Anda mengidentifikasi dan menyelesaikan masalah dengan cepat.

Indikator skor rendah

Pendekatan akses jaringan rentan terhadap kegagalan selama perubahan. Ada opsi pengujian terbatas, strategi penyebaran yang rumit, atau kemampuan pemantauan dan pemecahan masalah yang tidak memadai. Beberapa pihak diharuskan untuk berpartisipasi dalam sesi pemecahan masalah. Hal ini dapat menyebabkan peningkatan downtime dan mengurangi ketersediaan penawaran SaaS.

Pertanyaan penilaian diri

- Tindakan apa yang dilakukan untuk mengurangi risiko kegagalan perubahan saat memperbarui tumpukan jaringan?
- Apakah ada proses pengujian dan validasi menyeluruh?
- Seberapa cepat sistem dapat pulih dari perubahan yang gagal? Apakah ada proses rollback yang efisien?
- Apakah ada sistem pemantauan dan peringatan proaktif untuk mendeteksi dan mengatasi masalah dengan cepat selama dan setelah perubahan tumpukan jaringan?

- Berapa tingkat kegagalan perubahan historis untuk penerapan tumpukan jaringan. Pelajaran apa yang telah dipetik dari insiden masa lalu?
- Bagaimana pendekatan akses jaringan memfasilitasi atau membatasi implementasi perubahan. Apakah pendekatan meminimalkan gangguan layanan?
- Apa risiko memengaruhi ketersediaan penawaran SaaS di lingkungan produksi ketika Anda menerapkan perubahan yang melibatkan pendekatan akses jaringan?

Kualitas kode dan kinerja tim teknik

Pendekatan akses jaringan secara tidak langsung dapat mempengaruhi kualitas kode untuk penawaran SaaS. Kurangnya standardisasi dalam akses jaringan dapat memaksa tim teknik untuk mendukung beberapa pendekatan integrasi, yang dapat menyebabkan basis kode membengkak. Hal ini, pada gilirannya, dapat menghambat kemampuan tim untuk mengembangkan kedalaman dan kontrol atas kualitas kode yang diperlukan untuk mempertahankan tim teknik berkinerja tinggi.

Kriteria skor tinggi

Tim teknik tetap fokus berkat modularitas kode dan penggunaan kembali di seluruh pendekatan akses jaringan yang didukung. Pendekatan akses jaringan kompatibel dengan pipeline penyebaran yang ada dan strategi pengujian otomatis.

Indikator skor rendah

Kinerja tim teknik berkurang karena overhead yang terkait dengan integrasi dan pemeliharaan terlalu banyak pendekatan akses jaringan. Beberapa pendekatan secara signifikan meningkatkan kompleksitas, menghasilkan utang teknologi, atau memerlukan pengembangan solusi untuk mengatasi kemampuan yang hilang atau tidak mencukupi.

Pertanyaan penilaian diri

- Bagaimana pendekatan akses jaringan mengelola variabilitas jaringan?
- Apakah Anda perlu mengembangkan kode tambahan untuk menangani gangguan konektivitas?
- Apakah pendekatan akses jaringan baru terintegrasi secara mulus dengan pendekatan yang ada, atau apakah itu memerlukan pengembangan khusus yang signifikan?
- Sejauh mana perubahan yang diperlukan untuk mengadopsi pendekatan akses jaringan baru? Dapatkah basis kode yang ada dan pengujian otomatis digunakan secara efektif?

- Seberapa mudah atau sulitnya untuk menyebarkan atau menerapkan kembali layanan dengan pendekatan akses jaringan yang dipilih? Bisakah ini sering dilakukan? Apakah ada ketergantungan pada sumber daya ahli?
- Apakah pendekatan akses jaringan memfasilitasi atau mempersulit kepatuhan terhadap standar pengkodean dan praktik terbaik?
- Bagaimana pendekatan mempengaruhi fitur atau perbaikan baru? time-to-market

Pengurangan utang teknis

Evaluasi dampak pendekatan akses jaringan pada utang teknis harus mempertimbangkan skalabilitas, observabilitas, dan kemampuan keamanannya.

Kriteria skor tinggi

Pendekatan ini secara efektif merampingkan manajemen infrastruktur saat basis pelanggan berkembang. Ini menawarkan kemampuan observabilitas out-of-the-box yang kuat. Ini mendorong pemantauan dan pemeliharaan yang efisien.

Indikator skor rendah

Pendekatan akses jaringan tidak cukup mengamankan saluran komunikasi dan tidak memiliki alat yang cukup untuk pengamatan metrik kualitatif. Mungkin juga memerlukan pengembangan tambahan untuk manajemen infrastruktur karena basis pelanggan meningkat, atau mungkin memerlukan solusi untuk masalah keandalan.

Pertanyaan penilaian diri

- Bagaimana pendekatan akses jaringan mempengaruhi skalabilitas infrastruktur jangka panjang? Apakah ini memfasilitasi pertumbuhan yang mulus dengan investasi tambahan minimal?
- Seberapa komprehensif alat observabilitas yang disertakan? Apakah mereka memungkinkan pemantauan proaktif dan penyelesaian masalah?
- Apa dampak yang diantisipasi dari pendekatan akses jaringan pada pemeliharaan dan evolusi basis kode dari waktu ke waktu?
- Apakah pendekatan tersebut terintegrasi dengan baik dengan infrastruktur yang ada dan yang direncanakan. Apakah itu membutuhkan perubahan atau penambahan yang signifikan?

Skalabilitas, kapasitas, dan kinerja

Untuk menentukan kesesuaian pendekatan akses jaringan untuk penawaran SaaS, penting untuk menganalisis bagaimana ia mempertahankan kinerja optimal seiring dengan meningkatnya permintaan.

Kriteria skor tinggi

Pendekatan akses jaringan memfasilitasi ekspansi dengan mulus. Ini mempertahankan latensi rendah selama pemrosesan permintaan, dan secara efisien menangani lonjakan lalu lintas. Ini memberikan kinerja yang konsisten terlepas dari peningkatan tingkat lalu lintas, dan itu tidak memaksakan batasan operasional pada pertumbuhan.

Indikator skor rendah

Pendekatan akses jaringan tidak menskalakan secara efektif, mungkin karena keterbatasan bandwidth yang melekat atau kapasitas infrastruktur yang tidak mencukupi. Penyediaan dan manajemen sumber daya meningkatkan kompleksitas atau membuat dependensi. Kinerja layanan menurun karena peningkatan latensi, jitter, dan variabilitas throughput, terutama dalam kondisi jaringan yang padat.

Pertanyaan penilaian diri

- Bagaimana pendekatan akses jaringan mengakomodasi peningkatan jumlah penyewa dan volume data mereka?
- Apakah secara inheren dapat diskalakan untuk memenuhi tuntutan masa depan?
- Langkah-langkah apa yang ada untuk memastikan kinerja konsisten, bahkan selama periode lalu lintas puncak atau peristiwa penskalaan cepat?
- Bagaimana pendekatan ini menangani latensi jaringan dan jitter? Apakah ada mekanisme untuk mengoptimalkan throughput data dan meminimalkan penundaan?
- Dapatkah pendekatan akses jaringan beradaptasi dengan berbagai kondisi jaringan? Bisakah itu memberikan pengalaman penyewa tunggal untuk setiap pelanggan?
- Apa dampak dari pendekatan akses jaringan pada infrastruktur yang mendasarinya? Apakah itu memerlukan peningkatan atau perubahan signifikan pada sistem yang ada?

Metrik keunggulan operasional yang terkait dengan akses jaringan untuk penawaran SaaS

Bagian ini berisi metrik berikut:

- Ketahanan operasional dan pemulihan bencana
- Pemantauan kinerja layanan dan aplikasi

Ketahanan operasional dan pemulihan bencana

Pendekatan akses jaringan akan membantu penawaran SaaS menahan berbagai jenis gangguan dan dengan cepat pulih dari bencana apa pun.

Kriteria skor tinggi

Rencana pemulihan bencana yang ditetapkan dan diuji secara konsisten menunjukkan bahwa pendekatan akses jaringan memenuhi persyaratan pemulihan bencana. Pendekatan akses jaringan mendukung konfigurasi ketersediaan tinggi, dan mendukung mekanisme failover otomatis, cepat, dan andal.

Indikator skor rendah

Pendekatan akses jaringan membuat sulit untuk membangun strategi pemulihan bencana yang koheren. Anda mengamati waktu pemulihan yang berkepanjangan setelah gangguan. Kegagalan operasional yang sering terjadi pada infrastruktur jaringan berdampak pada penyampaian layanan.

Pertanyaan penilaian diri

- Kapan latihan pemulihan bencana terakhir, dan apa hasilnya?
- Berapa lama waktu yang dibutuhkan untuk memulihkan layanan penting setelah gangguan? Bagian mana dari infrastruktur jaringan yang perlu dipindahkan?
- Perbaikan apa yang dapat dilakukan pada infrastruktur jaringan untuk merampingkan rencana pemulihan bencana Anda?
- Apakah ada redundansi untuk komponen jaringan yang paling penting?
- Sudahkah Anda mengotomatiskan potensi pemindahan infrastruktur jaringan setelah pemadaman kritis?
- Bagaimana pendekatan akses jaringan mendukung toleransi dan keandalan kesalahan? Apakah ada mekanisme bawaan untuk menangani gangguan jaringan dan menjaga integritas data?

Pemantauan kinerja layanan dan aplikasi

Pendekatan akses jaringan dapat mempengaruhi alat pemantauan kinerja yang digunakan untuk memvalidasi operasi dan uptime layanan yang optimal. Bergantung pada layanannya, Anda mungkin memiliki akses ke metrik tingkat rendah (seperti tingkat penurunan paket) atau metrik tingkat yang lebih tinggi (seperti durasi sesi). Metrik tingkat rendah memberikan wawasan teknis terperinci tentang perilaku jaringan tetapi bisa rumit untuk ditafsirkan. Sebaliknya, metrik tingkat yang lebih tinggi sering menawarkan cara yang lebih langsung dan lebih mudah untuk mengukur pengalaman pengguna secara keseluruhan. Ini karena mereka menggabungkan dampak kondisi jaringan yang mendasarinya menjadi indikator kualitas layanan yang jelas.

Kriteria skor tinggi

Alat pemantauan komprehensif yang memberikan wawasan mendekati waktu nyata sudah tersedia. Anda memiliki sistem peringatan dan respons otomatis yang mengatasi masalah kinerja. Anda dapat memprediksi potensi kemacetan atau kegagalan layanan sebelum memengaruhi pengguna.

Indikator skor rendah

Gangguan layanan atau masalah kinerja yang sering terjadi tanpa diamati atau ditindaklanjuti. Kurangnya visibilitas ke kinerja layanan menghasilkan respons yang lambat terhadap kemacetan kinerja. Tim multi-pihak diperlukan untuk memecahkan masalah infrastruktur jaringan.

Pertanyaan penilaian diri

- Alat pemantauan dan metrik infrastruktur jaringan mana yang saat ini tersedia? Seberapa efektif mereka dalam mendeteksi anomali layanan?
- Seberapa cepat Anda dapat mengidentifikasi dan menyelesaikan masalah kinerja?
- Apakah Anda memiliki mekanisme yang memprediksi potensi masalah kinerja?
- Perbaikan apa yang dapat Anda lakukan untuk meningkatkan kemampuan observabilitas?

Metrik keamanan dan tata kelola yang terkait dengan akses jaringan untuk penawaran SaaS

Bagian ini berisi metrik berikut:

- Keamanan, kepatuhan, dan manajemen kerentanan

Keamanan, kepatuhan, dan manajemen kerentanan

Sangat penting bagi Anda untuk mengevaluasi aspek keamanan dari pendekatan akses jaringan, termasuk kepatuhan terhadap standar keamanan dan pengelolaan kerentanan.

Kriteria skor tinggi

Pendekatan akses jaringan membantu tim Anda mematuhi kerangka kerja keamanan, seperti Organisasi Internasional untuk Standardisasi (ISO) 27001, Kontrol Sistem dan Organisasi 2 (SOC 2), atau NIST. Itu membuatnya mudah untuk melakukan audit keamanan reguler. Enkripsi yang kuat dan mekanisme otentikasi sudah ada. Jaringan terisolasi, dan hanya sumber daya yang diperlukan yang terpapar pada infrastruktur pelanggan. Anda dapat melihat anomali jaringan dalam waktu dekat, tanpa overhead yang berlebihan.

Indikator skor rendah

Pendekatan akses jaringan rentan terhadap pelanggaran keamanan berulang atau kerentanan, dan tidak sesuai dengan standar keamanan utama. Anda sering mengamati deteksi tertunda dan respons terhadap insiden keamanan.

Pertanyaan penilaian diri

- Apakah ada pelanggaran keamanan baru-baru ini yang terkait dengan pendekatan akses jaringan yang dipilih, dan apa yang telah kita pelajari dari mereka?
- Bagaimana pendekatan akses jaringan Anda sesuai dengan standar keamanan global?
- Berapa lama waktu yang dibutuhkan untuk mendeteksi dan menanggapi ancaman keamanan? Bagaimana akses jaringan membantu atau membatasi kemampuan ini?
- Seberapa sering penilaian keamanan dilakukan pada pendekatan akses jaringan? Dapatkah Anda menggunakan perkakas umum untuk menilai keamanan pendekatan akses jaringan, atau apakah perangkat lunak khusus diperlukan?
- Tingkat keamanan apa yang melekat dalam pendekatan akses jaringan, dan bagaimana hal itu selaras dengan praktik terbaik industri dan persyaratan peraturan?

Ikhtisar layanan AWS jaringan untuk penawaran SaaS

Bagian ini membahas layanan AWS jaringan yang dirujuk dalam panduan ini. Ini juga membandingkan kemampuan mereka dan menjelaskan pertimbangan keamanan untuk setiap layanan.

Bagian ini berisi topik berikut:

- AWS layanan jaringan
- Membandingkan kemampuan layanan
- Fitur dan pertimbangan keamanan

AWS layanan jaringan

Berikut ini adalah hal-hal Layanan AWS yang dibahas secara konsisten dalam panduan ini.

AWS PrivateLink

AWS PrivateLinkadalah layanan cloud-native yang dapat menyediakan akses ke penawaran SaaS Anda jika pelanggan Anda sudah beroperasi di. AWS Cloud Pelanggan Anda terhubung ke penawaran SaaS melalui titik akhir VPC antarmuka. Ini adalah antarmuka jaringan endpoint yang disediakan dalam satu atau lebih subnet di pelanggan. Akun AWS Dalam skenario dalam panduan ini, lalu lintas berjalan melalui titik akhir VPC antarmuka dan tiba di Network Load Balancer di akun Anda. Network Load Balancer meneruskan lalu lintas ke aplikasi SaaS, yang telah Anda daftarkan sebagai layanan endpoint. Melalui titik akhir VPC sumber daya, juga AWS PrivateLink dapat membantu Anda mengakses sumber daya lain, seperti database.

Kisi VPC Amazon

Amazon VPC Lattice adalah layanan jaringan aplikasi yang membantu penyedia SaaS untuk secara aman dan efisien menawarkan layanan mereka kepada pelanggan yang beroperasi di beberapa dan. VPCs Akun AWS Pelanggan mengakses penawaran SaaS Anda melalui VPC Lattice, yang memberikan konektivitas jaringan yang konsisten, kontrol akses yang kuat, dan manajemen lalu lintas yang canggih. Dalam skenario ini, lalu lintas mengalir melalui VPC Lattice ke layanan aplikasi terdaftar Anda. Ini menyediakan komunikasi yang terukur dan aman, terlepas dari layanan komputasi mana yang Anda gunakan.

Peering VPC

VPC peering adalah koneksi jaringan antara dua virtual private cloud (VPCs) yang merutekan lalu lintas di antara mereka dengan menggunakan alamat atau IPv4 alamat pribadi. IPv6 Peering VPC biasanya digunakan antara entitas tepercaya, seperti yang ada dalam organisasi yang sama. Pelanggan Anda membuat permintaan peering ke salah satu dari Anda VPCs. Ketika Anda menerimanya, lalu lintas dapat mengalir di antara keduanya VPCs di kedua arah. Pendekatan koneksi ini menonjol karena keunikannya karena melibatkan komunikasi langsung antara dua VPCs tanpa layanan perantara atau infrastruktur untuk dikelola.

AWS Transit Gateway

AWS Transit Gatewayadalah hub transit jaringan terpusat yang dapat terhubung VPCs, koneksi jaringan pribadi virtual (VPN), AWS Direct Connect gateway, peralatan virtual pihak ketiga dalam VPC, dan gateway transit lainnya. Gateway transit dapat memiliki tabel rute yang berbeda untuk setiap lampiran. Ini memberikan fleksibilitas maksimum untuk routing, dan ini membantu Anda mengisolasi jaringan. Ini sering digunakan untuk menghubungkan banyak VPCs bersama-sama atau untuk inspeksi terpusat.

AWS Site-to-Site VPN

AWS Site-to-Site VPNdapat menggunakan teknologi keamanan protokol internet (IPsec) untuk membangun koneksi antara jaringan lokal, kantor jarak jauh, pabrik, penyedia cloud lainnya, dan jaringan AWS global. Koneksi dibuat dari gateway pribadi virtual atau gateway transit di VPC AWS Cloud ke gateway pelanggan berbasis fisik atau perangkat lunak, yang dapat berada di, lokal, atau di AWS Cloud cloud CSP lain. Koneksi dapat melalui Internet atau melalui AWS Direct Connect koneksi fisik. Dimungkinkan juga untuk memiliki koneksi Site-to-Site VPN yang dipercepat dengan menggunakan AWS Global Accelerator. Koneksi yang dipercepat merutekan lalu lintas ke lokasi AWS tepi, dan menawarkan pengurangan latensi dan peningkatan kinerja.

AWS Direct Connect

AWS Direct Connectmembangun koneksi pribadi berkecepatan tinggi antara pusat data lokal dan. AWS Cloud Dengan melewati internet publik, Direct Connect menyediakan koneksi latensi rendah yang lebih andal, aman, dan konsisten ke internet. AWS Cloud Pelanggan terhubung ke salah satu Direct Connect lokasi dan kemudian memilih koneksi host atau khusus AWS. Meskipun ini adalah pilihan arsitektur yang tidak umum untuk penawaran SaaS, ini bisa sangat cocok untuk penyedia SaaS yang memiliki sedikit konsumen perusahaan besar.

Membandingkan kemampuan layanan

Tabel berikut menguraikan kemampuan yang didukung dari Layanan AWS yang dibahas dalam panduan ini. Berikut ini adalah deskripsi kemampuan yang termasuk dalam tabel ini:

- Rentang CIDR yang tumpang tindih - Dapat menghubungkan dua atau lebih jaringan dengan rentang CIDR yang sama atau tumpang tindih
- Komunikasi dua arah - Dapat mendukung saluran komunikasi dua arah sehingga konsumen SaaS dapat mengekspos sumber daya internal, seperti database, ke penyedia SaaS
- IPv6— Dapat mendukung IPv6, baik single atau dual-stack
- Jumbo frame - Dapat mendukung jumbo frame dengan ukuran frame hingga 8.500 byte
- Hybrid-cloud — Dapat mendukung koneksi dengan jaringan lokal
- Multi-cloud — Dapat mendukung koneksi antar jaringan pada penyedia layanan cloud yang berbeda

Layanan atau pendekatan	Rentang CIDR yang tumpang tindih	Komunikasi dua arah	IPv6	Bingkai jumbo	Awan hibrida	Multi-awan
Pengintipan VPC	Tidak	Ya	Ya	Ya [5]	Tidak	Tidak
AWS PrivateLink	Ya	Ya [1]	Ya	Ya	Tidak [6]	Tidak [6]
Kisi VPC Amazon	Ya	Ya [1]	Ya	Ya	Tidak [6]	Tidak [6]
AWS Transit Gateway	Tidak	Ya	Ya	Ya	Ya [3]	Ya [3]
AWS Site-to-Site VPN	Tidak	Ya	Ya	Tidak	Ya	Ya

AWS Direct Connect	Tidak	Ya	Ya	Ya [2]	Ya	Ya
Akses internet publik [4]	Tidak berlaku	Tidak	Ya	Ya	Ya	Ya

1. Dengan sumber daya VPC di Amazon VPC Lattice
2. Hanya untuk antarmuka virtual pribadi dan transit
3. Dengan Site-to-Site VPN atau AWS Direct Connect lampiran
4. Sebagai istilah umum untuk AWS sumber daya yang membuat aplikasi dapat diakses publik, seperti Application Load Balancer
5. Hanya untuk mengintip koneksi dalam satu Wilayah AWS
6. Mungkin melalui koneksi Layer 3 yang sudah ada sebelumnya antara lingkungan

Fitur dan pertimbangan keamanan

Tabel berikut menguraikan fitur keamanan Layanan AWS yang dibahas dalam panduan ini.

- Cara otentikasi — Bagaimana Anda dapat memastikan bahwa hanya pelanggan Anda yang dapat terhubung ke layanan Anda. Tingkat otentikasi lain untuk permintaan masuk biasanya masih diperlukan, terutama di lingkungan penyewa bersama.
- Enkripsi dalam transit — Menjelaskan apakah enkripsi dalam transit disediakan secara default. Enkripsi asli menjelaskan enkripsi yang AWS menyediakan semua lalu lintas di dalam VPCs, di seluruh VPCs, atau di seluruh pusat data. Enkripsi tambahan menjelaskan enkripsi yang Anda kendalikan dan yang dapat dihentikan oleh layanan masing-masing.

Layanan atau pendekatan	Sarana otentikasi	Enkripsi dalam perjalanan
Pengintipan VPC	Anda memulai permintaan peering ke dan Akun AWS VPC pelanggan Anda atau menerima permintaan yang	Hanya enkripsi asli

	mereka mulai. Lihat Menerima atau menolak koneksi peering VPC.	
AWS PrivateLink	Anda memilih mana yang Akun AWS diizinkan untuk membuat titik akhir ke layanan Anda. Akun-akun ini dikenal sebagai prinsipal yang diizinkan. Lihat Menerima atau menolak permintaan koneksi.	Hanya enkripsi asli
Kisi VPC Amazon	Anda berbagi layanan VPC Lattice atau jaringan layanan dengan pelanggan Anda. Akun AWS Lihat Berbagi entitas Kisi VPC Anda.	Enkripsi asli dan enkripsi TLS tambahan
AWS Transit Gateway	Pelanggan Anda membuat permintaan lampiran peering dari mereka Akun AWS, atau Anda memulai permintaa n. Lihat Lampiran peering gateway Transit di Amazon VPC Transit Gateways.	Enkripsi asli dan IPsec enkripsi tambahan dengan lampiran VPN
AWS Site-to-Site VPN	Anda menggunakan kunci yang IPsec telah dibagikan sebelumnya atau sertifikat pribadi di perangkat pelanggan . Lihat opsi otentikasi AWS Site-to-Site VPN terowongan.	Enkripsi tambahan IPsec

AWS Direct Connect	Pelanggan Anda membuat permintaan antarmuka virtual dari mereka Akun AWS. Lihat antarmuka Direct Connect virtual dan antarmuka virtual yang dihosting.	Enkripsi Layer 2 tambahan dimungkinkan di situs yang dipilih. Lihat Direct Connect Lokasi.
Akses internet publik [1]	Diperlukan otentikasi khusus.	Enkripsi TLS tambahan dimungkinkan

1. Sebagai istilah umum untuk AWS sumber daya yang membuat aplikasi dapat diakses publik, seperti Application Load Balancer

Mengevaluasi opsi akses jaringan untuk penawaran SaaS

Metrik yang penting bagi organisasi Anda akan tergantung pada siapa pelanggan Anda, strategi bisnis Anda, dan tujuan organisasi Anda. Panduan ini menyajikan metrik yang dapat Anda gunakan untuk memilih pendekatan akses jaringan, tetapi Anda harus memprioritaskan metrik yang memenuhi persyaratan unik kasus penggunaan Anda.

Bagian ini berisi topik berikut:

Metrik evaluasi

Beberapa metrik konsisten di seluruh organisasi dan kasus penggunaan, dan ini adalah metrik yang dapat kami bantu untuk menilai Anda. Berikut ini adalah metrik ini:

- Kemudahan integrasi - Seberapa cepat dan mudah Anda dapat memasukkan pelanggan baru?
- Total biaya kepemilikan (TCO) — Apa struktur biaya? Di luar biaya infrastruktur tetap dan variabel, ada pertimbangan biaya tambahan utama yang terkait dengan overhead operasional, ketergantungan pada ahli, biaya penerapan perubahan, dan kepatuhan. Untuk informasi selengkapnya, lihat bagian Total biaya kepemilikan.
- Skalabilitas — Apakah pendekatan akses jaringan Anda dapat ditingkatkan untuk mendukung pertumbuhan perusahaan Anda? Menskalakan basis pelanggan Anda memiliki pertimbangan arsitektur dan organisasi yang penting. Pertimbangkan bagaimana Anda dapat menskalakan untuk mengakomodasi 5—100 kali lebih banyak pelanggan yang Anda dukung hari ini.
- Adaptabilitas — Dapatkah Anda menerapkan perubahan dengan mudah? Perubahan mungkin termasuk aplikasi baru, kemampuan baru, platform yang berbeda, atau jaringan yang berbeda.
- Isolasi jaringan — Berapa banyak infrastruktur jaringan yang Anda paparkan kepada pelanggan Anda? Apakah Anda menyediakan tingkat akses yang tepat, atau apakah Anda mengekspos seluruh jaringan? Jika Anda mengisolasi sumber daya jaringan lebih awal, akan lebih mudah untuk memberikan jaminan keamanan, privasi, dan kepatuhan nanti.
- Observabilitas — Apa kemampuan Anda untuk mendeteksi kegagalan atau degradasi layanan? Seberapa mudah dan cepat untuk mengidentifikasi masalah? Seberapa cepat (dan dengan biaya

overhead apa) Anda dapat membantu pelanggan Anda memahami poin kegagalan mereka dan membantu mereka menyelesaikannya?

- Waktu untuk memperbaiki — Berapa waktu tunggu antara deteksi kegagalan layanan atau degradasi dan melanjutkan operasi? Apa saja faktor yang mempengaruhi kemampuan ini?

Metrik lain unik untuk organisasi atau penawaran Anda karena berhubungan dengan operasi, strategi, atau tujuan bisnis Anda. Hanya Anda yang dapat menilai metrik ini. Berikut ini adalah metrik ini:

- Penyelarasan model bisnis — Apa model bisnis Anda, dan seberapa baik pendekatan akses individu selaras dengannya?
- Total addressable market (TAM) — Apa pasar Anda saat ini dan masa depan, dan seberapa baik itu dicakup oleh pendekatan akses jaringan?
- Return on Investment (ROI) — Peningkatan apa yang Anda harapkan dalam profitabilitas dan margin? Apakah manfaat finansial yang diharapkan cukup untuk memenuhi kebutuhan Anda akan akses layanan yang fleksibel dan fleksibel?
- Kepatuhan terhadap peraturan — Persyaratan peraturan seperti apa yang berlaku, dan di pasar mana?
- Perjanjian tingkat layanan (SLAs) - Apakah pelanggan membutuhkan penawaran SaaS Anda agar sangat tersedia? Komitmen seperti apa yang secara kontrak harus Anda pegang?

Total biaya kepemilikan

Bagian ini mengeksplorasi total biaya kepemilikan (TCO), yang merupakan salah satu metrik evaluasi yang digunakan untuk membandingkan pendekatan akses jaringan. TCO adalah metrik komposit yang terdiri dari biaya infrastruktur tetap dan variabel, overhead operasional, ketergantungan spesialis, biaya perubahan, dan biaya kepatuhan.

Peringkat TCO untuk setiap pendekatan akses jaringan mungkin berbeda untuk kasus penggunaan Anda. Misalnya, biaya perubahan untuk penyedia SaaS dengan layanan web sederhana dan lima penyewa berbeda dari penyedia SaaS dengan portofolio produk yang kompleks dan saling berhubungan dan ratusan atau ribuan penyewa. Selain itu, tidak semua komponen memiliki bobot yang sama. Misalnya, menyewa spesialis jaringan seringkali lebih mahal daripada biaya infrastruktur yang mendukung penyebaran individu layanan Anda. Gunakan nilai-nilai dalam tabel berikut untuk orientasi awal dan sebagai titik referensi untuk diskusi lebih lanjut.

Pendekatan akses	Biaya infrastruktur tetap	Biaya infrastruktur variabel	Overhead operasional	Ketergantungan spesialis	Biaya perubahan	Biaya kepatuhan
VPC mengintip	Tidak ada	Tidak ada	Tinggi	Rendah	Tinggi	Sedang
AWS PrivateLink	Rendah	Rendah	Rendah	Tidak ada	Rendah	Rendah
Kisi VPC Amazon	Sedang	Sedang	Rendah	Rendah	Rendah	Rendah
AWS Transit Gateway	Sedang	Sedang	Rendah	Rendah	Rendah	Sedang
AWS Site-to-Site VPN	Sedang	Tinggi	Tinggi	Sedang	Sedang	Rendah
AWS Direct Connect	Tinggi	Sedang	Sedang	Tinggi	Tinggi	Rendah
Akses internet publik	Rendah	Tinggi	Sedang	Rendah	Rendah	Tinggi

Biaya mengintip VPC

Tidak ada biaya infrastruktur langsung yang terkait dengan koneksi peering VPC. Ketika lalu lintas tetap berada dalam Availability Zone yang sama, tidak ada biaya transfer data. Namun, overhead operasional dapat menjadi signifikan karena manajemen dan kompleksitas tumbuh secara eksponensial dengan setiap koneksi peering tambahan. Beberapa pemahaman dasar tentang jaringan sudah cukup untuk mengatur koneksi peering, tetapi perubahan pada jaringan sulit

diterapkan dengan lebih dari segelintir koneksi peering. Biaya kepatuhan sedikit lebih tinggi karena kedua belah pihak mengekspos seluruh VPC satu sama lain, bukan layanan individual.

AWS PrivateLink biaya

AWS PrivateLink seringkali merupakan solusi hemat biaya dengan overhead operasional kecil. Ini karena penyedia SaaS harus mengelola hanya Network Load Balancer, dan konsumen harus mengelola hanya titik akhir VPC. Anda dapat membuat perubahan di kedua sisi secara transparan, yang mengurangi kolaborasi lintas organisasi yang mahal dan intensif sumber daya. Biaya kepatuhan cenderung rendah karena penyedia SaaS hanya mengekspos layanan yang mereka inginkan dan bukan seluruh jaringan.

Biaya Amazon VPC Lattice

Amazon VPC Lattice menawarkan struktur biaya yang seimbang dengan biaya infrastruktur tetap dan variabel yang moderat. Sebagai jaringan layanan yang dikelola sepenuhnya, ini secara signifikan mengurangi overhead operasional dengan mengotomatiskan penemuan layanan, manajemen lalu lintas, dan kontrol akses di beberapa. VPCs Ini menyederhanakan penerapan awal dan manajemen berkelanjutan dibandingkan dengan konfigurasi jaringan manual. Anda dapat menerapkan perubahan melalui kontrol berbasis kebijakan tanpa pembaruan perutean yang rumit, yang mengurangi ketergantungan pada spesialis jaringan. Biaya kepatuhan cenderung lebih rendah daripada pendekatan jaringan tradisional karena VPC Lattice menyediakan kontrol akses yang halus dan visibilitas komprehensif melalui kemampuan pemantauan dan pencatatan bawaan. Ini dapat mempermudah untuk menunjukkan kepatuhan terhadap peraturan.

AWS Transit Gateway biaya

AWS Transit Gateway memiliki biaya per jam dan pemrosesan data yang lebih besar daripada AWS PrivateLink, tetapi memiliki overhead operasional yang serupa. Anda harus memiliki pengetahuan yang lebih dalam tentang AWS Transit Gateway layanan dan perutean untuk mengatur semua tabel rute dengan benar. AWS Perubahan infrastruktur mungkin memerlukan perutean atau pembaruan DNS. Biaya kepatuhan serupa dengan peering VPC karena kedua belah pihak berpotensi mengekspos subjaringan atau keseluruhan satu sama lain. VPCs AWS Transit Gateway tabel rute juga perlu ditangani dengan hati-hati karena dibagikan oleh banyak konsumen, dan Anda tidak boleh mengizinkan lalu lintas di antara mereka.

AWS Site-to-Site VPN biaya

Karena Site-to-Site VPN pada dasarnya mengirimkan lalu lintas ke internet, biaya variabel tertinggi dibandingkan karena biaya transfer data. Meskipun ini adalah layanan jaringan pribadi virtual terkelola (VPN), ia hadir dengan overhead operasional yang signifikan, terutama pada gateway pelanggan. Penyediaan dan operasi membutuhkan pengetahuan lanjutan tentang jaringan, dan perubahan sering memerlukan tindakan dari kedua belah pihak. Biaya kepatuhan biasanya rendah karena tim keamanan sering menyetujui IPsec terowongan tanpa peninjauan tambahan.

AWS Direct Connect biaya

AWS Direct Connect datang dengan biaya infrastruktur tetap terbesar karena merupakan koneksi fisik pribadi langsung ke AWS Cloud. Pengetahuan khusus diperlukan untuk mengatur dan mengoperasikan sesi Border Gateway Protocol (BGP) (jika diperlukan), untuk mengoperasikan koneksi VPN, dan untuk melakukan rekayasa lalu lintas. Layanan ini mengurangi upaya untuk tim keamanan karena memadukan konektivitas pribadi dengan opsi tambahan memiliki Media Access Control Security (MACsec) dan IPsec enkripsi.

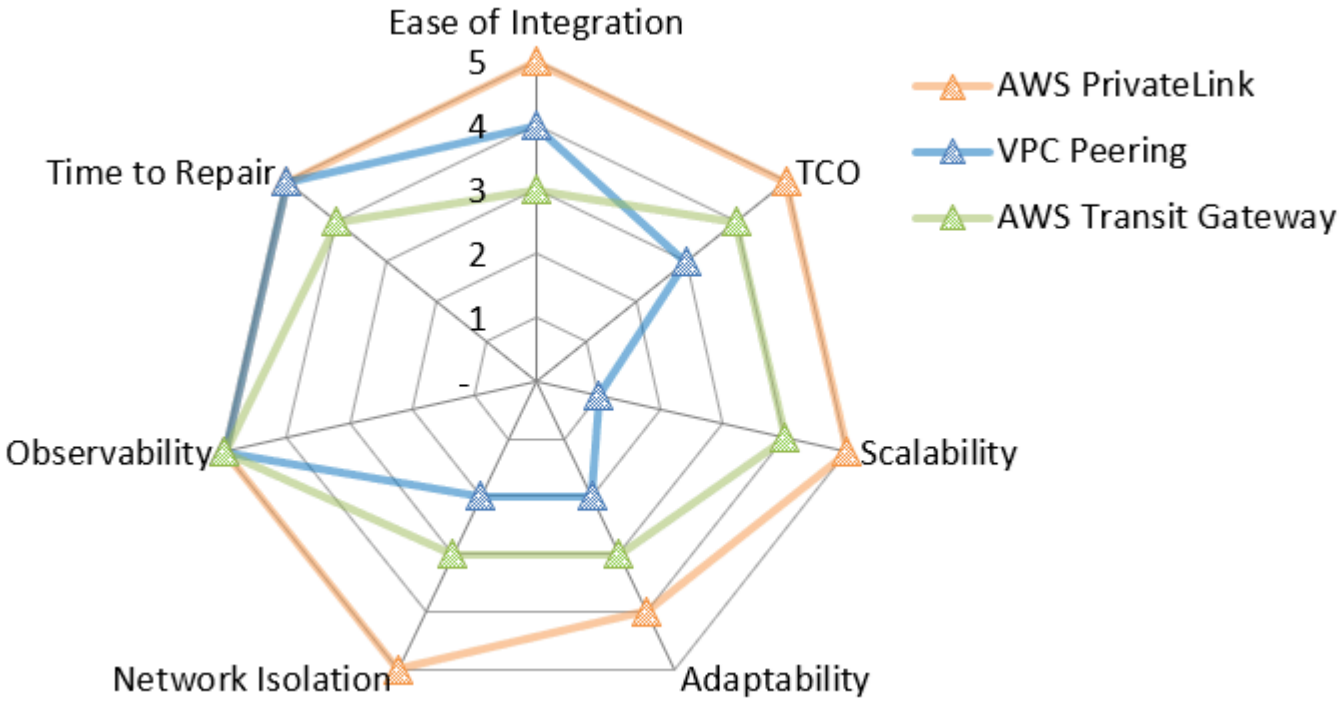
Biaya akses internet publik

Akses internet publik mengacu pada AWS sumber daya yang dapat Anda gunakan untuk membuat aplikasi dapat diakses publik, seperti Application Load Balancer. Untuk pendekatan ini, ada biaya variabel yang terkait dengan penyediaan akses ke layanan Anda, termasuk biaya untuk transfer data ke internet. Biaya overhead dan kepatuhan operasional dapat menjadi signifikan karena Anda mengekspos layanan ke Internet dan akan memerlukan mekanisme keamanan dan otentikasi tambahan. Namun, tidak ada perutean yang rumit yang terlibat, dan tidak ada pihak yang harus mengetahui detail tentang infrastruktur masing-masing.

Peta nilai jaringan

Untuk membantu Anda melihat gambaran besar dan membuat keputusan berdasarkan informasi, panduan ini menyertakan peta nilai jaringan untuk setiap skenario. Karena peringkat berbeda dari skenario ke skenario, layanan yang sama mungkin mendapat skor berbeda untuk dua skenario. Peta nilai adalah bagan radar, di mana skor sempurna hipotetis akan menjadi lima di semua kategori.

Misalnya, gambar berikut menunjukkan bagan radar sampel. Ini hanya mencakup metrik yang dapat kami bantu evaluasi. Kami menyarankan Anda membuat peta nilai Anda sendiri yang mencakup metrik tambahan yang hanya dapat Anda evaluasi.


Ease of Integration
5
4
3
2
1
-
TCO
Scalability
Adaptability
Network Isolation
Observability
Time to Repair
AWS PrivateLink
VPC Peering
AWS Transit Gateway

Skenario akses jaringan untuk penawaran SaaS di AWS Cloud

Bagian ini mencakup berbagai opsi akses jaringan untuk penawaran SaaS Anda di. AWS Cloud Ini membahas pendekatan dari perspektif konsumen Anda, yang mungkin memiliki kebutuhan konektivitas di dalam AWS Cloud, dari pusat data lokal, atau dari penyedia layanan cloud lainnya ()CSPs. Selain itu, Anda mungkin perlu mendukung akses dari berbagai jenis lingkungan konsumen.

Memahami persyaratan konektivitas jaringan di lingkungan yang beragam ini sangat penting untuk menciptakan strategi akses yang komprehensif. Keputusan arsitektur Anda harus memperhitungkan berbagai model keamanan, ekspektasi kinerja, dan kendala teknis sambil mempertahankan efisiensi operasional. Pendekatan yang tepat menyediakan konektivitas yang aman dan andal yang sesuai dengan pertumbuhan bisnis Anda dan meminimalkan kompleksitas implementasi dan overhead manajemen yang berkelanjutan.

Saat mengevaluasi opsi akses jaringan, pertimbangkan bagaimana setiap pendekatan memengaruhi total biaya kepemilikan Anda, termasuk tidak hanya biaya infrastruktur tetapi juga persyaratan overhead dan kepatuhan operasional. Beberapa pendekatan unggul dalam skalabilitas tetapi mungkin memperkenalkan kompleksitas, sementara yang lain memprioritaskan kemudahan integrasi dengan mengorbankan isolasi jaringan. Kemampuan teknis dan sumber daya konsumen Anda juga memainkan peran penting dalam menentukan solusi yang paling tepat.

Bagi konsumen di AWS Cloud, layanan seperti AWS PrivateLink menawarkan keuntungan yang signifikan dalam keamanan dan skalabilitas. Konsumen lokal mungkin mendapat manfaat dari AWS Direct Connect kinerja yang konsisten atau manfaat dari Site-to-Site VPN untuk konektivitas yang hemat biaya. Skenario multi-cloud seringkali memerlukan pertimbangan yang cermat terhadap tantangan interoperabilitas, dan Anda mungkin menggunakan arsitektur VPC transit untuk menstandarisasi pola akses. Dalam semua kasus, desain Anda harus mengantisipasi pertumbuhan konsumen dan lalu lintas di masa depan sehingga arsitektur jaringan Anda tetap tangguh dan mudah beradaptasi saat penawaran SaaS Anda berkembang.

Bagian ini berisi skenario berikut:

- Konsumen SaaS beroperasi AWS
- Konsumen layanan yang beroperasi di tempat
- Konsumen SaaS yang beroperasi di penyedia layanan cloud lainnya
- Mendukung lingkungan hibrida

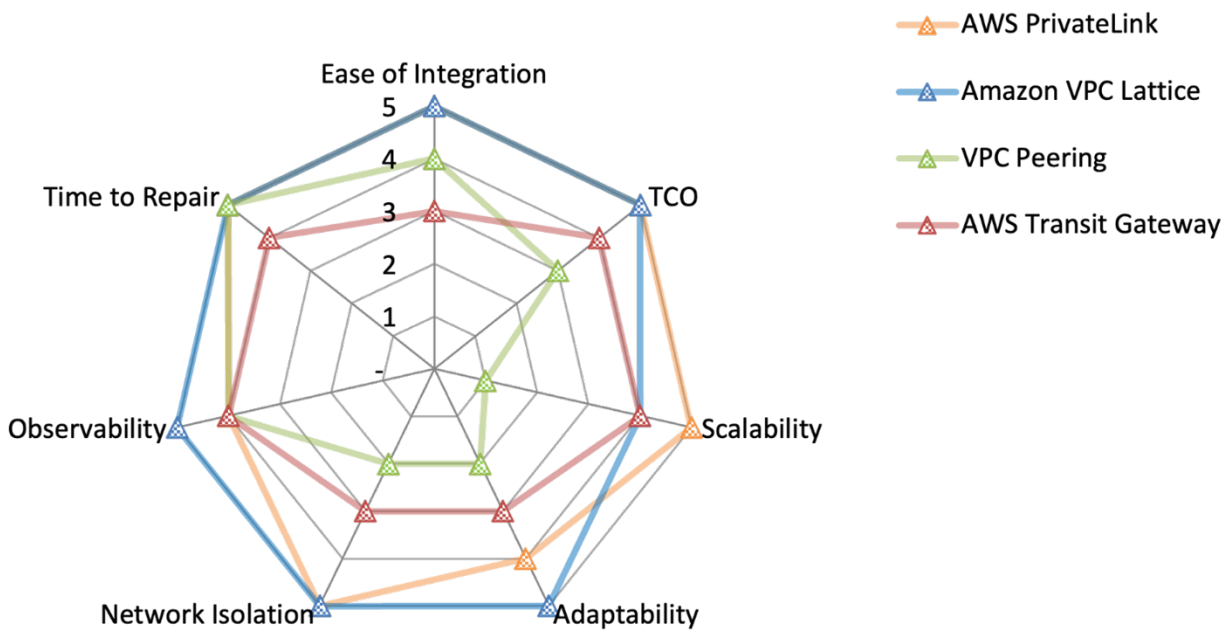
Konsumen SaaS beroperasi AWS

Bagian ini membahas opsi konektivitas jika Anda dan konsumen Anda beroperasi di. AWS Cloud Skenario ini menawarkan fleksibilitas terbesar karena banyak yang terintegrasi Layanan AWS secara native dan karena kedua belah pihak memiliki akses ke seluruh Layanan AWS portofolio.

Bagian ini membahas pendekatan akses jaringan berikut:

- Integrasi dengan AWS PrivateLink
- Berbagi layanan Amazon VPC Lattice
- Membuat koneksi peering VPC
- Menghubungkan VPCs dengan AWS Transit Gateway

Peta nilai jaringan berikut merangkum bagaimana masing-masing opsi ini mendapat skor untuk setiap metrik evaluasi. Untuk informasi selengkapnya tentang metrik evaluasi, lihat Metrik evaluasi dalam panduan ini. Dalam peta, lima mewakili skor terbaik, seperti TCO terendah, isolasi jaringan terbaik, atau waktu terendah untuk memperbaiki. Untuk informasi lebih lanjut tentang cara membaca bagan radar ini, lihat Peta nilai jaringan di panduan ini.



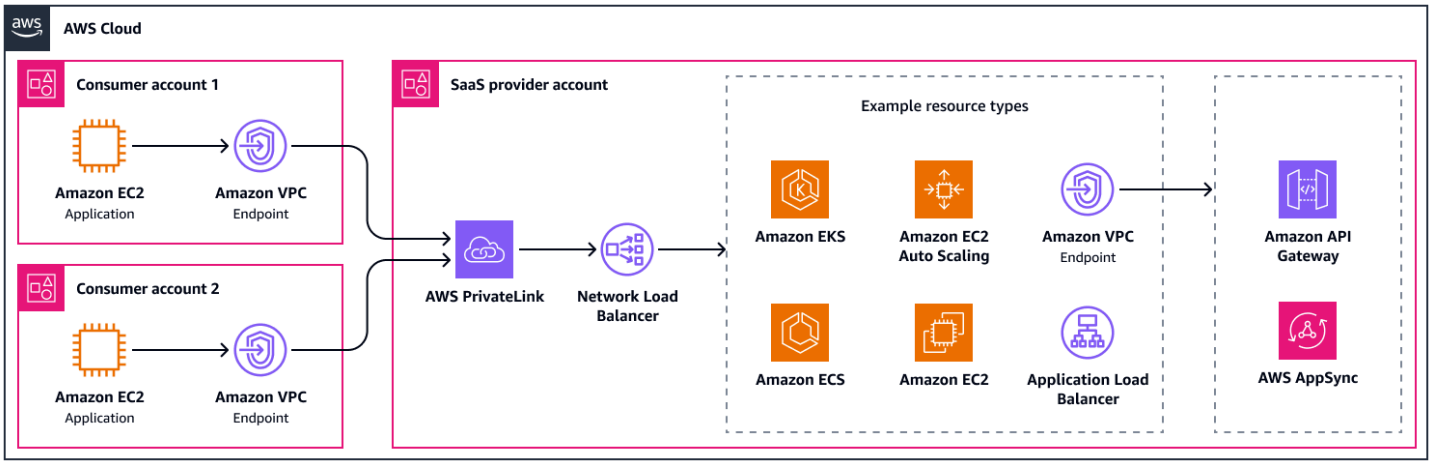
Bagan radar menunjukkan nilai-nilai berikut.

Metrik evaluasi	AWS PrivateLink	Kisi VPC Amazon	Peering VPC	AWS Transit Gateway
Kemudahan integrasi	5	5	4	3
TCO	5	5	3	4
Skalabilitas	5	4	1	4
Kemampuan beradaptasi	4	5	2	3
Isolasi jaringan	5	5	2	3
Observabilitas	4	5	4	4
Waktu untuk memperbaiki	5	5	5	4

Integrasi dengan AWS PrivateLink

AWS PrivateLinkadalah cara paling cloud-native untuk mengintegrasikan penawaran SaaS. Penyedia SaaS dapat meng-host aplikasi mereka baik di belakang Network Load Balancer. Network Load Balancer langsung terintegrasi dengan Application Load Balancer, Amazon Elastic Container Service (Amazon ECS) Service Elastic Container ECS), Amazon Elastic Kubernetes Service(Amazon EKS), dan grup Auto Scaling. Dimungkinkan juga untuk merutekan lalu lintas dari Network Load Balancer ke titik akhir VPC antarmuka di akun penyedia SaaS. Ini membantu Anda menggunakan API untuk menjangkau aplikasi, seperti melalui Amazon API Gateway atau AWS AppSync. Jika aplikasi Anda memerlukan akses ke sumber daya di lingkungan pelanggan yang tidak seimbang beban, seperti database, Anda dapat menggunakan titik akhir VPC sumber daya.

AWS PrivateLink mendukung bandwidth hingga 100 Gbps per Availability Zone. Diagram berikut menunjukkan konfigurasi dasar dengan beberapa kemungkinan integrasi. Ini menghubungkan dua akun konsumen ke akun penyedia SaaS melalui. AWS PrivateLink Ada titik akhir layanan di akun konsumen dan Network Load Balancer di akun penyedia SaaS.



Berikut ini adalah manfaat dari pendekatan ini:

- Kemudahan integrasi: Tidak diperlukan perubahan tabel rute
- Kemudahan integrasi: Anda dapat menawarkan layanan endpoint melalui AWS Marketplace
- Kemudahan integrasi: Titik akhir VPC mendukung nama DNS yang ramah
- Skalabilitas: Dapat diskalakan ke ribuan konsumen SaaS
- Adaptabilitas: Dukungan untuk rentang CIDR yang tumpang tindih
- Adaptabilitas: Support untuk IPv6
- Kemampuan beradaptasi: Dukungan Lintas Wilayah
- TCO: AWS PrivateLink adalah layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih sedikit upaya operasional
- Isolasi jaringan: Manfaat keamanan bagi konsumen SaaS karena lalu lintas tidak dapat dimulai dari penyedia SaaS
- Isolasi jaringan: Manfaat keamanan untuk penyedia SaaS karena mereka tidak mengekspos seluruh subnet atau VPC

Berikut ini adalah kelemahan dari pendekatan ini:

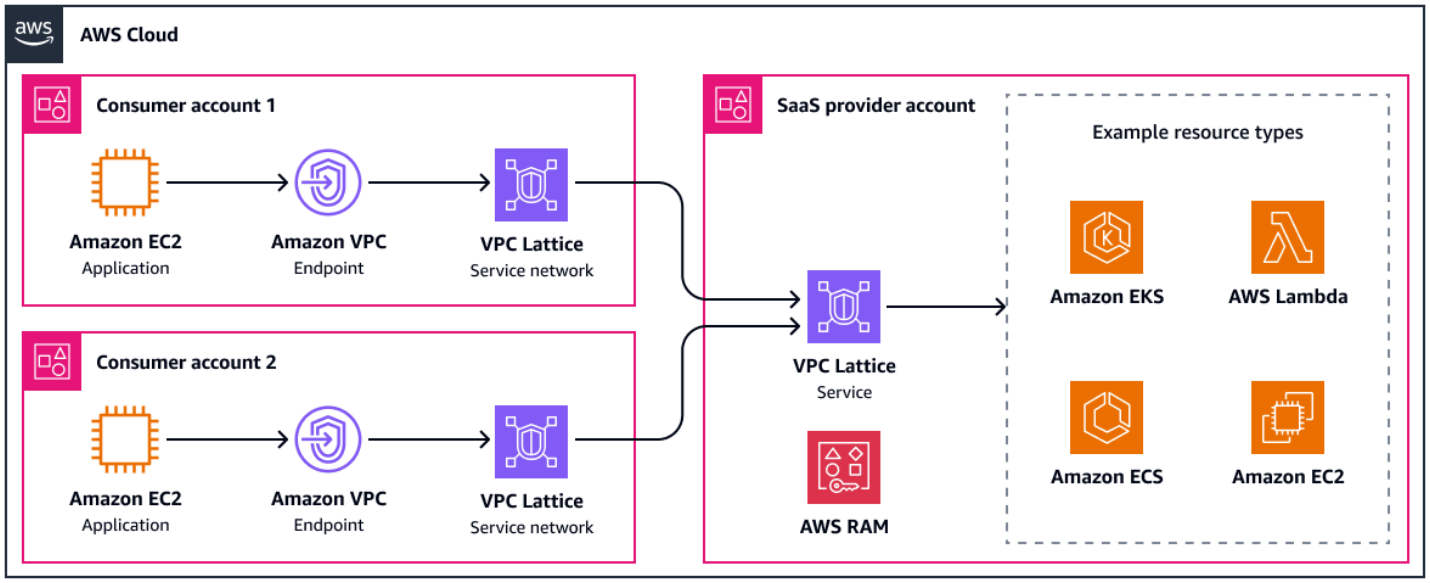
- Kemampuan beradaptasi: Penyedia SaaS harus menggunakan Availability Zone yang sama dengan konsumen
- Adaptabilitas: Dukungan hanya untuk koneksi yang diprakarsai klien, dan titik akhir VPC sumber daya diperlukan untuk komunikasi yang dimulai layanan
- Adaptabilitas: Network Load Balancer adalah satu-satunya integrasi langsung untuk AWS PrivateLink

Berbagi layanan Amazon VPC Lattice

Untuk menggunakan Amazon VPC Lattice sebagai opsi konektivitas untuk aplikasi SaaS Anda, pertama-tama Anda membuat satu atau beberapa layanan VPC Lattice yang mewakili komponen aplikasi SaaS Anda. Anda mengonfigurasi pendengar dan aturan perutean untuk mengarahkan lalu lintas ke target backend Anda, seperti instans, wadah, atau fungsi Amazon EC2. AWS Lambda Untuk informasi selengkapnya, lihat Menghubungkan layanan Saas dalam jaringan layanan VPC LatticeAWS (posting blog). Dari segi konsep, ini hampir sama dengan mengkonfigurasi Application Load Balancer. Kemudian, Anda membagikan layanan SaaS Anda secara aman dengan pelanggan Akun AWS atau organisasi dengan menggunakan AWS Resource Access Manager (AWS RAM), menentukan izin apa yang mereka miliki. Setelah pelanggan menerima pembagian sumber daya, mereka dapat mengaitkan layanan SaaS Anda dengan jaringan layanan VPC Lattice yang ada atau yang baru dibuat untuk mengaktifkan komunikasi. service-to-service

Setiap layanan VPC Lattice dapat mendukung hingga 10 Gbps dan 10.000 permintaan per detik per Availability Zone. Dengan menerapkan kebijakan autentikasi, pelanggan Anda dapat memiliki kontrol yang baik atas layanan dan sumber daya mana yang dapat mengakses aplikasi SaaS. Anda dapat menggunakan gateway sumber daya untuk mengakses sumber daya yang memerlukan koneksi TCP. Misalnya, ini mungkin kluster Amazon EKS yang Anda kelola, atau mungkin sumber daya yang dikelola pelanggan yang perlu diakses aplikasi Anda. Untuk informasi selengkapnya tentang penggunaan gateway sumber daya untuk penawaran SaaS, lihat Memperluas kemampuan SaaS Akun AWS di seluruh AWS PrivateLink penggunaan dukungan untuk sumber daya VPC (posting blog).AWS

Diagram berikut menunjukkan konfigurasi VPC Lattice tingkat tinggi dengan beberapa contoh integrasi. Ini menggunakan jaringan layanan yang dikelola pelanggan untuk mengakses aplikasi SaaS.



Berikut ini adalah manfaat dari pendekatan ini:

- Kemudahan integrasi: Tidak diperlukan perubahan tabel rute
- Kemudahan integrasi: Penemuan layanan di luar kotak
- Skalabilitas: Dapat diskalakan ke ribuan konsumen SaaS
- Adaptabilitas: Dukungan untuk rentang CIDR yang tumpang tindih
- Adaptabilitas: Support untuk IPv6
- Adaptabilitas: Terintegrasi dengan layanan AWS komputasi apa pun sebagai layanan VPC Lattice
- TCO: VPC Lattice adalah layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih sedikit upaya operasional
- TCO: Penyeimbangan beban bawaan dengan perutean lalu lintas tingkat lanjut
- Isolasi jaringan: Otorisasi berbutir halus dengan kebijakan autentikasi
- Isolasi jaringan: Manfaat keamanan bagi konsumen SaaS karena lalu lintas tidak dapat dimulai dari penyedia SaaS
- Isolasi jaringan: Manfaat keamanan untuk penyedia SaaS karena Anda tidak mengekspos seluruh subnet atau VPC

Berikut ini adalah kelemahan dari pendekatan ini:

- Adaptabilitas: Dukungan hanya untuk koneksi yang diprakarsai klien, dan gateway sumber daya diperlukan untuk komunikasi yang diprakarsai layanan

- Kemampuan beradaptasi: Tidak ada dukungan Lintas wilayah

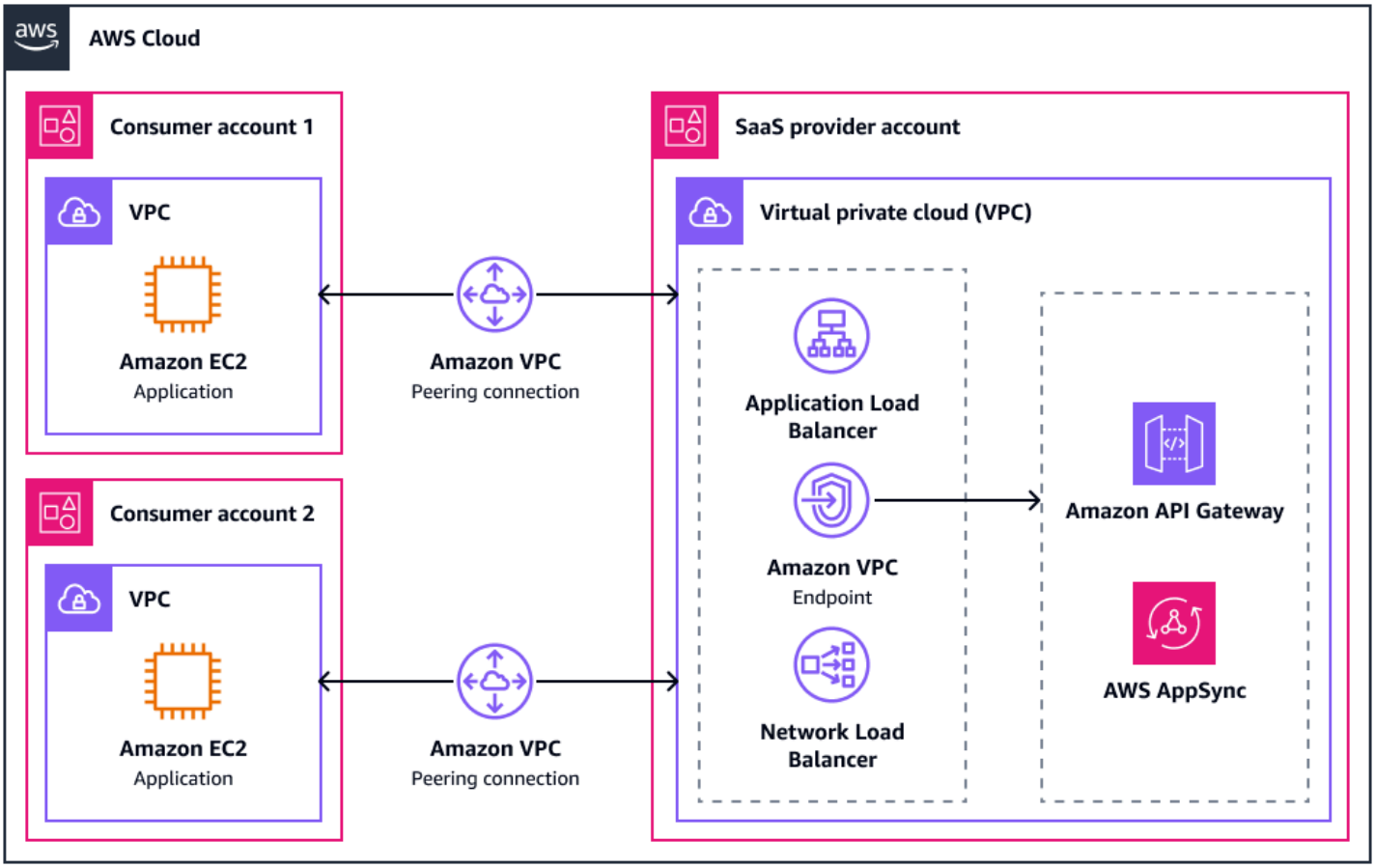
Membuat koneksi peering VPC

Saat Anda menggunakan peering VPC untuk menghubungkan VPC penyedia SaaS dengan VPC konsumen, kedua belah pihak dapat memulai koneksi. Ini memerlukan konfigurasi yang tepat dari grup keamanan, firewall, dan daftar kontrol akses jaringan (NACLs) di kedua akun. Jika tidak, lalu lintas yang tidak diinginkan mungkin masuk ke jaringan melalui koneksi peering. Anda dapat menggunakan grup keamanan untuk mereferensikan grup keamanan dari peered VPCs. Ini dapat membantu Anda mengontrol akses ke aplikasi Anda karena grup keamanan daftar izin memberikan kontrol akses yang lebih eksplisit dan terperinci dibandingkan dengan alamat IP daftar yang diizinkan.

Dengan VPC peering, penawaran SaaS dapat dicapai melalui layanan atau sumber daya yang digunakan di VPC. Sebagian besar aplikasi SaaS berada di belakang Application Load Balancer atau Network Load Balancer. AWS AppSync private APIs atau Amazon API Gateway private APIs adalah titik masuk umum lainnya ke aplikasi SaaS karena mereka dapat menjadi target melalui koneksi peering melalui titik akhir VPC antarmuka.

Setelah Anda membuat koneksi peering, Anda harus memperbarui tabel rute untuk VPCs di kedua akun untuk menentukan koneksi peering sebagai lompatan berikutnya untuk rentang CIDR masing-masing. Solusi ini direkomendasikan hanya untuk penyedia SaaS yang memiliki beberapa konsumen karena mengelola beberapa koneksi peering dengan cepat menjadi terlalu rumit.

Diagram berikut menunjukkan konfigurasi dasar dengan beberapa kemungkinan integrasi. VPCs di dua akun konsumen memiliki koneksi peering dengan VPC di akun penyedia SaaS.



Berikut ini adalah manfaat dari pendekatan ini:

- Waktu untuk memperbaiki: Tidak ada satu titik kegagalan untuk komunikasi
- Skalabilitas: Tidak ada batasan bandwidth selama pengintip VPC
- TCO: Tidak ada biaya untuk mengintip koneksi atau lalu lintas melalui koneksi peering dalam Availability Zone yang sama
- TCO: Tidak ada infrastruktur untuk dikelola
- Adaptabilitas: Support untuk IPv6
- Kemampuan beradaptasi: Didukung peering Antar Wilayah

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemampuan beradaptasi: Tidak ada dukungan untuk perutean transitif
- Kemampuan beradaptasi: Tidak ada dukungan untuk rentang CIDR yang tumpang tindih
- Skalabilitas: Skalabilitas terbatas (maksimum 125 koneksi peering per VPC)

- TCO: Kompleksitas tumbuh secara eksponensial dengan setiap koneksi peering tambahan
- TCO: Overhead dari mengelola tabel rute, mengintip koneksi sendiri, aturan kelompok keamanan, dan inspeksi lalu lintas
- Isolasi jaringan: Kontrol keamanan yang ketat diperlukan karena seluruh VPCs kedua belah pihak terpapar

Menghubungkan VPCs dengan AWS Transit Gateway

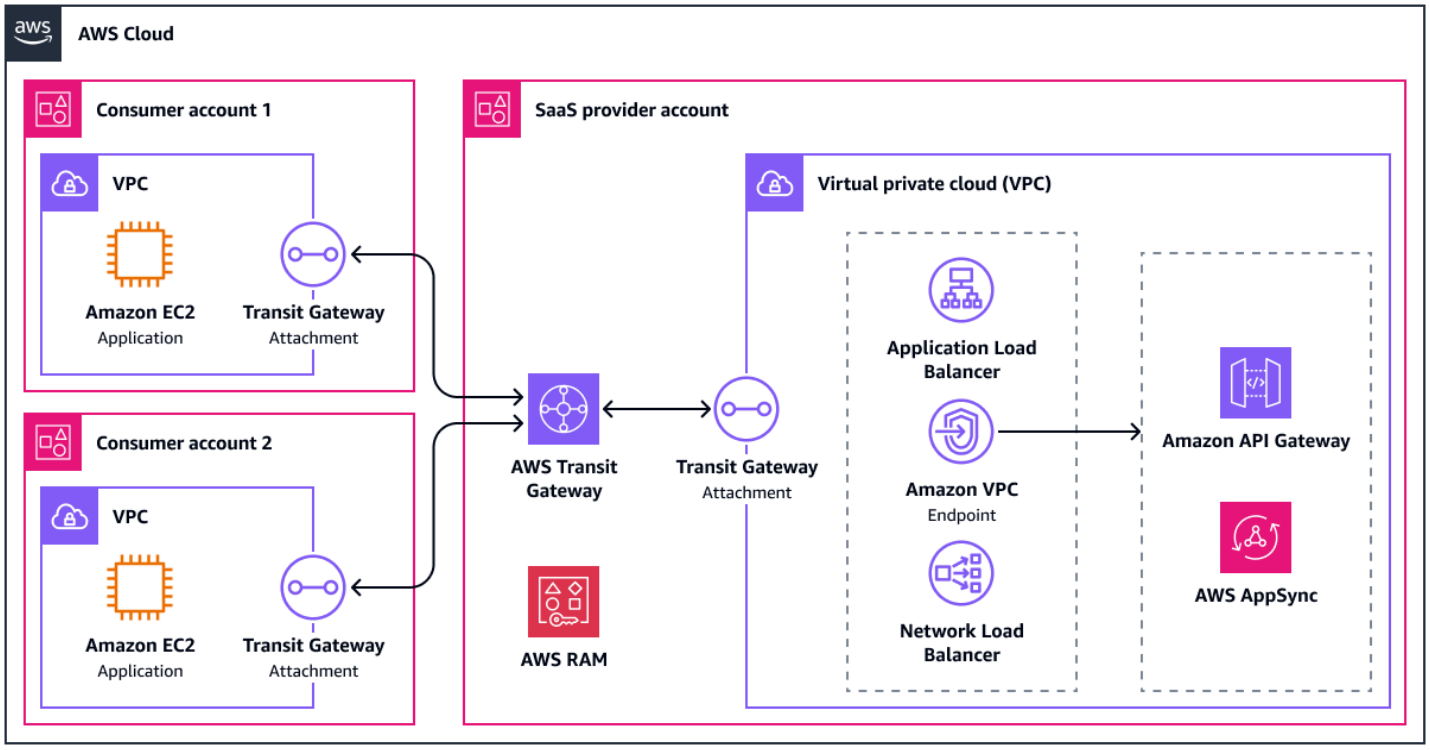
Saat Anda terhubung AWS Transit Gateway, itu membuat lampiran VPC dan menyebarkan antarmuka jaringan di subnet setiap Availability Zone yang seharusnya VPCs merutekan lalu lintas ke dan dari VPC. Disarankan untuk memiliki `/28` subnet khusus di setiap Availability Zone untuk lampiran VPC. Untuk informasi selengkapnya, lihat Praktik terbaik desain Gateway Transit VPC Amazon. VPCs Memerlukan tabel rute yang diperbarui untuk mengirim lalu lintas melalui antarmuka jaringan yang digunakan, dan tabel rute Transit Gateway perlu diperbarui sesuai dengan itu. Dalam konfigurasi multi-penyewa, Anda ingin VPC penyedia SaaS memiliki rute ke semua konsumen. VPCs Konsumen VPCs harus memiliki rute hanya ke VPC penyedia SaaS.

Transit Gateway sangat tersedia berdasarkan desain. Ini mendukung pemantauan dengan VPC Flow Logs, dan bandwidth maksimum untuk lampiran Transit Gateway adalah 100 Gbps per Availability Zone. Seperti peering VPC, pendekatan ini memungkinkan referensi grup keamanan lintas-VPC, yang menyederhanakan kontrol akses antar lingkungan.

Ada dua opsi utama untuk menghubungkan konsumen dengan penawaran SaaS Anda dengan Transit Gateway.

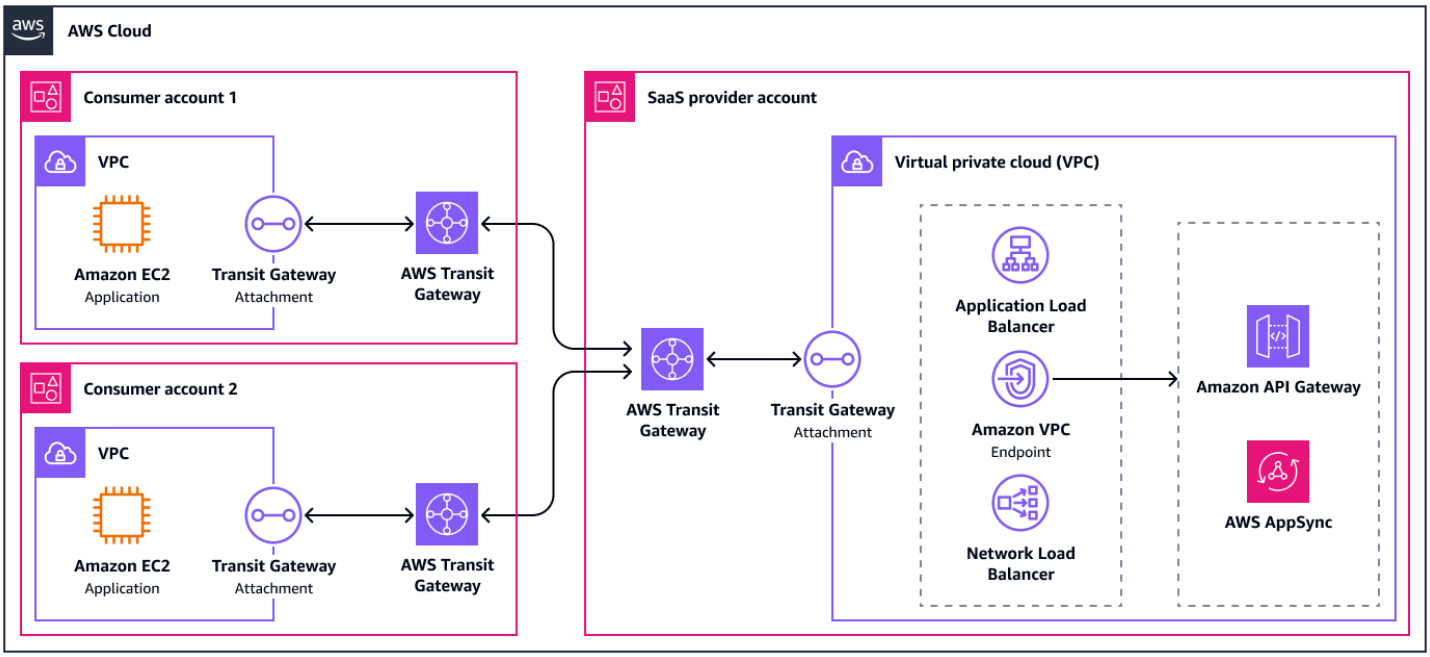
Opsi 1: Menggunakan RAM

Pada opsi pertama, penyedia layanan berbagi Transit Gateway dengan konsumen dengan menggunakan AWS Resource Access Manager (AWS RAM). Hal ini memungkinkan konsumen untuk menyebarkan lampiran VPC di akun mereka sendiri. Diagram berikut menunjukkan opsi ini pada tingkat tinggi.



Opsi 2: Gerbang transit peered

Opsi kedua adalah mengintip gateway transit Anda dengan gateway transit di akun konsumen. Ini memberi konsumen lebih banyak fleksibilitas karena mereka sekarang dapat sepenuhnya mengontrol tabel rute di dalam gateway transit mereka. Misalnya, mereka dapat mengatur inspeksi terpusat antara layanan dan beban kerja mereka. Kelemahan dari opsi ini hanya perutean statis antara gateway transit yang didukung. Diagram berikut menunjukkan opsi ini pada tingkat tinggi.



Berikut ini adalah manfaat dari pendekatan ini:

- Skalabilitas: Dukungan hingga 5.000 lampiran
- Skalabilitas: Satu tempat untuk mengelola dan memantau semua yang terhubung VPCs
- Kemampuan beradaptasi: Transit Gateway juga dapat dipasang ke VPNs, Direct Connect gateway, dan peralatan SD-WAN pihak ketiga
- Adaptabilitas: Arsitektur fleksibel, seperti menambahkan VPC inspeksi
- Adaptability: Support untuk routing transitif
- Kemampuan beradaptasi: Dapat mengintip gateway transit intra-wilayah dan antar wilayah
- Adaptabilitas: Support untuk IPv6
- TCO: AWS Transit Gateway adalah layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih sedikit upaya operasional
- TCO: TCO tumbuh secara linier dengan setiap lampiran gateway transit tambahan

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemudahan integrasi: Konfigurasi perutean membutuhkan pengetahuan jaringan tingkat lanjut
- Kemampuan beradaptasi: Tidak ada dukungan untuk rentang CIDR yang tumpang tindih
- TCO: Overhead dari mengelola entri tabel rute, aturan grup keamanan, dan inspeksi lalu lintas

- Keamanan: Kontrol keamanan yang ketat diperlukan karena seluruh VPCs kedua belah pihak terpapar

Konsumen layanan yang beroperasi di tempat

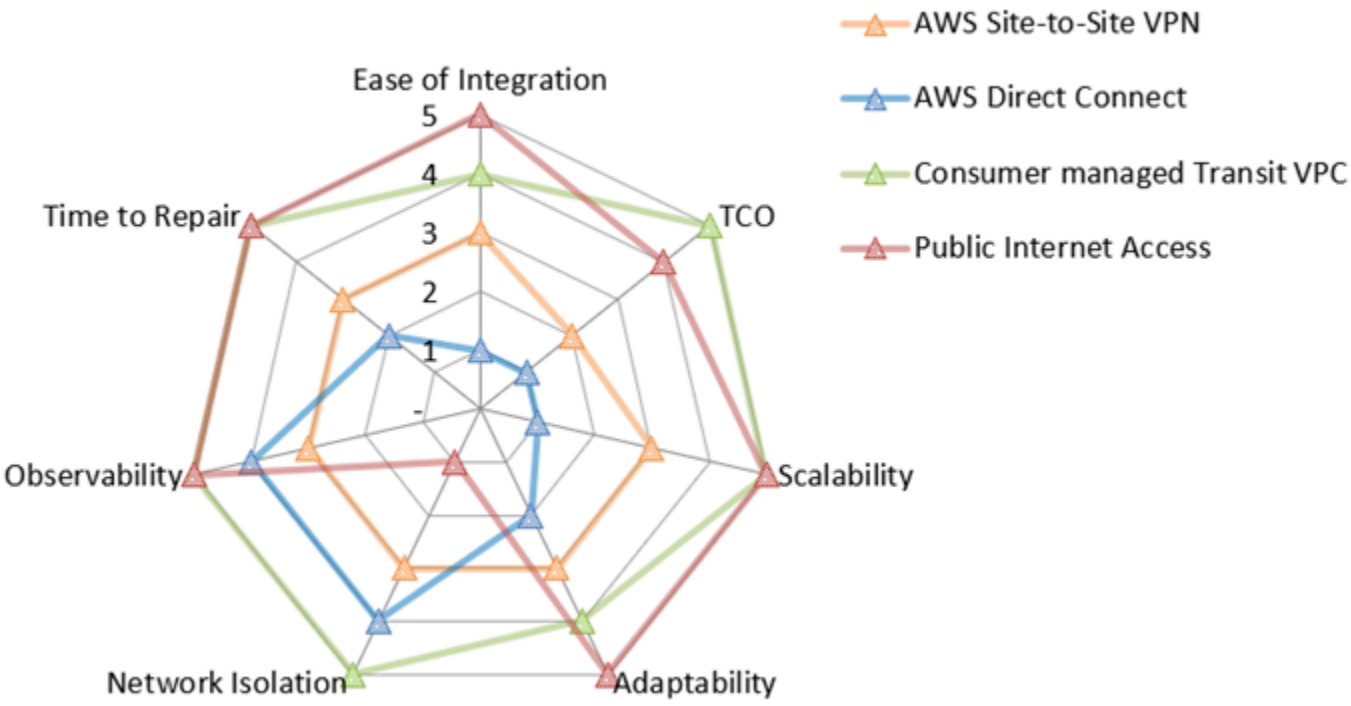
Bagian ini membahas opsi konektivitas antara beban kerja SaaS di AWS Cloud pusat data lokal dan SaaS. Banyak konsumen dengan persyaratan lokal, terutama di tingkat perusahaan, melihat cloud sebagai perpanjangan dari jaringan fisik mereka, dan mereka ingin mencerminkan hal itu dalam arsitektur mereka. Itu berarti konektivitas pribadi ke penawaran SaaS di cloud, baik melalui terowongan logis atau bahkan melalui koneksi fisik pribadi. Konsumen lain akan menerima konektivitas melalui internet publik, yang juga dibahas dalam bagian ini.

Bagian ini membahas pendekatan akses jaringan berikut:

- Menghubungkan dengan AWS Site-to-Site VPN
- Menghubungkan dengan AWS Direct Connect
- Menghubungkan dengan arsitektur VPC transit
- Menghubungkan melalui internet publik

Peta nilai jaringan berikut merangkum bagaimana masing-masing opsi ini mendapat skor untuk setiap metrik evaluasi. Untuk informasi selengkapnya tentang metrik evaluasi, lihat Metrik evaluasi dalam panduan ini. Dalam peta, lima mewakili skor terbaik, seperti TCO terendah, isolasi jaringan terbaik, atau waktu terendah untuk memperbaiki. Untuk informasi lebih lanjut tentang cara membaca bagan radar ini, lihat Peta nilai jaringan di panduan ini.

> **Note**
>
> Opsi VPC transit yang dikelola penyedia dikecualikan karena skornya sangat bergantung pada layanan mana yang dioperasikan.



Bagan radar menunjukkan nilai-nilai berikut.

Metrik evaluasi	AWS Site-to-Site VPN	AWS Direct Connect	VPC transit yang dikelola konsumen	Akses internet publik
Kemudahan integrasi	3	1	4	5
TCO	2	1	5	4
Skalabilitas	3	1	5	5
Kemampuan beradaptasi	3	2	4	5
Isolasi jaringan	3	4	5	1
Observabilitas	3	4	5	5
Waktu untuk memperbaiki	3	2	5	5

Menghubungkan dengan AWS Site-to-Site VPN

AWS Site-to-Site VPNkoneksi dapat berakhir baik pada gateway pribadi virtual atau gateway transit. Gateway pribadi virtual adalah titik akhir VPN di AWS sisi koneksi Site-to-Site VPN Anda yang dapat dilampirkan ke satu VPC. Transit Gateway adalah hub transit yang dapat digunakan untuk menghubungkan beberapa VPCs dan jaringan lokal. Ini juga dapat digunakan sebagai titik akhir VPN untuk AWS sisi koneksi Site-to-Site VPN. Bagian ini membahas kedua opsi.

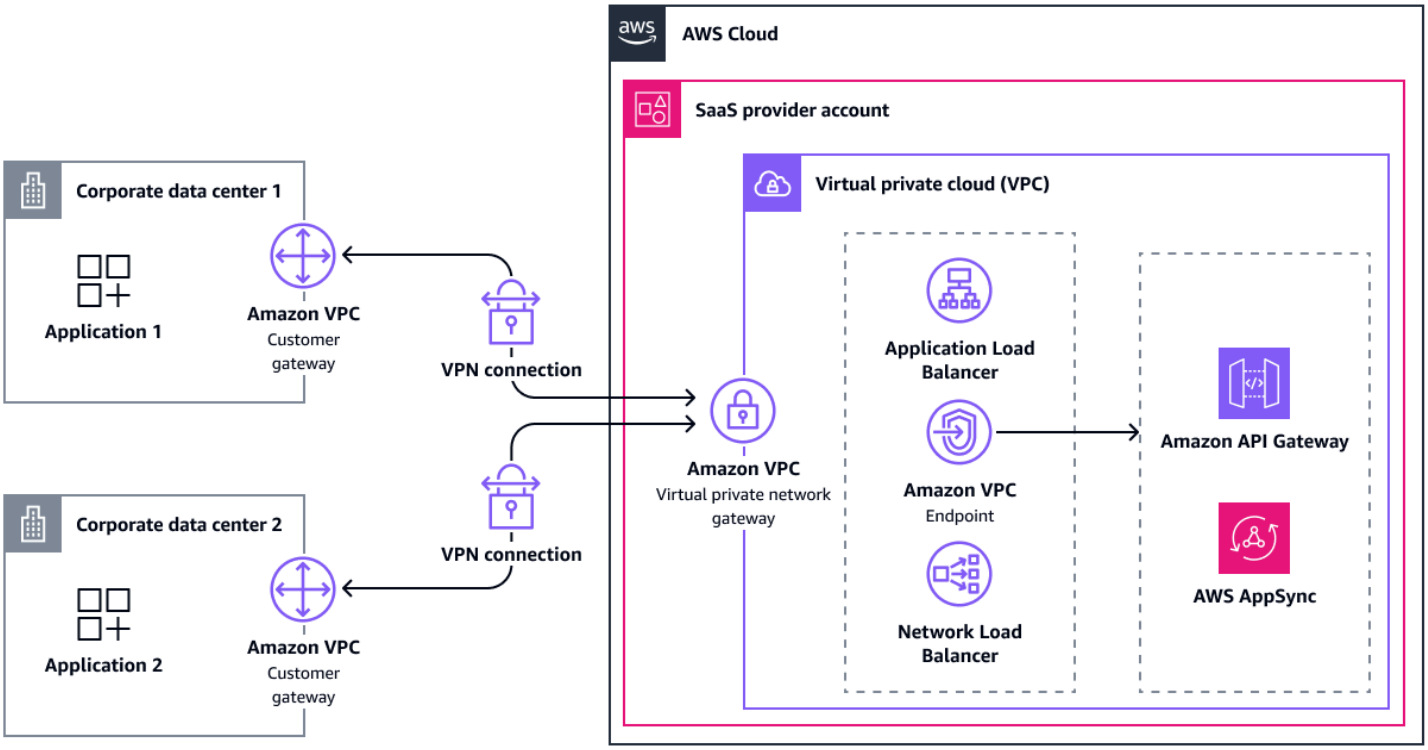
Koneksi melalui gateway pribadi virtual

Setelah Anda membuat gateway pribadi virtual, Anda melampirkannya ke VPC yang berisi penawaran SaaS Anda. Kemudian, Anda mengaktifkan propagasi rute untuk menyebarkan rute VPN ke tabel rute VPC. Rute tersebut dapat berupa rute dinamis statis atau BGP yang diiklankan.

Untuk ketersediaan tinggi, koneksi Site-to-Site VPN memiliki dua terowongan VPN yang berakhir di dua Availability Zone di AWS samping. Jika salah satu menjadi tidak tersedia, terowongan kedua dapat mengambil alih. Sebuah terowongan tunggal memungkinkan bandwidth maksimum 1,25 Gbps. Karena virtual private gateway tidak mendukung equal-cost multi-path routing (ECMP), Anda hanya dapat menggunakan satu terowongan pada satu waktu.

Untuk meningkatkan toleransi kesalahan, Anda dapat mengatur koneksi VPN kedua ke gateway pelanggan fisik kedua. Setelah koneksi dibuat, konsumen dapat mencapai sumber daya di VPC penyedia SaaS.

Diagram berikut menunjukkan arsitektur ini.



Berikut ini adalah manfaat dari pendekatan ini:

- Saatnya memperbaiki: Failover terkelola ke terowongan VPN sekunder
- Observabilitas: Integrasi untuk pemantauan aktif terkelola dengan menggunakan Network Synthetic Monitor
- Kemudahan integrasi: Dukungan perutean dinamis melalui BGP
- Adaptabilitas: Kompatibilitas dengan sebagian besar peralatan jaringan lokal
- Kemampuan beradaptasi: dukungan IPv6
- TCO: AWS Site-to-Site VPN adalah layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih sedikit upaya operasional
- TCO: Tidak ada biaya untuk gateway virtual, meskipun ada biaya untuk dua alamat publik IPv4 di masing-masing
- Isolasi jaringan: Memungkinkan komunikasi pribadi yang aman melalui internet

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemudahan integrasi: Konsumen harus mengkonfigurasi gateway pelanggan mereka

- Skalabilitas: Kurangnya dukungan ECMP membatasi bandwidth hingga 1,25 Gbps per gateway virtual
- Skalabilitas: Penskalaan terbatas karena peningkatan kompleksitas jaringan dan overhead operasional
- Adaptabilitas: IPv6 dukungan hanya untuk alamat IP bagian dalam terowongan VPN
- Kemampuan beradaptasi: Tidak ada perutean transitif
- TCO: Overhead operasional untuk memelihara, mengelola, dan mengonfigurasi berbagai koneksi VPN untuk penyedia SaaS

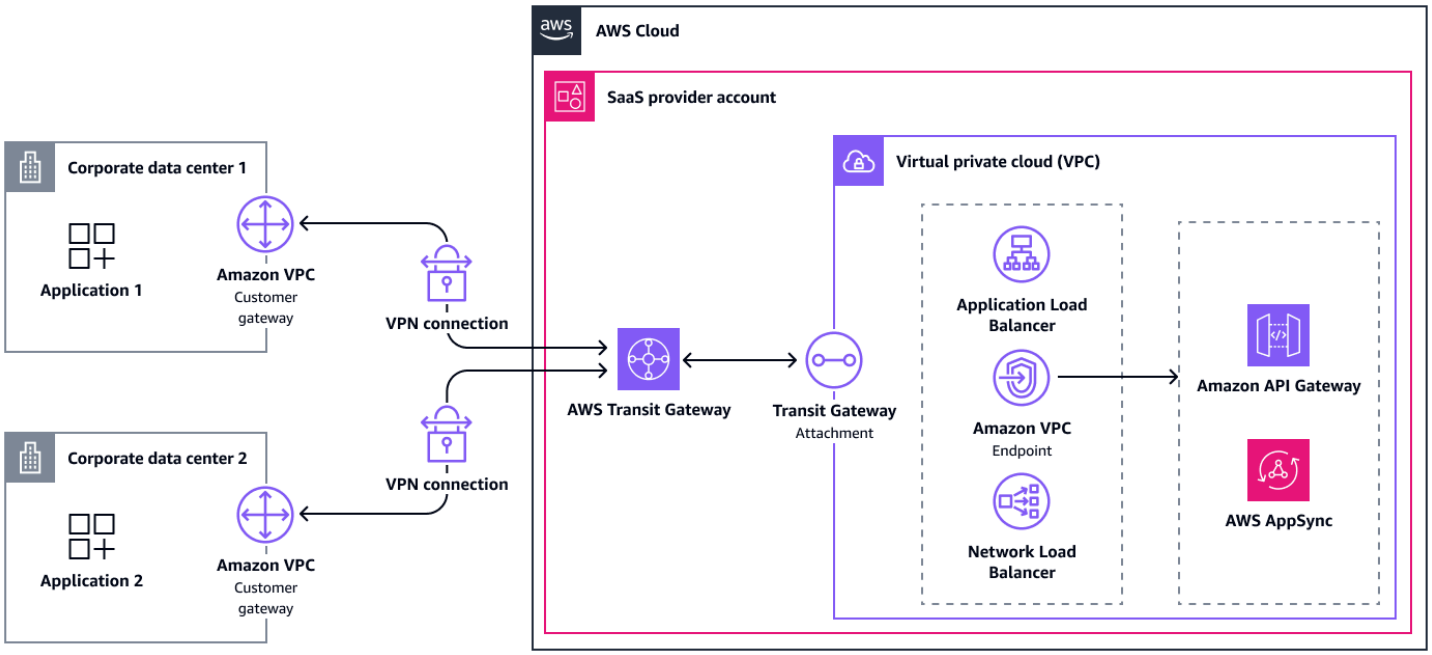
Koneksi melalui gateway transit

Koneksi melalui gateway transit mirip dengan gateway virtual. Namun, ada beberapa perbedaan yang perlu diingat.

Pertama, rute untuk lampiran VPN dapat secara otomatis disebarkan dalam tabel rute gateway transit, tetapi Anda harus menambahkan rute secara manual ke yang terlampir VPCs.

Dibandingkan dengan gateway virtual, Transit Gateway mendukung ECMP. Jika gateway pelanggan mendukung ECMP, ia dapat menggunakan kedua terowongan untuk mencapai total throughput maksimum 2,5 Gbps. Anda dapat membuat beberapa koneksi antara jaringan lokal yang sama ke gateway transit. Dengan menggunakan pendekatan ini, Anda dapat meningkatkan bandwidth maksimum hingga 2,5 Gbps per koneksi.

Diagram berikut menunjukkan arsitektur ini.



Berikut ini adalah manfaat dari pendekatan ini:

- Saatnya memperbaiki: Failover terkelola ke terowongan VPN sekunder
- Observabilitas: Integrasi untuk pemantauan aktif terkelola dengan menggunakan Network Synthetic Monitor
- Kemudahan integrasi: Dukungan perutean dinamis melalui BGP
- Skalabilitas: Dukungan ECMP memungkinkan penskalaan throughput VPN untuk memenuhi persyaratan bandwidth yang besar
- Skalabilitas: Sejumlah besar koneksi VPN yang didukung oleh gateway transit tunggal (hingga hampir 5.000)
- Skalabilitas: Satu tempat untuk mengelola dan memantau semua koneksi VPN
- Adaptabilitas: Kompatibilitas dengan sebagian besar peralatan jaringan lokal
- Kemampuan beradaptasi: dukungan IPv6
- Kemampuan beradaptasi: Mewarisi fleksibilitas AWS Transit Gateway
- TCO: AWS Transit Gateway adalah layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih sedikit upaya operasional
- TCO: Tidak ada biaya untuk gateway virtual, meskipun ada biaya untuk dua alamat publik IPv4 di masing-masing
- Isolasi jaringan: Memungkinkan komunikasi pribadi yang aman melalui internet

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemudahan integrasi: Konsumen harus mengkonfigurasi gateway pelanggan mereka
- Skalabilitas: Penskalaan terbatas karena peningkatan kompleksitas jaringan dan overhead operasional
- Adaptabilitas: IPv6 dukungan hanya untuk alamat IP bagian dalam terowongan VPN
- TCO: Overhead operasional untuk memelihara, mengelola, dan mengonfigurasi berbagai koneksi VPN untuk penyedia SaaS
- TCO: Biaya tambahan untuk penggunaan AWS Transit Gateway
- TCO: Kompleksitas tambahan mengelola tabel rute gateway transit

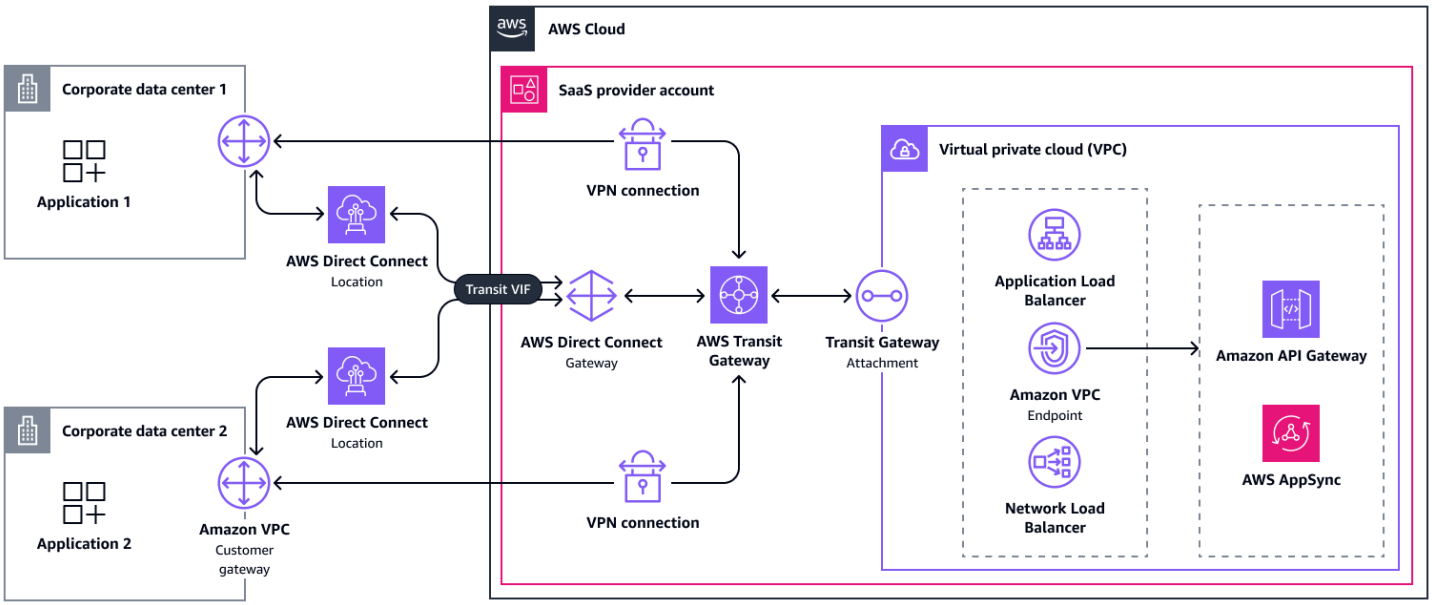
Menghubungkan dengan AWS Direct Connect

AWS Direct Connectmenghubungkan jaringan internal Anda ke Direct Connect lokasi melalui kabel serat optik Ethernet standar. Berbeda dengan opsi arsitektur lainnya, koneksi khusus tidak dapat dibuat dalam beberapa menit. Sebaliknya, proses ini bisa memakan waktu hingga beberapa hari jika semua persyaratan terpenuhi. Jika tidak, mungkin butuh waktu lebih lama. Oleh karena itu, kami menyarankan agar Anda menghubungi tim AWS akun Anda atau AWS Dukungan untuk bantuan dengan pendekatan ini. Secara opsional, Anda dapat memilih koneksi host yang disediakan oleh AWS Mitra dan dibagikan dengan pelanggan lain. Arsitekturnya sama. Anda dapat memilih Direct Connect karena mengurangi latensi, meningkatkan bandwidth, atau mematuhi persyaratan peraturan.

Untuk menggunakan Direct Connect koneksi, konsumen harus membuat antarmuka virtual publik, pribadi, atau transit. Ada berbagai pilihan arsitektur yang tersedia. Yang paling fleksibel untuk menghubungkan beberapa lokasi lokal ke lokasi AWS Cloud adalah antarmuka virtual transit yang terhubung ke Direct Connect gateway. Direct Connect Gateway adalah komponen logis global yang memungkinkan penyedia layanan untuk menghubungkan hingga enam gateway transit ke sana. Selanjutnya, Anda dapat menghubungkan hingga 30 antarmuka virtual ke gateway. Untuk skala, Anda dapat membuat Direct Connect gateway tambahan. Di akun penyedia SaaS, gateway transit kemudian dilampirkan ke VPCs, seperti yang dijelaskan sebelumnya.

Konsumen dapat terhubung menggunakan satu hingga empat Direct Connect koneksi dari total satu atau dua Direct Connect lokasi, tergantung pada tingkat ketahanan yang diinginkan. Untuk informasi selengkapnya, lihat Mengkonfigurasi Direct Connect untuk ketahanan maksimum. AWS Site-to-Site VPN Koneksi melalui internet mungkin juga berfungsi sebagai jalur cadangan berbiaya lebih rendah untuk koneksi Direct Connect . Koneksi Direct Connect khusus yang didukung dapat digunakan MACsecuntuk mengenkripsi tautan pada Lapisan 2 antara Direct Connect lokasi dan pusat data.

Adalah umum untuk memiliki koneksi Site-to-Site VPN untuk kerahasiaan tambahan data. Koneksi Site-to-Site VPN dapat berakhir pada gateway transit dengan menggunakan lampiran VPN normal. Diagram berikut menunjukkan arsitektur ini.



Berikut ini adalah manfaat dari pendekatan ini:

- Observabilitas: Integrasi untuk pemantauan aktif terkelola dengan menggunakan Network Synthetic Monitor
- Skalabilitas: Dukungan untuk peningkatan throughput bandwidth
- Kemampuan beradaptasi: dukungan IPv6
- TCO: Potensi untuk mengurangi transfer data
- TCO: Pengalaman jaringan yang konsisten
- Isolasi jaringan: Konektivitas pribadi yang dapat memenuhi persyaratan peraturan

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemudahan integrasi: Waktu dan upaya manual untuk mengatur
- Skalabilitas: Skalabilitas terbatas di luar puluhan Direct Connect koneksi karena ada beberapa kuota untuk dilacak
- Kemampuan beradaptasi: Opsi konfigurasi bergantung pada lokasi yang tersedia Direct Connect
- TCO: Direct Connect Pemeliharaan terjadwal dapat menyebabkan downtime yang memerlukan tindakan

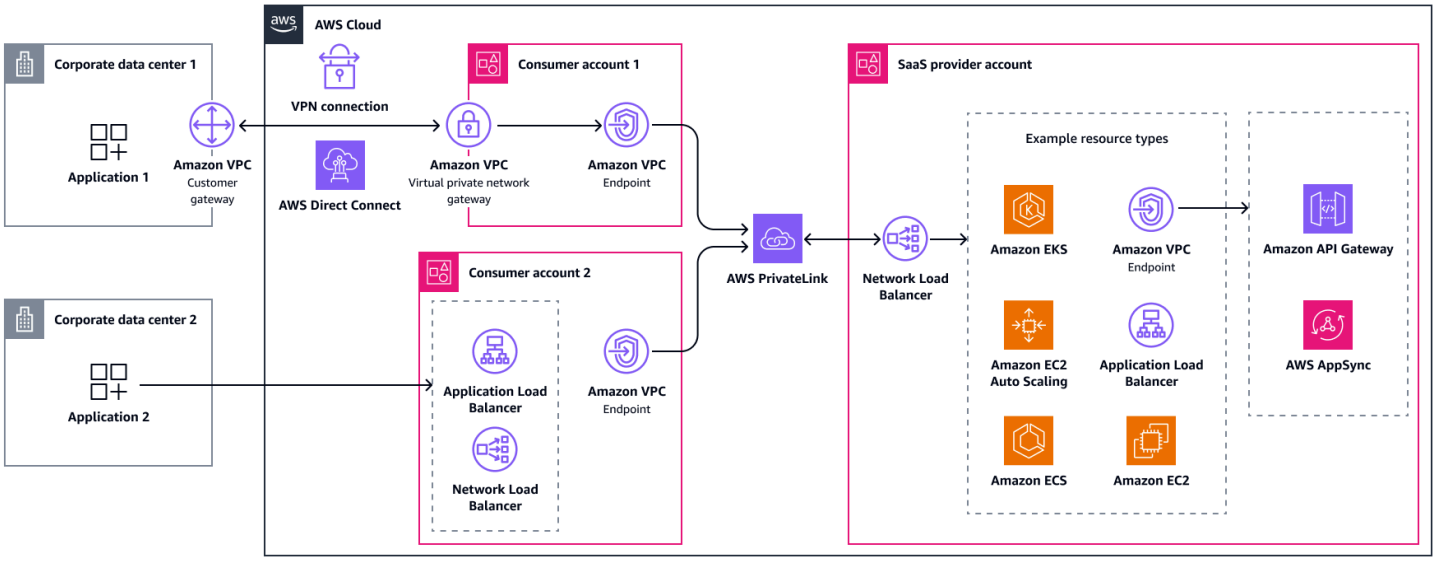
Menghubungkan dengan arsitektur VPC transit

Transit VPC adalah opsi arsitektur yang memberikan fleksibilitas kepada konsumen tentang cara terhubung AWS, dan memungkinkan penyedia SaaS mendapat manfaat dari memiliki akses terpadu ke layanan mereka melalui. AWS PrivateLink Konsumen terhubung dari tempat ke VPC transit yang hanya berisi titik masuk (seperti gateway pribadi virtual) dan titik akhir VPC antarmuka, yang merupakan sumber daya. AWS PrivateLink Transit VPCs harus dimiliki oleh penyedia SaaS atau oleh konsumen. Bagian ini membahas kedua opsi.

Anda dapat membuat VPC transit dan subnet dengan rentang CIDR yang kompatibel dengan pusat data lokal. Jika mereka memerlukan konektivitas pribadi, konsumen dapat terhubung ke VPC itu melalui AWS Direct Connect atau. AWS Site-to-Site VPN Anda juga dapat mengonfigurasi akses ke akun transit dari internet publik dengan menggunakan Application Load Balancer atau Network Load Balancer yang mengarah ke titik akhir VPC.

VPC transit yang dikelola konsumen

Dalam pendekatan ini, penyedia SaaS VPCs menyerahkan manajemen transit kepada konsumen. Dari sudut pandang teknis, arsitektur penyedia SaaS sama dengan saat menghubungkan ke konsumen secara menyeluruh. AWS Cloud AWS PrivateLink Dari perspektif penjualan dan produk, ini adalah upaya tambahan karena beberapa konsumen Akun AWS belum memilikinya. Mereka mungkin ragu untuk membuka dan mengoperasikan akun. Penyedia SaaS harus memberikan panduan kepada konsumen mereka tentang cara membuat Akun AWS dan menghubungkan pusat data lokal mereka. Diagram berikut menunjukkan campuran akses publik dan pribadi, di mana konsumen memiliki transit VPCs.



Berikut ini adalah manfaat dari pendekatan ini:

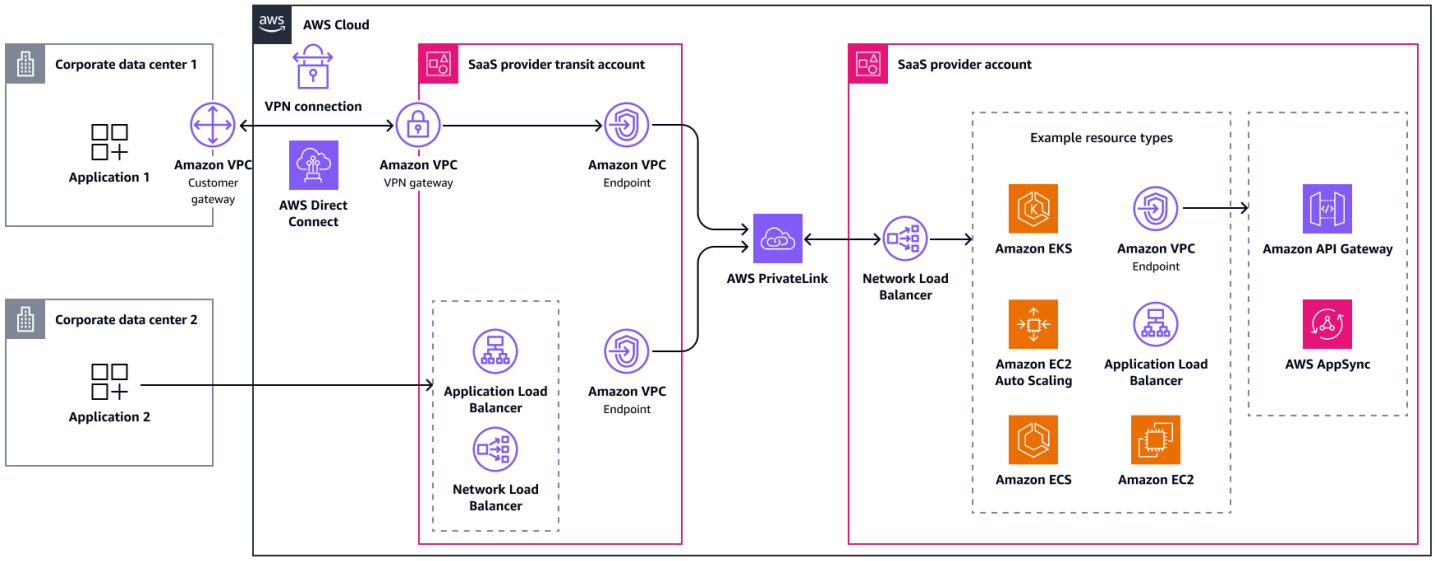
- Waktu untuk memperbaiki: Overhead operasional sebagian besar diturunkan ke konsumen SaaS
- Kemampuan beradaptasi: Konsumen SaaS dapat memilih dari berbagai opsi akses
- Kemampuan beradaptasi: Tidak ada konflik rentang CIDR, bahkan saat menggunakan VPN atau Site-to-Site Direct Connect
- Semua metrik: Penyedia layanan mewarisi manfaat AWS PrivateLink

Berikut ini adalah kelemahan dari pendekatan ini:

- Kemudahan integrasi: Konsumen SaaS membutuhkan setidaknya satu Akun AWS
- TCO: VPC transit adalah arsitektur, bukan layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih banyak upaya operasional

VPC transit yang dikelola penyedia

Pendekatan ini menggunakan teknologi yang sama, tetapi batasan akun dan tanggung jawab berubah. Di sini, penyedia SaaS memiliki transit VPCs, lebih disukai di akun terpisah dari penawaran SaaS. Decoupling ini mengurangi biaya, mengurangi risiko, dan memungkinkan akun transit untuk menskalakan secara independen. Untuk lingkungan yang memerlukan isolasi tingkat tinggi, Anda dapat membuat pemisahan tambahan antar penyewa dengan menggunakan subnet atau dengan membuat VPC transit terpisah untuk setiap konsumen. Konsumen kemudian dapat memilih cara terhubung ke VPC transit. Pendekatan ini memberikan lebih banyak opsi untuk memperluas total pasar yang dapat dialamatkan, tetapi memiliki TCO yang lebih tinggi untuk penyedia SaaS karena kebutuhan untuk mengoperasikan dan memantau komponen arsitektur tambahan.



Berikut ini adalah manfaat dari pendekatan ini:

- Kemampuan beradaptasi: Konsumen SaaS dapat memilih dari berbagai opsi akses
- Kemampuan beradaptasi: Konsumen SaaS tidak perlu memiliki Akun AWS
- Kemampuan beradaptasi: Tidak ada konflik rentang CIDR, bahkan saat menggunakan VPN atau Site-to-Site Direct Connect

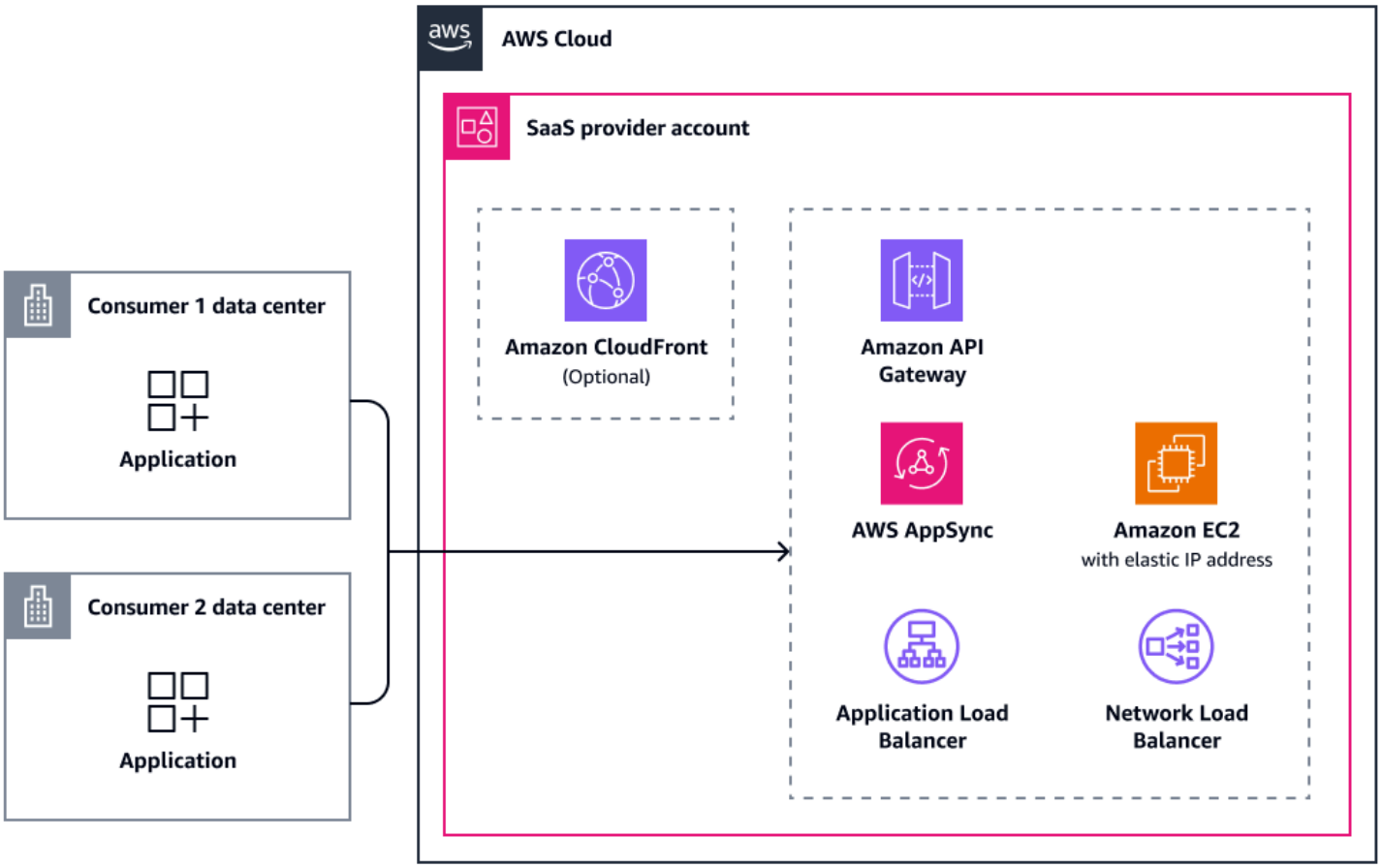
Berikut ini adalah kelemahan dari pendekatan ini:

- TCO: VPC transit adalah arsitektur, bukan layanan yang dikelola sepenuhnya, sehingga membutuhkan lebih banyak upaya operasional
- TCO: Penyedia SaaS perlu mengoperasikan dan memantau komponen arsitektur tambahan

Menghubungkan melalui internet publik

Akses internet publik juga merupakan pilihan yang valid untuk menyediakan akses ke penawaran SaaS, meskipun tidak menawarkan konektivitas pribadi dalam pengertian tradisional. Beberapa konsumen mungkin masih lebih memilih pendekatan akses publik karena tidak memerlukan infrastruktur jaringan tambahan antara mereka dan penyedia SaaS. Ini mengurangi kompleksitas, biaya, dan waktu integrasi dengan imbalan permukaan serangan yang meningkat. Mekanisme otentikasi dan otorisasi yang kuat dapat membantu mengurangi tingkat ancaman yang meningkat, dan Anda harus selalu mengenkripsi lalu lintas. Masih disarankan agar Anda memiliki lapisan keamanan tambahan dalam skenario ini, seperti dengan menggunakan AWS WAF.

Arsitektur dalam skenario ini sangat mudah. Konsumen terhubung ke host publik (penyedia SaaS) melalui internet. Aplikasi ini dapat di-host langsung di instance Amazon Elastic Compute Cloud (Amazon EC2) publik dengan alamat IP Elastis. Opsi yang lebih disukai adalah meng-hostingnya di belakang Application Load Balancer atau layanan serupa. Untuk kinerja yang lebih baik dan caching aset statis, Anda dapat menggunakan jaringan pengiriman konten, seperti Amazon CloudFront. Untuk menyajikan aplikasi dengan latensi minimum pada dua alamat IP Anycast statis global, Anda dapat menempatkan AWS Global Acceleratordi depan instans Amazon EC2, Network Load Balancer, atau Application Load Balancer. Selain itu CloudFront, Application Load Balancers, AWS AppSync, dan Amazon API Gateway semuanya terintegrasi dengan. AWS WAF Diagram berikut memberikan gambaran umum tentang opsi konektivitas akses internet publik.



Tabel berikut menjelaskan protokol dan integrasi yang didukung untuk skenario ini.

Layanan atau sumber daya	IPv6	AWS WAF integrasi	Bisa menjadi titik akhir Global Accelerator
Amazon CloudFront	Didukung	Didukung	Tidak Support

Amazon API Gateway	Didukung	Didukung	Tidak Support
AWS AppSync	Sebagian didukung	Didukung	Tidak Support
Amazon EC2 dengan alamat IP Elastis	Didukung	Tidak didukung	Didukung
Penyeimbang Beban Aplikasi	Didukung	Didukung	Didukung
Penyeimbang Beban Jaringan	Didukung	Tidak didukung	Didukung

Berikut ini adalah manfaat dari pendekatan ini:

- Kemudahan integrasi: Kesederhanaan dan aksesibilitas
- Skalabilitas: Skala tidak terbatas
- Kemampuan beradaptasi: Tidak ada konflik rentang CIDR yang mungkin
- Kemampuan beradaptasi: dukungan CloudFront

Berikut ini adalah kelemahan dari pendekatan ini:

- Isolasi jaringan: Tidak ada konektivitas pribadi
- Isolasi jaringan: Diperlukan langkah-langkah keamanan yang kuat

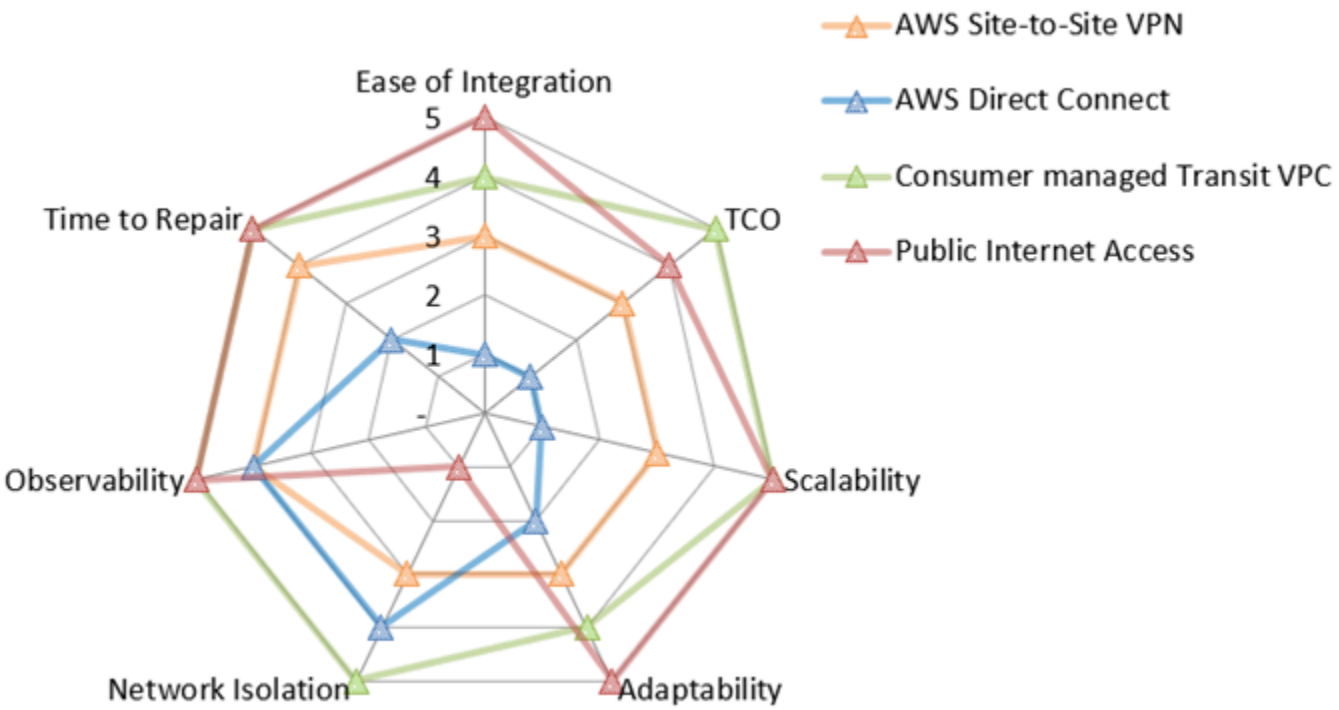
Manfaat dan kerugian lain berlaku, tergantung pada layanan yang Anda pilih.

Konsumen SaaS yang beroperasi di penyedia layanan cloud lainnya

Skenario ini menjelaskan solusi bagi konsumen di penyedia layanan cloud lainnya (CSPs). Skenario ini berbagi beberapa kesamaan dengan koneksi ke pusat data lokal. Faktanya, semua opsi konektivitas untuk lingkungan lokal sama-sama valid untuk konsumen di tempat lain CSPs, bahkan koneksi pribadi dengan AWS Direct Connect dimungkinkan dengan beberapa CSPs. Sebagian besar CSPs menawarkan dokumentasi dan dukungan tentang cara terhubung ke AWS Cloud melalui AWS Site-to-Site VPN atau AWS Direct Connect.

Saat memilih Site-to-Site VPN, konsumen dapat memperoleh manfaat dari gateway yang dikelola atau sumber daya serupa dari CSP masing-masing. Konsumen tidak perlu mengaturnya sendiri, seperti dalam skenario lokal. Ini memengaruhi beberapa metrik untuk Site-to-Site VPN, seperti peningkatan waktu untuk perbaikan dan pengamatan. Ini karena kedua ujung koneksi sekarang dikelola.

Peta nilai jaringan berikut merangkum bagaimana masing-masing opsi ini mendapat skor untuk setiap metrik evaluasi. Ini sangat mirip dengan peta nilai jaringan untuk koneksi lokal, meskipun nilai untuk Site-to-Site VPN berbeda. Untuk informasi selengkapnya tentang metrik evaluasi, lihat Metrik evaluasi di panduan ini. Dalam peta, lima mewakili skor terbaik, seperti TCO terendah, isolasi jaringan terbaik, atau waktu terendah untuk memperbaiki. Untuk informasi lebih lanjut tentang cara membaca bagan radar ini, lihat Peta nilai jaringan di panduan ini.



Bagan radar menunjukkan nilai-nilai berikut.

Metrik evaluasi	AWS Site-to-Site VPN	AWS Direct Connect	VPC transit yang dikelola konsumen	Akses internet publik
Kemudahan integrasi	3	1	4	5

TCO	3	1	5	4
Skalabilitas	3	1	5	5
Kemampuan beradaptasi	3	2	4	5
Isolasi jaringan	3	4	5	1
Observabilitas	4	4	5	5
Waktu untuk memperbaiki	4	2	5	5

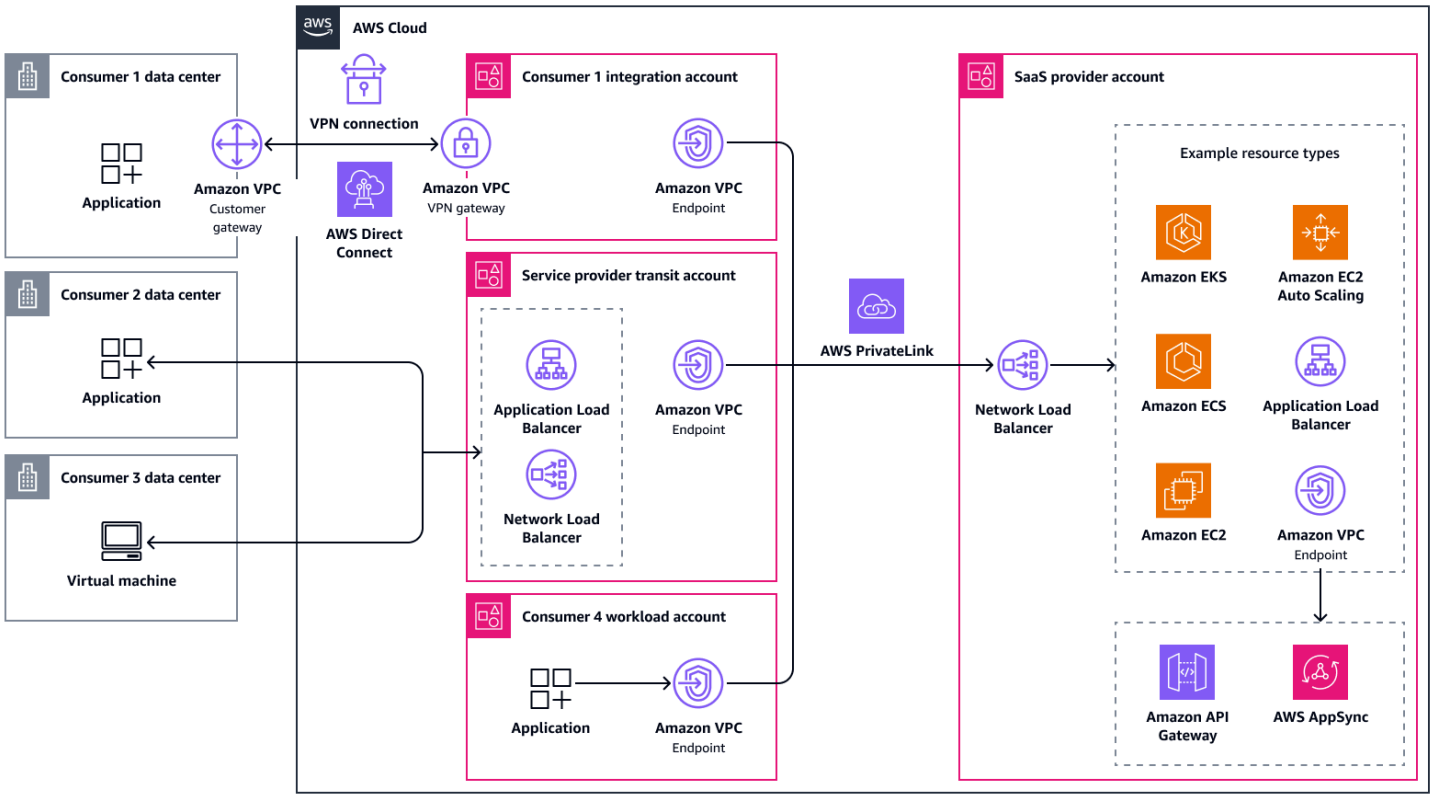
Mendukung lingkungan hibrida

Adalah umum bagi konsumen untuk datang dari lingkungan yang berbeda, masing-masing dengan kendala teknis dan keamanannya sendiri. Beberapa pelanggan dapat beroperasi sepenuhnya dari pusat data lokal yang memerlukan konektivitas aman melalui Internet atau melalui tautan jaringan khusus. Orang lain mungkin sudah menjalankan beban kerja di dalam AWS dan mengharapkan jalur jaringan pribadi latensi rendah. Kelompok ketiga mungkin bergantung pada yang lain CSPs, di mana konektivitas harus menjembatani jaringan cloud yang berbeda.

Terlepas dari itu, Anda harus bertujuan untuk akses jaringan standar ke aplikasi SaaS Anda untuk menyederhanakan arsitektur Anda dan mengurangi kompleksitas operasional. Dua dari pendekatan yang disajikan sebelumnya - akses internet publik dan transit VPCs - bekerja dengan baik di seluruh skenario ini. Akses internet publik menawarkan jalur orientasi tercepat dengan pengaturan minimal untuk pelanggan Anda. Transit VPCs menawarkan akses yang lebih terkontrol dan pribadi, sering digunakan AWS PrivateLink.

Saat merancang penawaran SaaS Anda, Anda dapat mengadopsi model akses jaringan tunggal atau menggabungkan beberapa pendekatan ke dalam penawaran berjenjang. Misalnya, Anda mungkin menawarkan tingkat penyebaran akses publik untuk pelanggan yang memprioritaskan kemudahan koneksi dan orientasi cepat, dan Anda mungkin menawarkan tingkat penyebaran akses pribadi untuk pelanggan yang memiliki kepatuhan ketat atau persyaratan kontrol keamanan. Tingkatan ini dilengkapi dengan profil biaya, kinerja, dan risiko yang berbeda. Dimungkinkan juga untuk menggabungkan kedua pendekatan menjadi satu arsitektur. Dalam hal ini, pastikan bahwa Anda memiliki langkah-langkah keamanan yang kuat sehingga jalur publik dan pribadi tetap terisolasi.

Diagram berikut menunjukkan pendekatan akses hybrid, di mana konsumen memiliki opsi untuk terhubung secara pribadi dari pusat data atau CSP mereka, secara publik, atau langsung melalui AWS PrivateLink (jika mereka memiliki beban kerja di). AWS Cloud



Skenario akses jaringan lanjutan untuk penawaran SaaS di AWS Cloud

Arsitektur yang dibahas di Skenario akses jaringan untuk penawaran SaaS di AWS Cloud bagian ini akan membantu Anda menemukan solusi untuk sebagian besar kasus penggunaan. Namun, ada beberapa skenario yang memiliki persyaratan teknis khusus. Banyak yang berada di luar cakupan panduan ini.

Bagian ini membahas persyaratan dan pertimbangan teknis lanjutan berikut:

- Komunikasi dua arah
- TCP, UDP, dan protokol berpemilik

Komunikasi dua arah

Dalam beberapa kasus, aplikasi memerlukan lalu lintas dua arah agar beroperasi seperti yang diharapkan. Kasus penggunaan umum adalah webhook atau layanan notifikasi. Umumnya, Anda dapat mencapai ini dengan memiliki WebSocket koneksi antara server dan klien. Koneksi ini membuat sesi TCP tetap terbuka dan memungkinkan kedua peserta untuk mengirim lalu lintas melalui koneksi. Sebagian besar layanan yang dibahas dalam panduan ini mendukung secara native WebSocket, termasuk Network Load Balancer, Application Load Balancer, Amazon API Gateway AWS PrivateLink, dan AWS AppSync (melalui titik akhir real-time pribadi).

Dalam kasus lain, aplikasi di sisi penyedia SaaS mungkin memerlukan akses ke sumber daya di sisi konsumen, seperti database. Ketika Anda terhubung melalui saluran dua arah, seperti AWS Site-to-Site VPN koneksi, itu bukan masalah.

Di sisi lain, AWS PrivateLink dan Elastic Load Balancing hanya mendukung lalu lintas searah. Jika Anda menggunakan layanan ini, Anda harus mengatur jalur jaringan lain untuk lalu lintas yang dimulai dari penawaran SaaS Anda. Misalnya, ini mungkin AWS PrivateLink koneksi tambahan yang berjalan ke arah sebaliknya.

TCP, UDP, dan protokol berpemilik

Banyak aplikasi dilayani melalui HTTP atau HTTPS, tetapi tidak semua. Beberapa mungkin menggunakan protokol Layer 7 lainnya di atas TCP, seperti Message Queuing Telemetry Support

(MQTT). Orang lain bahkan mungkin menggunakan UDP untuk melayani konsumen. Dalam kasus yang jarang terjadi, layanan menggunakan protokol berpemilik yang harus ditransmisikan di dalam paket (Layer 3). Untuk skenario ini, penting untuk memahami layanan mana yang mendukung penawaran SaaS Anda.

Untuk layanan Layer 3, Anda dapat menggunakan AWS PrivateLink dan Network Load Balancers, keduanya mendukung semua lalu lintas TCP dan UDP.

Untuk layanan Layer 7, Application Load Balancers dan Amazon CloudFront mendukung HTTP, HTTPS WebSocket, dan Google Remote Procedure Calls (gRPC). Demikian pula, Amazon API Gateway dan AWS AppSync masing-masing mendukung HTTP, HTTPS, dan WebSocket. Amazon CloudFront adalah satu-satunya layanan yang saat ini mendukung HTTP/3.

Anda dapat menggunakan Amazon VPC Lattice untuk menghubungkan aplikasi Layer 7 dan sumber daya Layer 3. Ini mendukung HTTP, HTTPS, gRPC, TCP, dan TLS passthrough.

Jika aplikasi dapat melayani lalu lintas hanya melalui Layer 3, sangat penting bahwa Anda menggunakan layanan AWS jaringan inti, seperti, AWS Transit Gateway, AWS Direct Connect AWS Site-to-Site VPN, dan VPC peering. Lalu lintas kemudian harus diarahkan langsung dari konsumen SaaS ke lapisan komputasi penawaran SaaS.

Anti-pola untuk akses jaringan di AWS Cloud

Anti-pola adalah solusi yang sering digunakan untuk masalah berulang di mana solusinya kontra-produktif, tidak efektif, atau kurang efektif daripada alternatif. Opsi desain yang disebutkan dalam bagian ini biasanya berfungsi, tetapi mereka datang dengan kerugian yang signifikan. Jika memungkinkan, mereka harus dihindari karena alternatif yang lebih baik tersedia.

Bagian ini membahas anti-pola dan tantangan berikut:

- Ketidakcocokan Zona Ketersediaan dengan AWS PrivateLink
- AWS Site-to-Site VPN hubungan antara Akun AWS

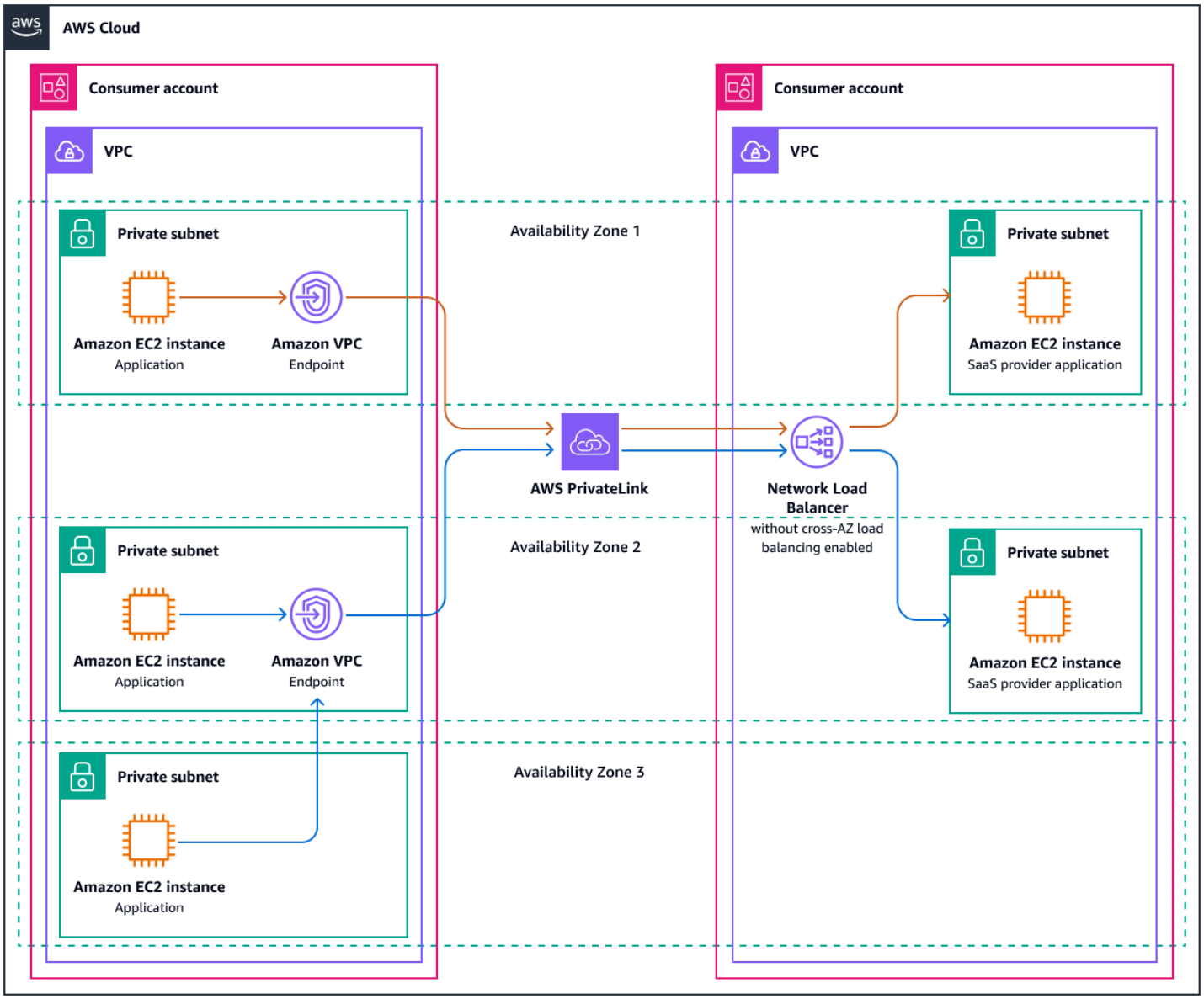
Ketidakcocokan Zona Ketersediaan dengan AWS PrivateLink

Saat menyediakan akses ke aplikasi melalui AWS PrivateLink, konsumen SaaS dapat membuat titik akhir VPC antarmuka hanya di Availability Zones tempat aplikasi digunakan. Misalnya, jika aplikasi diterapkan di `use1-az1` dan`use1-az2`, konsumen tidak dapat menerapkan titik akhir VPC di. `use1-az3` Kami menyarankan Anda menerapkan penawaran SaaS di setiap Availability Zone. Mayoritas Wilayah AWS memiliki tiga Availability Zone, meskipun beberapa memiliki lebih banyak. Untuk daftar lengkap, lihat Wilayah dan Availability Zone. Pertimbangkan jumlah Availability Zone saat memilih Wilayah AWS.

> **Note**
>
> Nama zona ketersediaan berbeda dari Availability Zone IDs. Untuk informasi selengkapnya, lihat Availability Zone IDs untuk AWS sumber daya Anda.

Jika penyedia SaaS memilih untuk tidak menerapkan di semua Availability Zone, ada beberapa konsekuensinya. Asumsikan penawaran SaaS diterapkan di `use1-az1` dan`use1-az2`, tetapi konsumen menggunakan ketiga Availability Zone, termasuk. `use1-az3` Titik akhir VPC antarmuka digunakan di sisi konsumen di `use1-az1` dan`use1-az2`, dan sekarang aplikasi `use1-az3` perlu mengakses salah satu titik akhir ini. Pertama-tama, lalu lintas harus diizinkan dari subnet di Availability Zone yang tak tertandingi ke titik akhir VPC masing-masing. Konsumen dapat memutuskan untuk menggunakan nama AWS PrivateLink DNS regional, yang dapat menyelesaikan ke titik akhir VPC dan yang mendistribusikan lalu lintas secara merata di antara keduanya. Atau konsumen mungkin memilih untuk mengirim lalu lintas langsung ke titik akhir, seperti`use1-az2`.

Ini menghasilkan 67% lalu lintas yang tiba di sisi penyedia `use1-az2` dan 33% masuk. `use1-az1` Gambar berikut menggambarkan skenario ini.



Dengan jumlah konsumen yang signifikan dan distribusi lalu lintas yang tidak merata, beban kerja mungkin mengalami masalah kapasitas di satu Availability Zone dan berada di bawah kapasitas di zona lain. Untuk mengatasi masalah itu, penyedia SaaS dapat memutuskan untuk memuat keseimbangan lalu lintas secara merata di sisi mereka dengan mengaktifkan penyeimbangan beban lintas zona pada Network Load Balancer. Ini menimbulkan biaya tambahan.

Jika hanya satu Availability Zone yang dicocokkan oleh penyedia layanan, maka semua lalu lintas akan masuk melalui satu titik akhir. Ini menciptakan ketidakseimbangan yang lebih besar. Akibatnya, penawaran SaaS tidak lagi tersedia untuk konsumen. Tidak masalah bagi konsumen jika aplikasi

disajikan melalui Availability Zone tambahan yang tidak mereka gunakan sendiri. Dalam kasus terburuk, penyedia SaaS mungkin tidak dapat melayani konsumen yang tidak menggunakan Zona Ketersediaan yang sama.

Dalam kasus yang jarang terjadi bahwa tidak ada opsi yang layak bagi penyedia SaaS untuk menyediakan aplikasi mereka di semua Availability Zone, dimungkinkan juga untuk membuat subnet hanya di Availability Zone yang hilang dan kemudian memperluas layanan ke Availability Zone kosong tersebut. Penyeimbangan beban lintas zona kemudian dapat mendistribusikan lalu lintas masuk melalui titik akhir aplikasi aktual di Availability Zone lainnya.

AWS Site-to-Site VPN hubungan antara Akun AWS

Perusahaan yang bermigrasi dari lingkungan lokal ke cloud terkadang mencoba mengangkat dan menggeser seluruh jaringan. Hal ini dapat menyebabkan masalah karena ada perbedaan yang signifikan antara praktik jaringan lokal dan cloud. Jika pergeseran pola pikir ini tidak terjadi, hal-hal seperti AWS Site-to-Site VPN koneksi dari satu VPC ke VPC lain dapat terjadi. Pendekatan ini gagal memanfaatkan layanan jaringan yang dibangun khusus di AWS Cloud, yang menyederhanakan manajemen dan meningkatkan kinerja. Beradaptasi dengan desain cloud-native membantu mengurangi overhead operasional dan menghasilkan konektivitas yang lebih andal dan terukur di antaranya. VPCs

Jika Anda berpikir untuk menyediakan opsi konektivitas ini sebagai penyedia SaaS, tanyakan pada diri Anda atau konsumen mengapa AWS Site-to-Site VPN harus digunakan. Kemudian, bekerja mundur dari persyaratan tersebut untuk menemukan opsi konektivitas yang lebih baik. Bagian Membandingkan kemampuan layanan dari panduan ini berisi matriks yang dapat Anda gunakan untuk membantu mengidentifikasi opsi. Kemudian, Anda dapat mengerjakan bagian yang relevan dari panduan ini untuk menemukan pendekatan arsitektur yang membahas kasus penggunaan Anda.

Langkah selanjutnya

Panduan ini menjelaskan berbagai pendekatan akses jaringan dalam skenario yang berbeda, dan ini menjelaskan manfaat dan kelemahan masing-masing arsitektur. Anda harus memahami mengapa memilih pendekatan akses jaringan seharusnya tidak menjadi diskusi teknologi murni. Keselarasan antara bisnis dan teknologi sangat penting. Langkah dan rekomendasi berikut dapat membantu Anda menilai dan membakukan strategi arsitektur jaringan Anda dengan mengevaluasi kemampuan saat ini, menganalisis kebutuhan pasar, dan menerapkan kontrol tata kelola.

Bagian ini berisi topik berikut:

- Menilai arsitektur dan kemampuan saat ini
- Analisis pasar dan pelanggan
- Penyelarasan strategis
- Standardisasi
- Tata kelola
- Pengulangan

Menilai arsitektur dan kemampuan saat ini

Tinjau arsitektur jaringan saat ini terhadap sumber data yang relevan, seperti kerangka penilaian diri dalam panduan ini, persyaratan peraturan saat ini, dan keadaan pasar saat ini (baik dalam hal pelanggan Anda maupun analisis kompetitif). Misalnya, pertimbangkan untuk menggunakan AWS Well-Architected Framework, yang didasarkan pada pengalaman puluhan tahun menjalankan sistem produksi dalam skala besar di. AWS Cloud

Tinjau setiap pengecualian potensial, satu kali, dan keputusan produk historis. Jadilah penasaran, tantang mereka, dan jangan secara otomatis menganggap validitasnya. Persyaratan pelanggan dari tahun lalu mungkin tidak lagi berlaku. Asumsi yang menantang menciptakan peluang untuk menyederhanakan dan mengurangi kompleksitas arsitektur Anda.

Secara sederhana, dokumentasikan pengamatan sehingga dapat diakses dan dipahami oleh beragam peran dalam organisasi Anda. Tangkap di mana keadaan saat ini berbeda dari keadaan target, apa keadaan target, dampaknya, dan kapan pengamatan dilakukan. Merekam informasi ini membantu organisasi Anda membuat keputusan berdasarkan data baru.

Analisis pasar dan pelanggan

Kumpulkan wawasan tentang tren pasar. Apa cara konsumen yang saat ini disukai untuk mengakses penawaran SaaS seperti milik Anda? Apakah Anda masih bertemu pelanggan Anda di mana mereka berada? Apakah kelompok atau perilaku pelanggan berubah? Apakah eksekutif Anda mengarahkan kapal menuju pasar baru, geografi dengan persyaratan peraturan khusus, atau tingkat pelanggan baru? Apakah bisnis Anda, atau model operasi berubah? Misalnya, apakah Anda mempertimbangkan untuk memberi label putih pada layanan Anda? Apakah rencana pertumbuhan Anda termasuk bekerja dengan mitra sehingga layanan Anda tersedia bagi pelanggan ketika mereka terhubung dengan mitra tersebut?

Penyelarasan strategis

Ketika Anda memahami kemampuan Anda saat ini, arsitektur saat ini, pasar, dan pelanggan, hubungi pertemuan penyelarasan strategis. Dengan pemangku kepentingan produk, bisnis, dan teknologi yang relevan, tantang persyaratan mana yang masih berlaku dan persyaratan baru mana yang perlu dipertimbangkan. Temukan peluang untuk mengurangi kompleksitas dengan menjatuhkan persyaratan yang tidak lagi diperlukan. Ini bukan desain oleh komite; tim teknik perlu mempersiapkan dan memiliki detail arsitektur dan implementasi yang sebenarnya. Namun, pertemuan ini harus menjelaskan mengapa ini adalah serangkaian persyaratan yang memaksimalkan manfaat bagi pelanggan dan organisasi Anda.

Standardisasi

Untuk menarik pelanggan, mungkin tergoda untuk membiarkan masing-masing dengan bebas memilih cara terhubung ke layanan Anda. Bagaimanapun, solusi apa pun mungkin bekerja secara teknis, dan Anda mungkin juga memiliki pengetahuan dan sumber daya untuk mengelola dan mengoperasikan semuanya. Ini dapat bekerja dengan baik sampai titik tertentu, tetapi seiring skala bisnis Anda, menjadi sulit untuk dikelola. Tumpukan observabilitas Anda perlu mendukung metrik dari berbagai solusi, dan teknisi keandalan situs Anda juga harus dapat memahaminya. Anda memerlukan up-to-date dokumentasi untuk setiap pendekatan konektivitas. Perubahan besar dalam aplikasi Anda perlu dievaluasi terhadap setiap pendekatan akses yang Anda tawarkan. Anda perlu menulis dan memelihara otomatisasi dan infrastruktur sebagai kode (IAc) untuk setiap pendekatan akses. Overhead tambahan untuk tidak menstandarisasi akses ke layanan Anda harus dipertimbangkan terhadap fleksibilitas yang ingin Anda tawarkan kepada pelanggan Anda.

Jika Anda membutuhkan bintang utara untuk memandu pengambilan keputusan Anda, kami sarankan standardisasi. Standarisasi bagaimana pelanggan Anda berinteraksi dengan layanan yang Anda berikan biasanya merupakan satu-satunya tindakan paling berdampak yang dapat Anda ambil untuk meningkatkan banyak metrik keberhasilan di seluruh organisasi Anda. Standardisasi memudahkan tim produk untuk memahami struktur biaya layanan Anda dan membuat keputusan produk berbasis data. Lebih mudah bagi tim operasi untuk memecahkan masalah dan mengotomatiskan bagian dari proses pemecahan masalah di lingkungan yang dikembangkan, diluncurkan, dan dioperasikan sesuai dengan standar yang telah ditentukan. Ini dapat membantu Anda mendeteksi anomali, perilaku tak terduga, atau tindakan oleh aktor jahat. Standardisasi juga mengurangi utang teknis. Dibutuhkan siklus yang lebih sedikit bagi tim teknik untuk menguji dan meluncurkan perubahan pada produksi. Ini juga dapat meningkatkan kecepatan Anda ke pasar, meningkatkan keberhasilan orientasi layanan mandiri, dan mengurangi risiko peraturan.

Oleh karena itu, kami sarankan Anda juga meninjau satu kali yang mungkin ada hari ini. Hitung jumlah siklus operasional yang Anda habiskan untuk mendukung pelanggan yang sudah ada. Bandingkan hasil Anda dengan data historis, dan nilai apakah skala pendekatan Anda saat ini untuk tahun-tahun mendatang. Setiap kali ada kebutuhan untuk mengalihkan dari standar, tantang persyaratan di balik permintaan tersebut. Mengevaluasi dampak, dan menyeimbangkan manfaat langsung dengan komitmen jangka panjang.

Dalam kasus di mana penyesuaian tidak dapat dihindari tetapi bertentangan dengan standar Anda, pertimbangkan model tanggung jawab bersama. Dalam model ini, produk Anda sebagian besar terlindung dari perubahan yang diminta, dan penyesuaian terjadi di lingkungan minimalis dan berdedikasi. Sebagai contoh, lihat Menghubungkan dengan arsitektur VPC transit bagian.

Tata kelola

Untuk memenuhi persyaratan peraturan dan standar internal Anda sendiri, tata kelola sangat penting. Dengan tata kelola yang tepat, Anda dapat mengontrol di mana dan bagaimana menegakkan standar. Anda juga menetapkan kontrol untuk mendeteksi perbedaan dari standar dan memberi tahu pemilik sumber daya tentang tindakan korektif yang diperlukan. AWS Organizations, AWS Config, AWS CloudTrail, dan AWS Control Towerbeberapa dari banyak Layanan AWS yang dapat membantu Anda mengelola dan mengatur beban kerja Anda di. AWS Cloud

Pengulangan

Dengan menggunakan pembelajaran dari upaya awal Anda, siapkan proses yang ringan dan berulang agar tetap selaras di masa depan. Tentukan peran mana yang Anda butuhkan input,

seberapa sering, seberapa akurat data yang dibutuhkan, bagaimana data akan dibagikan, dan siapa yang akan menindaklanjutinya.

Sumber daya

AWS dokumentasi

- Mengintegrasikan layanan pihak ketiga dalam AWS Cloud(PanduanAWS Preskriptif)
- Otorisasi SaaS multi-penyewa dan kontrol akses API (Panduan Preskriptif)AWS
- Kelola penyewa di beberapa produk SaaS pada satu bidang kontrol AWS (Panduan Preskriptif)
- Apa itu AWS Direct Connect? (Direct Connect dokumentasi)
- Apa itu AWS PrivateLink? (Dokumentasi Amazon VPC)
- Apa itu AWS Site-to-Site VPN? (AWS Site-to-Site VPN dokumentasi)
- Apa itu AWS Transit Gateway? (Dokumentasi Amazon VPC)
- Apa itu VPC peering? (Dokumentasi Amazon VPC)

AWS Sumber daya lainnya

- Opsi Konektivitas Amazon Virtual Private Cloud (AWS Whitepaper)
- AWS re:invent 2021 - Bagaimana memilih penyeimbang beban yang tepat untuk beban kerja Anda () AWS YouTube
- Apa itu SaaS? (AWS situs web)
- AWS Program Pabrik SaaS (program)AWS Partner
- Panduan untuk Arsitektur Multi-Tenant pada AWS(Perpustakaan Solusi)AWS

Riwayat dokumen

Tabel berikut menjelaskan perubahan signifikan pada panduan ini. Jika Anda ingin diberi tahu tentang pembaruan masa depan, Anda dapat berlangganan umpan RSS.

Perubahan	Deskripsi	Tanggal
Publikasi awal	—	September 12, 2025

AWS Glosarium Panduan Preskriptif

Berikut ini adalah istilah yang umum digunakan dalam strategi, panduan, dan pola yang disediakan oleh Panduan AWS Preskriptif. Untuk menyarankan entri, silakan gunakan tautan Berikan umpan balik di akhir glosarium.

Nomor

7 Rs

Tujuh strategi migrasi umum untuk memindahkan aplikasi ke cloud. Strategi ini dibangun di atas 5 Rs yang diidentifikasi Gartner pada tahun 2011 dan terdiri dari yang berikut:

- Refactor/re-architect — Pindahkan aplikasi dan modifikasi arsitekturnya dengan memanfaatkan sepenuhnya fitur cloud-native untuk meningkatkan kelincahan, kinerja, dan skalabilitas. Ini biasanya melibatkan porting sistem operasi dan database. Contoh: Migrasikan database Oracle lokal Anda ke Amazon Aurora Edition. PostgreSQL-Compatible
- Replatform (angkat dan bentuk ulang) — Pindahkan aplikasi ke cloud, dan perkenalkan beberapa tingkat pengoptimalan untuk memanfaatkan kemampuan cloud. Contoh: Memigrasikan database Oracle lokal Anda ke Amazon Relational Database Service (Amazon RDS) untuk Oracle di. AWS Cloud
- Pembelian kembali (drop and shop) - Beralih ke produk yang berbeda, biasanya dengan beralih dari lisensi tradisional ke model SaaS. Contoh: Migrasikan sistem manajemen hubungan pelanggan (CRM) Anda ke. Salesforce.com
- Rehost (lift dan shift) — Pindahkan aplikasi ke cloud tanpa membuat perubahan apa pun untuk memanfaatkan kemampuan cloud. Contoh: Migrasikan database Oracle lokal Anda ke Oracle pada instans EC2 di. AWS Cloud
- Relokasi (hypervisor-level lift and shift) — Pindahkan infrastruktur ke cloud tanpa membeli perangkat keras baru, menulis ulang aplikasi, atau memodifikasi operasi yang ada. Anda memigrasikan server dari platform lokal ke layanan cloud untuk platform yang sama. Contoh: Migrasikan Microsoft Hyper-V aplikasi ke AWS.
- Pertahankan (kunjungi kembali) - Simpan aplikasi di lingkungan sumber Anda. Ini mungkin termasuk aplikasi yang memerlukan refactoring besar, dan Anda ingin menunda pekerjaan itu sampai nanti, dan aplikasi lama yang ingin Anda pertahankan, karena tidak ada pembenaran bisnis untuk memigrasikannya.

- Pensiun — Menonaktifkan atau menghapus aplikasi yang tidak lagi diperlukan di lingkungan sumber Anda.

A

A2A () Agent-to-Agent

Protokol stateful untuk kolaborasi agen-ke-agen yang mendukung delegasi tugas dan transfer negara.

ABAC

Lihat kontrol akses berbasis atribut.

layanan abstrak

Lihat layanan terkelola.

ASAM

Lihat atomisitas, konsistensi, isolasi, daya tahan.

migrasi aktif-aktif

Metode migrasi database di mana basis data sumber dan target tetap sinkron (dengan menggunakan alat replikasi dua arah atau operasi penulisan ganda), dan kedua database menangani transaksi dari menghubungkan aplikasi selama migrasi. Metode ini mendukung migrasi dalam batch kecil yang terkontrol alih-alih memerlukan pemotongan satu kali. Ini lebih fleksibel tetapi membutuhkan lebih banyak pekerjaan daripada migrasi aktif-pasif.

migrasi aktif-pasif

Metode migrasi database di mana database sumber dan target disimpan dalam sinkron, tetapi hanya database sumber yang menangani transaksi dari menghubungkan aplikasi sementara data direplikasi ke database target. Basis data target tidak menerima transaksi apa pun selama migrasi.

Agen

Sistem AI yang dapat secara mandiri bernalar, merencanakan, dan mengambil tindakan menggunakan alat untuk mencapai tujuan.

Agen Ops

Praktik operasional untuk membangun, menguji, menyebarkan, dan menjalankan agen AI dalam produksi dalam skala besar.

fungsi agregat

Fungsi SQL yang beroperasi pada sekelompok baris dan menghitung nilai pengembalian tunggal untuk grup. Contoh fungsi agregat meliputi `SUM` dan`MAX`.

AI

Lihat kecerdasan buatan.

AIOps

Lihat operasi kecerdasan buatan.

anonimisasi

Proses menghapus informasi pribadi secara permanen dalam kumpulan data. Anonimisasi dapat membantu melindungi privasi pribadi. Data anonim tidak lagi dianggap sebagai data pribadi.

anti-pola

Solusi yang sering digunakan untuk masalah berulang di mana solusinya kontra-produktif, tidak efektif, atau kurang efektif daripada alternatif.

kontrol aplikasi

Pendekatan keamanan yang memungkinkan penggunaan hanya aplikasi yang disetujui untuk membantu melindungi sistem dari malware.

portofolio aplikasi

Kumpulan informasi rinci tentang setiap aplikasi yang digunakan oleh organisasi, termasuk biaya untuk membangun dan memelihara aplikasi, dan nilai bisnisnya. Informasi ini adalah kunci untuk penemuan portofolio dan proses analisis dan membantu mengidentifikasi dan memprioritaskan aplikasi yang akan dimigrasi, dimodernisasi, dan dioptimalkan.

kecerdasan buatan (AI)

Bidang ilmu komputer yang didedikasikan untuk menggunakan teknologi komputasi untuk melakukan fungsi kognitif yang biasanya terkait dengan manusia, seperti belajar, memecahkan masalah, dan mengenali pola. Untuk informasi lebih lanjut, lihat Apa itu Kecerdasan Buatan?

operasi kecerdasan buatan (AIOps)

Proses menggunakan teknik pembelajaran mesin untuk memecahkan masalah operasional, mengurangi insiden operasional dan intervensi manusia, dan meningkatkan kualitas layanan. Untuk informasi selengkapnya tentang cara AIOps digunakan dalam strategi AWS migrasi, lihat panduan integrasi operasi.

enkripsi asimetris

Algoritma enkripsi yang menggunakan sepasang kunci, kunci publik untuk enkripsi dan kunci pribadi untuk dekripsi. Anda dapat berbagi kunci publik karena tidak digunakan untuk dekripsi, tetapi akses ke kunci pribadi harus sangat dibatasi.

atomisitas, konsistensi, isolasi, daya tahan (ACID)

Satu set properti perangkat lunak yang menjamin validitas data dan keandalan operasional database, bahkan dalam kasus kesalahan, kegagalan daya, atau masalah lainnya.

kontrol akses berbasis atribut (ABAC)

Praktik membuat izin berbutir halus berdasarkan atribut pengguna, seperti departemen, peran pekerjaan, dan nama tim. Untuk informasi selengkapnya, lihat ABAC untuk AWS dokumentasi AWS Identity and Access Management (IAM).

sumber data otoritatif

Lokasi di mana Anda menyimpan versi utama data, yang dianggap sebagai sumber informasi yang paling dapat diandalkan. Anda dapat menyalin data dari sumber data otoritatif ke lokasi lain untuk tujuan memproses atau memodifikasi data, seperti menganonimkan, menyunting, atau membuat nama samaran.

Zona Ketersediaan

Lokasi berbeda di dalam Wilayah AWS yang terisolasi dari kegagalan di Availability Zone lainnya dan menyediakan konektivitas jaringan latensi rendah yang murah ke Availability Zone lainnya di Wilayah yang sama.

AWS Kerangka Adopsi Cloud (AWS CAF)

Kerangka pedoman dan praktik terbaik AWS untuk membantu organisasi mengembangkan rencana yang efisien dan efektif untuk bergerak dengan sukses ke cloud. AWS CAF mengatur panduan ke dalam enam area fokus yang disebut perspektif: bisnis, orang, tata kelola, platform, keamanan, dan operasi. Perspektif bisnis, orang, dan tata kelola fokus pada keterampilan dan proses bisnis; perspektif platform, keamanan, dan operasi fokus pada keterampilan dan proses teknis. Misalnya, perspektif masyarakat menargetkan pemangku kepentingan yang menangani

sumber daya manusia (SDM), fungsi kepegawaian, dan manajemen orang. Untuk perspektif ini, AWS CAF memberikan panduan untuk pengembangan, pelatihan, dan komunikasi orang untuk membantu mempersiapkan organisasi untuk adopsi cloud yang sukses. Untuk informasi lebih lanjut, lihat situs web AWS CAF dan whitepaper AWS CAF.

AWS Kerangka Kualifikasi Beban Kerja (AWS WQF)

Alat yang mengevaluasi beban kerja migrasi database, merekomendasikan strategi migrasi, dan memberikan perkiraan kerja. AWS WQF disertakan dengan AWS Schema Conversion Tool ()AWS SCT. Ini menganalisis skema database dan objek kode, kode aplikasi, dependensi, dan karakteristik kinerja, dan memberikan laporan penilaian.

B

bot buruk

Bot yang dimaksudkan untuk mengganggu atau menyebabkan kerugian bagi individu atau organisasi.

BCP

Lihat perencanaan kontinuitas bisnis.

grafik perilaku

Pandangan interaktif yang terpadu tentang perilaku dan interaksi sumber daya dari waktu ke waktu. Anda dapat menggunakan grafik perilaku dengan Amazon Detective untuk memeriksa upaya logon yang gagal, panggilan API yang mencurigakan, dan tindakan serupa. Untuk informasi selengkapnya, lihat Data dalam grafik perilaku di dokumentasi Detektif.

sistem big-endian

Sistem yang menyimpan byte paling signifikan terlebih dahulu. Lihat juga endianness.

klasifikasi biner

Sebuah proses yang memprediksi hasil biner (salah satu dari dua kelas yang mungkin). Misalnya, model ML Anda mungkin perlu memprediksi masalah seperti “Apakah email ini spam atau bukan spam?” atau “Apakah produk ini buku atau mobil?”

filter mekar

Struktur data probabilistik dan efisien memori yang digunakan untuk menguji apakah suatu elemen adalah anggota dari suatu himpunan.

blue/green penyebaran

Strategi penyebaran tempat Anda membuat dua lingkungan yang terpisah namun identik. Anda menjalankan versi aplikasi saat ini di satu lingkungan (biru) dan versi aplikasi baru di lingkungan lain (hijau). Strategi ini membantu Anda dengan cepat memutar kembali dengan dampak minimal.

bot

Aplikasi perangkat lunak yang menjalankan tugas otomatis melalui internet dan mensimulasikan aktivitas atau interaksi manusia. Beberapa bot berguna atau bermanfaat, seperti perayap web yang mengindeks informasi di internet. Beberapa bot lain, yang dikenal sebagai bot buruk, dimaksudkan untuk mengganggu atau membahayakan individu atau organisasi.

botnet

Jaringan bot yang terinfeksi oleh malware dan berada di bawah kendali satu pihak, yang dikenal sebagai bot herder atau operator bot. Botnet adalah mekanisme paling terkenal untuk skala bot dan dampaknya.

cabang

Area berisi repositori kode. Cabang pertama yang dibuat dalam repositori adalah cabang utama. Anda dapat membuat cabang baru dari cabang yang ada, dan Anda kemudian dapat mengembangkan fitur atau memperbaiki bug di cabang baru. Cabang yang Anda buat untuk membangun fitur biasanya disebut sebagai cabang fitur. Saat fitur siap dirilis, Anda menggabungkan cabang fitur kembali ke cabang utama. Untuk informasi selengkapnya, lihat Tentang cabang (GitHub dokumentasi).

akses break-glass

Dalam keadaan luar biasa dan melalui proses yang disetujui, cara cepat bagi pengguna untuk mendapatkan akses ke Akun AWS yang biasanya tidak memiliki izin untuk mengaksesnya. Untuk informasi lebih lanjut, lihat indikator Implementasikan prosedur break-glass dalam panduan. AWS Well-Architected

strategi brownfield

Infrastruktur yang ada di lingkungan Anda. Saat mengadopsi strategi brownfield untuk arsitektur sistem, Anda merancang arsitektur di sekitar kendala sistem dan infrastruktur saat ini. Jika Anda memperluas infrastruktur yang ada, Anda dapat memadukan strategi brownfield dan greenfield.

cache penyangga

Area memori tempat data yang paling sering diakses disimpan.

kemampuan bisnis

Apa yang dilakukan bisnis untuk menghasilkan nilai (misalnya, penjualan, layanan pelanggan, atau pemasaran). Arsitektur layanan mikro dan keputusan pengembangan dapat didorong oleh kemampuan bisnis. Untuk informasi selengkapnya, lihat bagian Terorganisir di sekitar kemampuan bisnis dari Menjalankan layanan mikro kontainer di whitepaper. AWS

perencanaan kelangsungan bisnis (BCP)

Rencana yang membahas dampak potensial dari peristiwa yang mengganggu, seperti migrasi skala besar, pada operasi dan memungkinkan bisnis untuk melanjutkan operasi dengan cepat.

C

KAFE

Lihat Kerangka Adopsi AWS Cloud.

penyebaran kenari

Rilis versi yang lambat dan bertahap untuk pengguna akhir. Ketika Anda yakin, Anda menyebarkan versi baru dan mengganti versi saat ini secara keseluruhan.

CCoE

Lihat Cloud Center of Excellence.

CDC

Lihat mengubah pengambilan data.

ubah pengambilan data (CDC)

Proses melacak perubahan ke sumber data, seperti tabel database, dan merekam metadata tentang perubahan tersebut. Anda dapat menggunakan CDC untuk berbagai tujuan, seperti mengaudit atau mereplikasi perubahan dalam sistem target untuk mempertahankan sinkronisasi.

rekayasa kekacauan

Sengaja memperkenalkan kegagalan atau peristiwa yang mengganggu untuk menguji ketahanan sistem. Anda dapat menggunakan AWS Fault Injection Service (AWS FIS) untuk melakukan eksperimen yang menekankan AWS beban kerja Anda dan mengevaluasi responsnya.

CI/CD

Lihat integrasi berkelanjutan dan pengiriman berkelanjutan.

klasifikasi

Proses kategorisasi yang membantu menghasilkan prediksi. Model ML untuk masalah klasifikasi memprediksi nilai diskrit. Nilai diskrit selalu berbeda satu sama lain. Misalnya, model mungkin perlu mengevaluasi apakah ada mobil dalam gambar atau tidak.

Pengembang Warga

Pengguna bisnis yang membuat aplikasi AI menggunakan platform tanpa code/low kode tanpa keterampilan teknis khusus.

Enkripsi sisi klien

Enkripsi data secara lokal, sebelum target Layanan AWS menerimanya.

Cloud Center of Excellence (CCoE)

Tim multi-disiplin yang mendorong upaya adopsi cloud di seluruh organisasi, termasuk mengembangkan praktik terbaik cloud, memobilisasi sumber daya, menetapkan jadwal migrasi, dan memimpin organisasi melalui transformasi skala besar. Untuk informasi selengkapnya, lihat posting CCoE di Blog Strategi AWS Cloud Perusahaan.

komputasi cloud

Teknologi cloud yang biasanya digunakan untuk penyimpanan data jarak jauh dan manajemen perangkat IoT. Cloud computing umumnya terhubung ke teknologi edge computing.

model operasi cloud

Dalam organisasi TI, model operasi yang digunakan untuk membangun, mematangkan, dan mengoptimalkan satu atau lebih lingkungan cloud. Untuk informasi selengkapnya, lihat Membangun Model Operasi Cloud Anda.

tahap adopsi cloud

Empat fase yang biasanya dilalui organisasi ketika mereka bermigrasi ke AWS Cloud:

- Proyek — Menjalankan beberapa proyek terkait cloud untuk bukti konsep dan tujuan pembelajaran
- Foundation — Melakukan investasi dasar untuk meningkatkan adopsi cloud Anda (misalnya, membuat landing zone, mendefinisikan CCoE, membuat model operasi)
- Migrasi — Migrasi aplikasi individual
- Re-invention — Mengoptimalkan produk dan layanan, dan berinovasi di cloud

Tahapan ini didefinisikan oleh Stephen Orban dalam posting blog The Journey Toward Cloud-First & the Stages of Adoption di blog Strategi AWS Cloud Perusahaan. Untuk informasi tentang bagaimana kaitannya dengan strategi AWS migrasi, lihat panduan kesiapan migrasi.

CMDB

Lihat database manajemen konfigurasi.

repositori kode

Lokasi di mana kode sumber dan aset lainnya, seperti dokumentasi, sampel, dan skrip, disimpan dan diperbarui melalui proses kontrol versi. Repositori cloud umum termasuk GitHub atau. Bitbucket Cloud Setiap versi kode disebut cabang. Dalam struktur layanan mikro, setiap repositori dikhususkan untuk satu bagian fungsionalitas. Satu CI/CD pipa dapat menggunakan beberapa repositori.

cache dingin

Cache buffer yang kosong, tidak terisi dengan baik, atau berisi data basi atau tidak relevan. Ini mempengaruhi kinerja karena instance database harus membaca dari memori utama atau disk, yang lebih lambat daripada membaca dari cache buffer.

data dingin

Data yang jarang diakses dan biasanya historis. Saat menanyakan jenis data ini, kueri lambat biasanya dapat diterima. Memindahkan data ini ke tingkat atau kelas penyimpanan yang berkinerja lebih rendah dan lebih murah dapat mengurangi biaya.

visi komputer (CV)

Bidang AI yang menggunakan pembelajaran mesin untuk menganalisis dan mengekstrak informasi dari format visual seperti gambar dan video digital. Misalnya, Amazon SageMaker AI menyediakan algoritma pemrosesan gambar untuk CV.

konfigurasi drift

Untuk beban kerja, konfigurasi berubah dari status yang diharapkan. Ini dapat menyebabkan beban kerja menjadi tidak patuh, dan biasanya bertahap dan tidak disengaja.

database manajemen konfigurasi (CMDB)

Repositori yang menyimpan dan mengelola informasi tentang database dan lingkungan TI, termasuk komponen perangkat keras dan perangkat lunak dan konfigurasinya. Anda biasanya menggunakan data dari CMDB dalam penemuan portofolio dan tahap analisis migrasi.

paket kesesuaian

Kumpulan AWS Config aturan dan tindakan remediasi yang dapat Anda kumpulkan untuk menyesuaikan kepatuhan dan pemeriksaan keamanan Anda. Anda dapat menerapkan paket kesesuaian sebagai entitas tunggal di Akun AWS dan Region, atau di seluruh organisasi, dengan menggunakan templat YAMM. Untuk informasi selengkapnya, lihat Paket kesesuaian dalam dokumentasi. AWS Config

integrasi berkelanjutan dan pengiriman berkelanjutan (CI/CD)

Proses mengotomatiskan sumber, membangun, menguji, pementasan, dan tahap produksi dari proses rilis perangkat lunak. CI/CD biasanya digambarkan sebagai pipa. CI/CD dapat membantu Anda mengotomatiskan proses, meningkatkan produktivitas, meningkatkan kualitas kode, dan memberikan lebih cepat. Untuk informasi lebih lanjut, lihat Manfaat pengiriman berkelanjutan. CD juga dapat berarti penerapan berkelanjutan. Untuk informasi selengkapnya, lihat Continuous Delivery vs Continuous Deployment.

CV

Lihat visi komputer.

D

data saat istirahat

Data yang stasioner di jaringan Anda, seperti data yang ada di penyimpanan.

klasifikasi data

Proses untuk mengidentifikasi dan mengkategorikan data dalam jaringan Anda berdasarkan kekritisan dan sensitivitasnya. Ini adalah komponen penting dari setiap strategi manajemen risiko keamanan siber karena membantu Anda menentukan perlindungan dan kontrol retensi yang tepat untuk data. Klasifikasi data adalah komponen pilar keamanan dalam AWS Well-Architected Framework. Untuk informasi selengkapnya, lihat Klasifikasi data.

penyimpangan data

Variasi yang berarti antara data produksi dan data yang digunakan untuk melatih model ML, atau perubahan yang berarti dalam data input dari waktu ke waktu. Penyimpangan data dapat mengurangi kualitas, akurasi, dan keadilan keseluruhan dalam prediksi model ML.

data dalam transit

Data yang aktif bergerak melalui jaringan Anda, seperti antara sumber daya jaringan.

jala data

Kerangka arsitektur yang menyediakan kepemilikan data terdistribusi dan terdesentralisasi dengan manajemen dan tata kelola terpusat.

minimalisasi data

Prinsip pengumpulan dan pemrosesan hanya data yang sangat diperlukan. Mempraktikkan minimalisasi data di dalamnya AWS Cloud dapat mengurangi risiko privasi, biaya, dan jejak karbon analitik Anda.

perimeter data

Satu set pagar pembatas pencegahan di AWS lingkungan Anda yang membantu memastikan bahwa hanya identitas tepercaya yang mengakses sumber daya tepercaya dari jaringan yang diharapkan. Untuk informasi selengkapnya, lihat Membangun perimeter data pada AWS.

prapemrosesan data

Untuk mengubah data mentah menjadi format yang mudah diuraikan oleh model ML Anda. Preprocessing data dapat berarti menghapus kolom atau baris tertentu dan menangani nilai yang hilang, tidak konsisten, atau duplikat.

asal data

Proses melacak asal dan riwayat data sepanjang siklus hidupnya, seperti bagaimana data dihasilkan, ditransmisikan, dan disimpan.

subjek data

Individu yang datanya dikumpulkan dan diproses.

gudang data

Sistem manajemen data yang mendukung intelijen bisnis, seperti analitik. Gudang data biasanya berisi sejumlah besar data historis, dan biasanya digunakan untuk kueri dan analisis.

bahasa definisi database (DDL)

Pernyataan atau perintah untuk membuat atau memodifikasi struktur tabel dan objek dalam database.

bahasa manipulasi basis data (DHTML)

Pernyataan atau perintah untuk memodifikasi (memasukkan, memperbarui, dan menghapus) informasi dalam database.

DDL

Lihat bahasa definisi database.

ansambel yang dalam

Untuk menggabungkan beberapa model pembelajaran mendalam untuk prediksi. Anda dapat menggunakan ansambel dalam untuk mendapatkan prediksi yang lebih akurat atau untuk memperkirakan ketidakpastian dalam prediksi.

pembelajaran mendalam

Subbidang ML yang menggunakan beberapa lapisan jaringan saraf tiruan untuk mengidentifikasi pemetaan antara data input dan variabel target yang diinginkan.

pertahanan-mendalam

Pendekatan keamanan informasi di mana serangkaian mekanisme dan kontrol keamanan dilapisi dengan cermat di seluruh jaringan komputer untuk melindungi kerahasiaan, integritas, dan ketersediaan jaringan dan data di dalamnya. Saat Anda mengadopsi strategi ini AWS, Anda menambahkan beberapa kontrol pada lapisan AWS Organizations struktur yang berbeda untuk membantu mengamankan sumber daya. Misalnya, pendekatan defense-in-depth mungkin menggabungkan otentikasi multi-faktor, segmentasi jaringan, dan enkripsi.

administrator yang didelegasikan

Di AWS Organizations, layanan yang kompatibel dapat mendaftarkan akun AWS anggota untuk mengelola akun organisasi dan mengelola izin untuk layanan tersebut. Akun ini disebut administrator yang didelegasikan untuk layanan itu. Untuk informasi selengkapnya dan daftar layanan yang kompatibel, lihat Layanan yang berfungsi dengan AWS Organizations AWS Organizations dokumentasi.

deployment

Proses pembuatan aplikasi, fitur baru, atau perbaikan kode tersedia di lingkungan target. Deployment melibatkan penerapan perubahan dalam basis kode dan kemudian membangun dan menjalankan basis kode itu di lingkungan aplikasi.

lingkungan pengembangan

Lihat lingkungan.

kontrol detektif

Kontrol keamanan yang dirancang untuk mendeteksi, mencatat, dan memperingatkan setelah suatu peristiwa terjadi. Kontrol ini adalah garis pertahanan kedua, memperingatkan Anda tentang peristiwa keamanan yang melewati kontrol pencegahan yang ada. Untuk informasi selengkapnya, lihat Kontrol Detektif dalam Menerapkan kontrol keamanan pada. AWS

pemetaan aliran nilai pengembangan (DVSM)

Sebuah proses yang digunakan untuk mengidentifikasi dan memprioritaskan kendala yang mempengaruhi kecepatan dan kualitas dalam siklus hidup pengembangan perangkat lunak. DVSM memperluas proses pemetaan aliran nilai yang awalnya dirancang untuk praktik manufaktur ramping. Ini berfokus pada langkah-langkah dan tim yang diperlukan untuk menciptakan dan memindahkan nilai melalui proses pengembangan perangkat lunak.

kembar digital

Representasi virtual dari sistem dunia nyata, seperti bangunan, pabrik, peralatan industri, atau jalur produksi. Kembar digital mendukung pemeliharaan prediktif, pemantauan jarak jauh, dan optimalisasi produksi.

tabel dimensi

Dalam skema bintang, tabel yang lebih kecil yang berisi atribut data tentang data kuantitatif dalam tabel fakta. Atribut tabel dimensi biasanya bidang teks atau angka diskrit yang berperilaku seperti teks. Atribut ini biasanya digunakan untuk pembatasan kueri, pemfilteran, dan pelabelan set hasil.

musibah

Peristiwa yang mencegah beban kerja atau sistem memenuhi tujuan bisnisnya di lokasi utama yang digunakan. Peristiwa ini dapat berupa bencana alam, kegagalan teknis, atau akibat dari tindakan manusia, seperti kesalahan konfigurasi yang tidak disengaja atau serangan malware.

pemulihan bencana (DR)

Strategi dan proses yang Anda gunakan untuk meminimalkan downtime dan kehilangan data yang disebabkan oleh bencana. Untuk informasi selengkapnya, lihat Disaster Recovery of Workloads on AWS: Recovery in the Cloud in the AWS Well-Architected Framework.

DML~

Lihat bahasa manipulasi database.

desain berbasis domain

Pendekatan untuk mengembangkan sistem perangkat lunak yang kompleks dengan menghubungkan komponennya ke domain yang berkembang, atau tujuan bisnis inti, yang dilayani setiap komponen. Konsep ini diperkenalkan oleh Eric Evans dalam bukunya, Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston: Addison-Wesley Professional, 2003). Untuk informasi tentang cara menggunakan desain berbasis domain dengan pola gambar pencekik, lihat Memodernisasi layanan web Microsoft ASP.NET (ASMX) lama secara bertahap menggunakan container dan Amazon API Gateway.

DR

Lihat pemulihan bencana.

deteksi drift

Melacak penyimpangan dari konfigurasi dasar. Misalnya, Anda dapat menggunakan AWS CloudFormation untuk mendeteksi penyimpangan dalam sumber daya sistem, atau Anda dapat menggunakannya AWS Control Tower untuk mendeteksi perubahan di landing zone yang mungkin memengaruhi kepatuhan terhadap persyaratan tata kelola.

DVSM

Lihat pemetaan aliran nilai pengembangan.

E

EDA

Lihat analisis data eksplorasi.

EDI

Lihat pertukaran data elektronik.

komputasi tepi

Teknologi yang meningkatkan daya komputasi untuk perangkat pintar di tepi jaringan IoT. Jika dibandingkan dengan komputasi awan, komputasi tepi dapat mengurangi latensi komunikasi dan meningkatkan waktu respons.

pertukaran data elektronik (EDI)

Pertukaran otomatis dokumen bisnis antar organisasi. Untuk informasi selengkapnya, lihat Apa itu Pertukaran Data Elektronik.

enkripsi

Proses komputasi yang mengubah data plaintext, yang dapat dibaca manusia, menjadi ciphertext.

kunci enkripsi

String kriptografi dari bit acak yang dihasilkan oleh algoritma enkripsi. Panjang kunci dapat bervariasi, dan setiap kunci dirancang agar tidak dapat diprediksi dan unik.

endianness

Urutan byte disimpan dalam memori komputer. Big-endian sistem menyimpan byte paling signifikan terlebih dahulu. Little-endian sistem menyimpan byte paling tidak signifikan terlebih dahulu.

titik akhir

Lihat titik akhir layanan.

layanan endpoint

Layanan yang dapat Anda host di cloud pribadi virtual (VPC) untuk dibagikan dengan pengguna lain. Anda dapat membuat layanan endpoint dengan AWS PrivateLink dan memberikan izin kepada prinsipal lain Akun AWS atau ke AWS Identity and Access Management (IAM). Akun atau prinsipal ini dapat terhubung ke layanan endpoint Anda secara pribadi dengan membuat titik akhir VPC antarmuka. Untuk informasi selengkapnya, lihat Membuat layanan titik akhir di dokumentasi Amazon Virtual Private Cloud (Amazon VPC).

perencanaan sumber daya perusahaan (ERP)

Sistem yang mengotomatiskan dan mengelola proses bisnis utama (seperti akuntansi, MES, dan manajemen proyek) untuk suatu perusahaan.

enkripsi amplop

Proses mengenkripsi kunci enkripsi dengan kunci enkripsi lain. Untuk informasi selengkapnya, lihat Enkripsi amplop dalam dokumentasi AWS Key Management Service (AWS KMS).

lingkungan

Sebuah contoh dari aplikasi yang sedang berjalan. Berikut ini adalah jenis lingkungan yang umum dalam komputasi awan:

- Development Environment — Sebuah contoh dari aplikasi yang berjalan yang hanya tersedia untuk tim inti yang bertanggung jawab untuk memelihara aplikasi. Lingkungan pengembangan digunakan untuk menguji perubahan sebelum mempromosikannya ke lingkungan atas. Jenis lingkungan ini kadang-kadang disebut sebagai lingkungan pengujian.

- lingkungan yang lebih rendah — Semua lingkungan pengembangan untuk aplikasi, seperti yang digunakan untuk build awal dan pengujian.
- lingkungan produksi — Sebuah contoh dari aplikasi yang berjalan yang dapat diakses oleh pengguna akhir. Dalam sebuah CI/CD pipeline, lingkungan produksi adalah lingkungan penyebaran terakhir.
- lingkungan atas — Semua lingkungan yang dapat diakses oleh pengguna selain tim pengembangan inti. Ini dapat mencakup lingkungan produksi, lingkungan praproduksi, dan lingkungan untuk pengujian penerimaan pengguna.

epik

Dalam metodologi tangkas, kategori fungsional yang membantu mengatur dan memprioritaskan pekerjaan Anda. Epik memberikan deskripsi tingkat tinggi tentang persyaratan dan tugas implementasi. Misalnya, epos keamanan AWS CAF mencakup manajemen identitas dan akses, kontrol detektif, keamanan infrastruktur, perlindungan data, dan respons insiden. Untuk informasi selengkapnya tentang epos dalam strategi AWS migrasi, lihat panduan implementasi program.

ERP

Lihat perencanaan sumber daya perusahaan.

analisis data eksplorasi (EDA)

Proses menganalisis dataset untuk memahami karakteristik utamanya. Anda mengumpulkan atau mengumpulkan data dan kemudian melakukan penyelidikan awal untuk menemukan pola, mendeteksi anomali, dan memeriksa asumsi. EDA dilakukan dengan menghitung statistik ringkasan dan membuat visualisasi data.

F

tabel fakta

Tabel tengah dalam skema bintang. Ini menyimpan data kuantitatif tentang operasi bisnis. Biasanya, tabel fakta berisi dua jenis kolom: kolom yang berisi ukuran dan yang berisi kunci asing ke tabel dimensi.

gagal cepat

Filosofi yang menggunakan pengujian yang sering dan bertahap untuk mengurangi siklus hidup pengembangan. Ini adalah bagian penting dari pendekatan tangkas.

batas isolasi kesalahan

Dalam AWS Cloud, batas seperti Availability Zone, Wilayah AWS, control plane, atau data plane yang membatasi efek kegagalan dan membantu meningkatkan ketahanan beban kerja. Untuk informasi selengkapnya, lihat Batas Isolasi AWS Kesalahan.

cabang fitur

Lihat cabang.

fitur

Data input yang Anda gunakan untuk membuat prediksi. Misalnya, dalam konteks manufaktur, fitur bisa berupa gambar yang diambil secara berkala dari lini manufaktur.

pentingnya fitur

Seberapa signifikan fitur untuk prediksi model. Ini biasanya dinyatakan sebagai skor numerik yang dapat dihitung melalui berbagai teknik, seperti Shapley Additive Explanations (SHAP) dan gradien terintegrasi. Untuk informasi lebih lanjut, lihat Interpretabilitas model pembelajaran mesin dengan. AWS

transformasi fitur

Untuk mengoptimalkan data untuk proses ML, termasuk memperkaya data dengan sumber tambahan, menskalakan nilai, atau mengekstrak beberapa set informasi dari satu bidang data. Hal ini memungkinkan model ML untuk mendapatkan keuntungan dari data. Misalnya, jika Anda memecah tanggal “2021-05-27 00:15:37" menjadi “2021", “Mei”, “Kamis”, dan “15", Anda dapat membantu algoritme pembelajaran mempelajari pola bernuansa yang terkait dengan komponen data yang berbeda.

beberapa tembakan mendorong

Menyediakan LLM dengan sejumlah kecil contoh yang menunjukkan tugas dan output yang diinginkan sebelum memintanya untuk melakukan tugas serupa. Teknik ini adalah aplikasi pembelajaran dalam konteks, di mana model belajar dari contoh (bidikan) yang tertanam dalam petunjuk. Few-shot prompt bisa efektif untuk tugas-tugas yang membutuhkan pemformatan, penalaran, atau pengetahuan domain tertentu. Lihat juga bidikan nol.

FGAC

Lihat kontrol akses berbutir halus.

kontrol akses berbutir halus (FGAC)

Penggunaan beberapa kondisi untuk mengizinkan atau menolak permintaan akses.

migrasi flash-cut

Metode migrasi database yang menggunakan replikasi data berkelanjutan melalui pengambilan data perubahan untuk memigrasikan data dalam waktu sesingkat mungkin, alih-alih menggunakan pendekatan bertahap. Tujuannya adalah untuk menjaga downtime seminimal mungkin.

FM

Lihat model pondasi.

model pondasi (FM)

Jaringan saraf pembelajaran mendalam yang besar yang telah melatih kumpulan data besar-besaran data umum dan tidak berlabel. FM mampu melakukan berbagai tugas umum, seperti memahami bahasa, menghasilkan teks dan gambar, dan berbicara dalam bahasa alami. Untuk informasi selengkapnya, lihat Apa itu Model Foundation.

Gerbang FM

Perantara terpusat yang mengontrol dan menormalkan akses ke model pondasi. Juga dikenal sebagai gateway LLM.

G

AI generatif

Subset model AI yang telah dilatih pada sejumlah besar data dan yang dapat menggunakan prompt teks sederhana untuk membuat konten dan artefak baru, seperti gambar, video, teks, dan audio. Untuk informasi lebih lanjut, lihat Apa itu AI Generatif.

pemblokiran geografis

Lihat pembatasan geografis.

pembatasan geografis (pemblokiran geografis)

Di Amazon CloudFront, opsi untuk mencegah pengguna di negara tertentu mengakses distribusi konten. Anda dapat menggunakan daftar izinkan atau daftar blokir untuk menentukan negara yang disetujui dan dilarang. Untuk informasi selengkapnya, lihat Membatasi distribusi geografis konten Anda dalam dokumentasi. CloudFront

Alur kerja Gitflow

Pendekatan di mana lingkungan bawah dan atas menggunakan cabang yang berbeda dalam repositori kode sumber. Alur kerja Gitflow dianggap warisan, dan alur kerja berbasis batang adalah pendekatan modern yang lebih disukai.

gambar emas

Sebuah snapshot dari sistem atau perangkat lunak yang digunakan sebagai template untuk menyebarkan instance baru dari sistem atau perangkat lunak itu. Misalnya, di bidang manufaktur, gambar emas dapat digunakan untuk menyediakan perangkat lunak pada beberapa perangkat dan membantu meningkatkan kecepatan, skalabilitas, dan produktivitas dalam operasi manufaktur perangkat.

strategi greenfield

Tidak adanya infrastruktur yang ada di lingkungan baru. Saat mengadopsi strategi greenfield untuk arsitektur sistem, Anda dapat memilih semua teknologi baru tanpa batasan kompatibilitas dengan infrastruktur yang ada, juga dikenal sebagai brownfield. Jika Anda memperluas infrastruktur yang ada, Anda dapat memadukan strategi brownfield dan greenfield.

pagar pembatas

Aturan tingkat tinggi yang membantu mengatur sumber daya, kebijakan, dan kepatuhan di seluruh unit organisasi (OU). Pagar pembatas preventif menegakkan kebijakan untuk memastikan keselarasan dengan standar kepatuhan. Mereka diimplementasikan dengan menggunakan kebijakan kontrol layanan dan batas izin IAM. Detective guardrails mendeteksi pelanggaran kebijakan dan masalah kepatuhan, dan menghasilkan peringatan untuk remediasi. Mereka diimplementasikan dengan menggunakan AWS Config, AWS Security Hub CSPM, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector, dan pemeriksaan khusus AWS Lambda .

pagar pembatas (AI)

Mekanisme keamanan yang menyaring, memvalidasi, dan membatasi input dan output agen untuk membantu memastikan perilaku AI yang bertanggung jawab dan aman.

H

HA

Lihat ketersediaan tinggi.

migrasi database heterogen

Memigrasi database sumber Anda ke database target yang menggunakan mesin database yang berbeda (misalnya, Oracle ke Amazon Aurora). Migrasi heterogen biasanya merupakan bagian dari upaya arsitektur ulang, dan mengubah skema dapat menjadi tugas yang kompleks. AWS menyediakan AWS SCT yang membantu dengan konversi skema.

ketersediaan tinggi (HA)

Kemampuan beban kerja untuk beroperasi terus menerus, tanpa intervensi, jika terjadi tantangan atau bencana. Sistem HA dirancang untuk gagal secara otomatis, secara konsisten memberikan kinerja berkualitas tinggi, dan menangani beban dan kegagalan yang berbeda dengan dampak kinerja minimal.

modernisasi sejarawan

Pendekatan yang digunakan untuk memodernisasi dan meningkatkan sistem teknologi operasional (OT) untuk melayani kebutuhan industri manufaktur dengan lebih baik. Sejarawan adalah jenis database yang digunakan untuk mengumpulkan dan menyimpan data dari berbagai sumber di pabrik.

data penahanan

Sebagian dari data historis berlabel yang ditahan dari kumpulan data yang digunakan untuk melatih model pembelajaran mesin. Anda dapat menggunakan data penahanan untuk mengevaluasi kinerja model dengan membandingkan prediksi model dengan data penahanan.

manusia-dalam-lingkaran (HiTL)

Pola alur kerja di mana eksekusi agen berhenti untuk peninjauan dan persetujuan manusia pada titik keputusan kritis.

migrasi database homogen

Memigrasi database sumber Anda ke database target yang berbagi mesin database yang sama (misalnya, Microsoft SQL Server ke Amazon RDS for SQL Server). Migrasi homogen biasanya merupakan bagian dari upaya rehosting atau replatforming. Anda dapat menggunakan utilitas database asli untuk memigrasi skema.

data panas

Data yang sering diakses, seperti data real-time atau data translasi terbaru. Data ini biasanya memerlukan tingkat atau kelas penyimpanan berkinerja tinggi untuk memberikan respons kueri yang cepat.

perbaikan terbaru

Perbaikan mendesak untuk masalah kritis dalam lingkungan produksi. Karena urgensinya, perbaikan terbaru biasanya dibuat di luar alur kerja DevOps rilis biasa.

periode hypercare

Segera setelah cutover, periode waktu ketika tim migrasi mengelola dan memantau aplikasi yang dimigrasi di cloud untuk mengatasi masalah apa pun. Biasanya, periode ini panjangnya 1-4 hari. Pada akhir periode hypercare, tim migrasi biasanya mentransfer tanggung jawab untuk aplikasi ke tim operasi cloud.

I

IAc

Lihat infrastruktur sebagai kode.

kebijakan berbasis identitas

Kebijakan yang dilampirkan pada satu atau beberapa prinsip IAM yang mendefinisikan izin mereka dalam lingkungan. AWS Cloud

aplikasi idle

Aplikasi yang memiliki penggunaan CPU dan memori rata-rata antara 5 dan 20 persen selama periode 90 hari. Dalam proyek migrasi, adalah umum untuk menghentikan aplikasi ini atau mempertahankannya di tempat.

IIoT

Lihat Internet of Things industri.

infrastruktur yang tidak dapat diubah

Model yang menyebarkan infrastruktur baru untuk beban kerja produksi alih-alih memperbarui, menambal, atau memodifikasi infrastruktur yang ada. Infrastruktur yang tidak dapat diubah secara inheren lebih konsisten, andal, dan dapat diprediksi daripada infrastruktur yang dapat berubah. Untuk informasi selengkapnya, lihat praktik terbaik Deploy using immutable infrastructure in the Framework. AWS Well-Architected

masuk (masuknya) VPC

Dalam arsitektur AWS multi-akun, VPC yang menerima, memeriksa, dan merutekan koneksi jaringan dari luar aplikasi. Arsitektur Referensi AWS Keamanan merekomendasikan pengaturan

akun Jaringan Anda dengan VPC masuk, keluar, dan inspeksi untuk melindungi antarmuka dua arah antara aplikasi Anda dan internet yang lebih luas.

migrasi inkremental

Strategi cutover di mana Anda memigrasikan aplikasi Anda dalam bagian-bagian kecil alih-alih melakukan satu cutover penuh. Misalnya, Anda mungkin hanya memindahkan beberapa layanan mikro atau pengguna ke sistem baru pada awalnya. Setelah Anda memverifikasi bahwa semuanya berfungsi dengan baik, Anda dapat secara bertahap memindahkan layanan mikro atau pengguna tambahan hingga Anda dapat menonaktifkan sistem lama Anda. Strategi ini mengurangi risiko yang terkait dengan migrasi besar.

Industri 4.0

Sebuah istilah yang diperkenalkan oleh Klaus Schwab pada tahun 2016 untuk merujuk pada modernisasi proses manufaktur melalui kemajuan dalam konektivitas, data real-time, otomatisasi, analitik, dan. AI/ML

infrastruktur

Semua sumber daya dan aset yang terkandung dalam lingkungan aplikasi.

infrastruktur sebagai kode (IAc)

Proses penyediaan dan pengelolaan infrastruktur aplikasi melalui satu set file konfigurasi. IAc dirancang untuk membantu Anda memusatkan manajemen infrastruktur, menstandarisasi sumber daya, dan menskalakan dengan cepat sehingga lingkungan baru dapat diulang, andal, dan konsisten.

Internet of Things industri (IIoT)

Penggunaan sensor dan perangkat yang terhubung ke internet di sektor industri, seperti manufaktur, energi, otomotif, perawatan kesehatan, ilmu kehidupan, dan pertanian. Untuk informasi selengkapnya, lihat Membangun strategi transformasi digital Internet of Things (IIoT) industri.

inspeksi VPC

Dalam arsitektur AWS multi-akun, VPC terpusat yang mengelola inspeksi lalu lintas jaringan antara VPC (dalam hal yang sama atau berbeda Wilayah AWS), internet, dan jaringan lokal. Arsitektur Referensi AWS Keamanan merekomendasikan pengaturan akun Jaringan Anda dengan VPC masuk, keluar, dan inspeksi untuk melindungi antarmuka dua arah antara aplikasi Anda dan internet yang lebih luas.

Internet of Things (IoT)

Jaringan objek fisik yang terhubung dengan sensor atau prosesor tertanam yang berkomunikasi dengan perangkat dan sistem lain melalui internet atau melalui jaringan komunikasi lokal. Untuk informasi selengkapnya, lihat Apa itu IoT?

interpretabilitas

Karakteristik model pembelajaran mesin yang menggambarkan sejauh mana manusia dapat memahami bagaimana prediksi model bergantung pada inputnya. Untuk informasi lebih lanjut, lihat Interpretabilitas model pembelajaran mesin dengan. AWS

IoT

Lihat Internet of Things.

Perpustakaan informasi TI (ITIL)

Serangkaian praktik terbaik untuk memberikan layanan TI dan menyelaraskan layanan ini dengan persyaratan bisnis. ITIL menyediakan dasar untuk ITSM.

Manajemen layanan TI (ITSM)

Kegiatan yang terkait dengan merancang, menerapkan, mengelola, dan mendukung layanan TI untuk suatu organisasi. Untuk informasi tentang mengintegrasikan operasi cloud dengan alat ITSM, lihat panduan integrasi operasi.

ITIL

Lihat perpustakaan informasi TI.

ITSM

Lihat manajemen layanan TI.

L

kontrol akses berbasis label (LBAC)

Implementasi kontrol akses wajib (MAC) di mana pengguna dan data itu sendiri masing-masing secara eksplisit diberi nilai label keamanan. Persimpangan antara label keamanan pengguna dan label keamanan data menentukan baris dan kolom mana yang dapat dilihat oleh pengguna.

landing zone

Landing zone adalah AWS lingkungan multi-akun yang dirancang dengan baik yang dapat diskalakan dan aman. Ini adalah titik awal dari mana organisasi Anda dapat dengan cepat meluncurkan dan menyebarkan beban kerja dan aplikasi dengan percaya diri dalam lingkungan keamanan dan infrastruktur mereka. Untuk informasi selengkapnya tentang zona pendaratan, lihat Menyiapkan lingkungan multi-akun AWS yang aman dan dapat diskalakan.

model bahasa besar (LLM)

Model AI pembelajaran mendalam yang dilatih sebelumnya pada sejumlah besar data. LLM dapat melakukan beberapa tugas, seperti menjawab pertanyaan, meringkas dokumen, menerjemahkan teks ke bahasa lain, dan menyelesaikan kalimat. Untuk informasi lebih lanjut, lihat Apa itu LLM.

migrasi besar

Migrasi 300 atau lebih server.

LBAC

Lihat kontrol akses berbasis label.

hak istimewa paling sedikit

Praktik keamanan terbaik untuk memberikan izin minimum yang diperlukan untuk melakukan tugas. Untuk informasi selengkapnya, lihat Menerapkan izin hak istimewa terkecil dalam dokumentasi IAM.

angkat dan geser

Lihat 7 Rs.

sistem endian kecil

Sebuah sistem yang menyimpan byte paling tidak signifikan terlebih dahulu. Lihat juga endianness.

LLM

Lihat model bahasa besar.

lingkungan yang lebih rendah

Lihat lingkungan.

M

pembelajaran mesin (ML)

Jenis kecerdasan buatan yang menggunakan algoritma dan teknik untuk pengenalan pola dan pembelajaran. ML menganalisis dan belajar dari data yang direkam, seperti data Internet of Things (IoT), untuk menghasilkan model statistik berdasarkan pola. Untuk informasi selengkapnya, lihat Machine Learning.

cabang utama

Lihat cabang.

malware

Perangkat lunak yang dirancang untuk membahayakan keamanan atau privasi komputer. Malware dapat mengganggu sistem komputer, membocorkan informasi sensitif, atau mendapatkan akses yang tidak sah. Contoh malware termasuk virus, worm, ransomware, Trojan horse, spyware, dan keyloggers.

layanan terkelola

Layanan AWS yang AWS mengoperasikan lapisan infrastruktur, sistem operasi, dan platform, dan Anda mengakses titik akhir untuk menyimpan dan mengambil data. Amazon Simple Storage Service (Amazon S3) dan Amazon DynamoDB adalah contoh layanan terkelola. Ini juga dikenal sebagai layanan abstrak.

sistem eksekusi manufaktur (MES)

Sistem perangkat lunak untuk melacak, memantau, mendokumentasikan, dan mengendalikan proses produksi yang mengubah bahan baku menjadi produk jadi di lantai toko.

PETA

Lihat Program Percepatan Migrasi.

MCP

Lihat Protokol Konteks Model.

Protokol Konteks Model (MCP)

Protokol stateless untuk komunikasi agen -to- alat.

Server MCP

Layanan yang mengekspos satu atau lebih alat melalui Protokol Konteks Model.

mekanisme

Proses lengkap di mana Anda membuat alat, mendorong adopsi alat, dan kemudian memeriksa hasilnya untuk melakukan penyesuaian. Mekanisme adalah siklus yang memperkuat dan meningkatkan dirinya sendiri saat beroperasi. Untuk informasi selengkapnya, lihat Membangun mekanisme dalam AWS Well-Architected Kerangka Kerja.

akun anggota

Semua Akun AWS selain akun manajemen yang merupakan bagian dari organisasi di AWS Organizations. Akun dapat menjadi anggota dari hanya satu organisasi pada suatu waktu.

MES

Lihat sistem eksekusi manufaktur.

Transportasi Telemetri Antrian Pesan (MQTT)

Protokol komunikasi mesin-ke-mesin (M2M) yang ringan, berdasarkan pola publish/subscribe, untuk perangkat IoT yang dibatasi sumber daya.

layanan mikro

Layanan kecil dan independen yang berkomunikasi melalui API yang terdefinisi dengan baik dan biasanya dimiliki oleh tim kecil yang mandiri. Misalnya, sistem asuransi mungkin mencakup layanan mikro yang memetakan kemampuan bisnis, seperti penjualan atau pemasaran, atau subdomain, seperti pembelian, klaim, atau analitik. Manfaat layanan mikro termasuk kelincahan, penskalaan yang fleksibel, penyebaran yang mudah, kode yang dapat digunakan kembali, dan ketahanan. Untuk informasi selengkapnya, lihat Mengintegrasikan layanan mikro dengan menggunakan layanan tanpa AWS server.

arsitektur microservices

Pendekatan untuk membangun aplikasi dengan komponen independen yang menjalankan setiap proses aplikasi sebagai layanan mikro. Layanan mikro ini berkomunikasi melalui antarmuka yang terdefinisi dengan baik dengan menggunakan API ringan. Setiap layanan mikro dalam arsitektur ini dapat diperbarui, digunakan, dan diskalakan untuk memenuhi permintaan fungsi tertentu dari suatu aplikasi. Untuk informasi selengkapnya, lihat Menerapkan layanan mikro di AWS.

Program Percepatan Migrasi (MAP)

AWS Program yang menyediakan dukungan konsultasi, pelatihan, dan layanan untuk membantu organisasi membangun fondasi operasional yang kuat untuk pindah ke cloud, dan untuk membantu mengimbangi biaya awal migrasi. MAP mencakup metodologi migrasi untuk

mengeksekusi migrasi lama dengan cara metodis dan seperangkat alat untuk mengotomatisasi dan mempercepat skenario migrasi umum.

migrasi dalam skala

Proses memindahkan sebagian besar portofolio aplikasi ke cloud dalam gelombang, dengan lebih banyak aplikasi bergerak pada tingkat yang lebih cepat di setiap gelombang. Fase ini menggunakan praktik terbaik dan pelajaran yang dipetik dari fase sebelumnya untuk mengimplementasikan pabrik migrasi tim, alat, dan proses untuk merampingkan migrasi beban kerja melalui otomatisasi dan pengiriman tangkas. Ini adalah fase ketiga dari strategi AWS migrasi.

pabrik migrasi

Cross-functional tim yang merampingkan migrasi beban kerja melalui pendekatan otomatis dan gesit. Tim pabrik migrasi biasanya mencakup operasi, analis dan pemilik bisnis, insinyur migrasi, pengembang, dan DevOps profesional yang bekerja di sprint. Antara 20 dan 50 persen portofolio aplikasi perusahaan terdiri dari pola berulang yang dapat dioptimalkan dengan pendekatan pabrik. Untuk informasi selengkapnya, lihat diskusi tentang pabrik migrasi dan panduan Pabrik Migrasi Cloud di kumpulan konten ini.

metadata migrasi

Informasi tentang aplikasi dan server yang diperlukan untuk menyelesaikan migrasi. Setiap pola migrasi memerlukan satu set metadata migrasi yang berbeda. Contoh metadata migrasi termasuk subnet target, grup keamanan, dan akun. AWS

pola migrasi

Tugas migrasi berulang yang merinci strategi migrasi, tujuan migrasi, dan aplikasi atau layanan migrasi yang digunakan. Contoh: Rehost migrasi ke Amazon EC2 AWS dengan Layanan Migrasi Aplikasi.

Penilaian Portofolio Migrasi (MPA)

Alat online yang menyediakan informasi untuk memvalidasi kasus bisnis untuk bermigrasi ke. AWS Cloud MPA menyediakan penilaian portofolio terperinci (ukuran kanan server, harga, perbandingan TCO, analisis biaya migrasi) serta perencanaan migrasi (analisis data aplikasi dan pengumpulan data, pengelompokan aplikasi, prioritas migrasi, dan perencanaan gelombang). Alat MPA (memerlukan login) tersedia gratis untuk semua AWS konsultan dan konsultan APN Partner.

Penilaian Kesiapan Migrasi (MRA)

Proses mendapatkan wawasan tentang status kesiapan cloud organisasi, mengidentifikasi kekuatan dan kelemahan, dan membangun rencana aksi untuk menutup kesenjangan yang diidentifikasi, menggunakan CAF. AWS Untuk informasi selengkapnya, lihat panduan kesiapan migrasi. MRA adalah tahap pertama dari strategi AWS migrasi.

strategi migrasi

Pendekatan yang digunakan untuk memigrasikan beban kerja ke. AWS Cloud Untuk informasi lebih lanjut, lihat entri 7 Rs di glosarium ini dan lihat Memobilisasi organisasi Anda untuk mempercepat migrasi skala besar.

ML

Lihat pembelajaran mesin.

modernisasi

Mengubah aplikasi usang (warisan atau monolitik) dan infrastrukturnya menjadi sistem yang gesit, elastis, dan sangat tersedia di cloud untuk mengurangi biaya, mendapatkan efisiensi, dan memanfaatkan inovasi. Untuk informasi selengkapnya, lihat Strategi untuk memodernisasi aplikasi di. AWS Cloud

penilaian kesiapan modernisasi

Evaluasi yang membantu menentukan kesiapan modernisasi aplikasi organisasi; mengidentifikasi manfaat, risiko, dan dependensi; dan menentukan seberapa baik organisasi dapat mendukung keadaan masa depan aplikasi tersebut. Hasil penilaian adalah cetak biru arsitektur target, peta jalan yang merinci fase pengembangan dan tonggak untuk proses modernisasi, dan rencana aksi untuk mengatasi kesenjangan yang diidentifikasi. Untuk informasi lebih lanjut, lihat Mengevaluasi kesiapan modernisasi untuk aplikasi di. AWS Cloud

aplikasi monolitik (monolit)

Aplikasi yang berjalan sebagai layanan tunggal dengan proses yang digabungkan secara ketat. Aplikasi monolitik memiliki beberapa kelemahan. Jika satu fitur aplikasi mengalami lonjakan permintaan, seluruh arsitektur harus diskalakan. Menambahkan atau meningkatkan fitur aplikasi monolitik juga menjadi lebih kompleks ketika basis kode tumbuh. Untuk mengatasi masalah ini, Anda dapat menggunakan arsitektur microservices. Untuk informasi lebih lanjut, lihat Mengurai monolit menjadi layanan mikro.

MPA

Lihat Penilaian Portofolio Migrasi.

MQTT

Lihat Transportasi Telemetri Antrian Pesan.

klasifikasi multiclass

Sebuah proses yang membantu menghasilkan prediksi untuk beberapa kelas (memprediksi satu dari lebih dari dua hasil). Misalnya, model ML mungkin bertanya "Apakah produk ini buku, mobil, atau telepon?" atau "Kategori produk mana yang paling menarik bagi pelanggan ini?"

infrastruktur yang bisa berubah

Model yang memperbarui dan memodifikasi infrastruktur yang ada untuk beban kerja produksi. Untuk meningkatkan konsistensi, keandalan, dan prediktabilitas, AWS Well-Architected Framework merekomendasikan penggunaan infrastruktur yang tidak dapat diubah sebagai praktik terbaik.

O

OAC

Lihat kontrol akses asal.

OAI

Lihat identitas akses asal.

OCM

Lihat manajemen perubahan organisasi.

migrasi offline

Metode migrasi di mana beban kerja sumber diturunkan selama proses migrasi. Metode ini melibatkan waktu henti yang diperpanjang dan biasanya digunakan untuk beban kerja kecil dan tidak kritis.

OI

Lihat integrasi operasi.

OLA

Lihat perjanjian tingkat operasional.

migrasi online

Metode migrasi di mana beban kerja sumber disalin ke sistem target tanpa diambil offline. Aplikasi yang terhubung ke beban kerja dapat terus berfungsi selama migrasi. Metode ini melibatkan waktu henti nol hingga minimal dan biasanya digunakan untuk beban kerja produksi yang kritis.

OPC-UA

Lihat Komunikasi Proses Terbuka - Arsitektur Terpadu.

Komunikasi Proses Terbuka - Arsitektur Terpadu () OPC-UA

Protokol komunikasi mesin-ke-mesin (M2M) untuk otomasi industri. OPC-UA menyediakan standar interoperabilitas dengan enkripsi data, otentikasi, dan skema otorisasi.

perjanjian tingkat operasional (OLA)

Perjanjian yang menjelaskan apa yang dijanjikan kelompok TI fungsional untuk diberikan satu sama lain, untuk mendukung perjanjian tingkat layanan (SLA).

Tinjauan Kesiapan Operasional (ORR)

Daftar pertanyaan dan praktik terbaik terkait yang membantu Anda memahami, mengevaluasi, mencegah, atau mengurangi ruang lingkup insiden dan kemungkinan kegagalan. Untuk informasi selengkapnya, lihat Ulasan Kesiapan Operasional (ORR) dalam Kerangka Kerja. AWS Well-Architected

teknologi operasional (OT)

Sistem perangkat keras dan perangkat lunak yang bekerja dengan lingkungan fisik untuk mengendalikan operasi industri, peralatan, dan infrastruktur. Di bidang manufaktur, integrasi sistem OT dan teknologi informasi (TI) adalah fokus utama untuk transformasi Industri 4.0.

integrasi operasi (OI)

Proses modernisasi operasi di cloud, yang melibatkan perencanaan kesiapan, otomatisasi, dan integrasi. Untuk informasi selengkapnya, lihat panduan integrasi operasi.

jejak organisasi

Jejak yang dibuat oleh AWS CloudTrail itu mencatat semua peristiwa untuk semua Akun AWS dalam organisasi di AWS Organizations. Jejak ini dibuat di setiap Akun AWS bagian organisasi dan melacak aktivitas di setiap akun. Untuk informasi selengkapnya, lihat Membuat jejak untuk organisasi dalam CloudTrail dokumentasi.

manajemen perubahan organisasi (OCM)

Kerangka kerja untuk mengelola transformasi bisnis utama yang mengganggu dari perspektif orang, budaya, dan kepemimpinan. OCM membantu organisasi mempersiapkan, dan transisi ke, sistem dan strategi baru dengan mempercepat adopsi perubahan, mengatasi masalah transisi, dan mendorong perubahan budaya dan organisasi. Dalam strategi AWS migrasi, kerangka kerja ini disebut percepatan orang, karena kecepatan perubahan yang diperlukan dalam proyek adopsi cloud. Untuk informasi lebih lanjut, lihat panduan OCM.

kontrol akses asal (OAC)

Di CloudFront, opsi yang disempurnakan untuk membatasi akses untuk mengamankan konten Amazon Simple Storage Service (Amazon S3) Anda. OAC mendukung semua bucket S3 di semua Wilayah AWS, enkripsi sisi server dengan AWS KMS (SSE-KMS), dan dinamis `PUT` dan `DELETE` permintaan ke bucket S3.

identitas akses asal (OAI)

Di CloudFront, opsi untuk membatasi akses untuk mengamankan konten Amazon S3 Anda. Saat Anda menggunakan OAI, CloudFront buat prinsipal yang dapat diautentikasi oleh Amazon S3. Prinsipal yang diautentikasi dapat mengakses konten dalam bucket S3 hanya melalui distribusi tertentu. CloudFront Lihat juga OAC, yang menyediakan kontrol akses yang lebih terperinci dan ditingkatkan.

ORR

Lihat tinjauan kesiapan operasional.

OT

Lihat teknologi operasional.

keluar (jalan keluar) VPC

Dalam arsitektur AWS multi-akun, VPC yang menangani koneksi jaringan yang dimulai dari dalam aplikasi. Arsitektur Referensi AWS Keamanan merekomendasikan pengaturan akun Jaringan Anda dengan VPC masuk, keluar, dan inspeksi untuk melindungi antarmuka dua arah antara aplikasi Anda dan internet yang lebih luas.

P

batas izin

Kebijakan manajemen IAM yang dilampirkan pada prinsipal IAM untuk menetapkan izin maksimum yang dapat dimiliki pengguna atau peran. Untuk informasi selengkapnya, lihat Batas izin dalam dokumentasi IAM.

Informasi Identifikasi Pribadi (PII)

Informasi yang, jika dilihat secara langsung atau dipasangkan dengan data terkait lainnya, dapat digunakan untuk menyimpulkan identitas individu secara wajar. Contoh PII termasuk nama, alamat, dan informasi kontak.

PII

Lihat informasi yang dapat diidentifikasi secara pribadi.

buku pedoman

Serangkaian langkah yang telah ditentukan sebelumnya yang menangkap pekerjaan yang terkait dengan migrasi, seperti mengirimkan fungsi operasi inti di cloud. Buku pedoman dapat berupa skrip, runbook otomatis, atau ringkasan proses atau langkah-langkah yang diperlukan untuk mengoperasikan lingkungan modern Anda.

PLC

Lihat pengontrol logika yang dapat diprogram.

PLM

Lihat manajemen siklus hidup produk.

kebijakan

Objek yang dapat menentukan izin (lihat kebijakan berbasis identitas), menentukan kondisi akses (lihat kebijakanberbasis sumber daya), atau menentukan izin maksimum untuk semua akun dalam organisasi di (lihat kebijakan kontrol layanan). AWS Organizations

persistensi poliglot

Secara independen memilih teknologi penyimpanan data microservice berdasarkan pola akses data dan persyaratan lainnya. Jika layanan mikro Anda memiliki teknologi penyimpanan data yang sama, mereka dapat menghadapi tantangan implementasi atau mengalami kinerja yang buruk. Layanan mikro lebih mudah diimplementasikan dan mencapai kinerja dan skalabilitas yang lebih baik jika mereka menggunakan penyimpanan data yang paling sesuai dengan kebutuhan mereka.

penilaian portofolio

Proses menemukan, menganalisis, dan memprioritaskan portofolio aplikasi untuk merencanakan migrasi. Untuk informasi selengkapnya, lihat Mengevaluasi kesiapan migrasi.

predikat

Kondisi kueri yang mengembalikan `true` atau`false`, biasanya terletak di `WHERE` klausa.

predikat pushdown

Teknik pengoptimalan kueri database yang menyaring data dalam kueri sebelum transfer. Ini mengurangi jumlah data yang harus diambil dan diproses dari database relasional, dan meningkatkan kinerja kueri.

kontrol preventif

Kontrol keamanan yang dirancang untuk mencegah suatu peristiwa terjadi. Kontrol ini adalah garis pertahanan pertama untuk membantu mencegah akses tidak sah atau perubahan yang tidak diinginkan ke jaringan Anda. Untuk informasi selengkapnya, lihat Kontrol pencegahan dalam Menerapkan kontrol keamanan pada. AWS

principal

Entitas AWS yang dapat melakukan tindakan dan mengakses sumber daya. Entitas ini biasanya merupakan pengguna root untuk Akun AWS, peran IAM, atau pengguna. Untuk informasi selengkapnya, lihat Prinsip dalam istilah dan konsep Peran dalam dokumentasi IAM.

privasi berdasarkan desain

Pendekatan rekayasa sistem yang memperhitungkan privasi melalui seluruh proses pengembangan.

zona host pribadi

Container yang menyimpan informasi tentang bagaimana Anda ingin Amazon Route 53 merespons kueri DNS untuk domain dan subdomainnya dalam satu atau beberapa VPC. Untuk informasi selengkapnya, lihat Bekerja dengan zona yang dihosting pribadi di dokumentasi Route 53.

kontrol proaktif

Kontrol keamanan yang dirancang untuk mencegah penyebaran sumber daya yang tidak sesuai. Kontrol ini memindai sumber daya sebelum disediakan. Jika sumber daya tidak sesuai dengan kontrol, maka itu tidak disediakan. Untuk informasi selengkapnya, lihat panduan referensi Kontrol

dalam AWS Control Tower dokumentasi dan lihat Kontrol proaktif dalam Menerapkan kontrol keamanan pada AWS.

manajemen siklus hidup produk (PLM)

Manajemen data dan proses untuk suatu produk di seluruh siklus hidupnya, mulai dari desain, pengembangan, dan peluncuran, melalui pertumbuhan dan kematangan, hingga penurunan dan penghapusan.

lingkungan produksi

Lihat lingkungan.

pengontrol logika yang dapat diprogram (PLC)

Di bidang manufaktur, komputer yang sangat andal dan mudah beradaptasi yang memantau mesin dan mengotomatiskan proses manufaktur.

rantai cepat

Menggunakan output dari satu prompt LLM sebagai input untuk prompt berikutnya untuk menghasilkan respons yang lebih baik. Teknik ini digunakan untuk memecah tugas yang kompleks menjadi subtugas, atau untuk secara iteratif memperbaiki atau memperluas respons awal. Ini membantu meningkatkan akurasi dan relevansi respons model dan memungkinkan hasil yang lebih terperinci dan dipersonalisasi.

pseudonimisasi

Proses penggantian pengenal pribadi dalam kumpulan data dengan nilai placeholder. Pseudonimisasi dapat membantu melindungi privasi pribadi. Data pseudonim masih dianggap sebagai data pribadi.

publish/subscribe (pub/sub)

Pola yang memungkinkan komunikasi asinkron antara layanan mikro untuk meningkatkan skalabilitas dan daya tanggap. Misalnya, dalam MES berbasis layanan mikro, layanan mikro dapat mempublikasikan pesan peristiwa ke saluran yang dapat berlangganan layanan mikro lainnya. Sistem dapat menambahkan layanan mikro baru tanpa mengubah layanan penerbitan.

Q

rencana kueri

Serangkaian langkah, seperti instruksi, yang digunakan untuk mengakses data dalam sistem database relasional SQL.

regresi rencana kueri

Ketika pengoptimal layanan database memilih rencana yang kurang optimal daripada sebelum perubahan yang diberikan ke lingkungan database. Hal ini dapat disebabkan oleh perubahan statistik, kendala, pengaturan lingkungan, pengikatan parameter kueri, dan pembaruan ke mesin database.

R

Matriks RACI

Lihat bertanggung jawab, akuntabel, dikonsultasikan, diinformasikan (RACI).

LAP

Lihat Retrieval Augmented Generation.

ransomware

Perangkat lunak berbahaya yang dirancang untuk memblokir akses ke sistem komputer atau data sampai pembayaran dilakukan.

Matriks RASCI

Lihat bertanggung jawab, akuntabel, dikonsultasikan, diinformasikan (RACI).

RCAC

Lihat kontrol akses baris dan kolom.

replika baca

Salinan database yang digunakan untuk tujuan read-only. Anda dapat merutekan kueri ke replika baca untuk mengurangi beban pada database utama Anda.

arsitek ulang

Lihat 7 Rs.

tujuan titik pemulihan (RPO)

Jumlah waktu maksimum yang dapat diterima sejak titik pemulihan data terakhir. Ini menentukan apa yang dianggap sebagai kehilangan data yang dapat diterima antara titik pemulihan terakhir dan gangguan layanan.

tujuan waktu pemulihan (RTO)

Penundaan maksimum yang dapat diterima antara gangguan layanan dan pemulihan layanan.

refactor

Lihat 7 Rs.

Region

Kumpulan AWS sumber daya di wilayah geografis. Masing-masing Wilayah AWS terisolasi dan independen dari yang lain untuk memberikan toleransi kesalahan, stabilitas, dan ketahanan. Untuk informasi selengkapnya, lihat Menentukan Wilayah AWS akun yang dapat digunakan.

regresi

Teknik ML yang memprediksi nilai numerik. Misalnya, untuk memecahkan masalah “Berapa harga rumah ini akan dijual?” Model ML dapat menggunakan model regresi linier untuk memprediksi harga jual rumah berdasarkan fakta yang diketahui tentang rumah (misalnya, luas persegi).

rehost

Lihat 7 Rs.

melepaskan

Dalam proses penyebaran, tindakan mempromosikan perubahan pada lingkungan produksi.

memindahkan

Lihat 7 Rs.

memplatform ulang

Lihat 7 Rs.

pembelian kembali

Lihat 7 Rs.

ketahanan

Kemampuan aplikasi untuk melawan atau pulih dari gangguan. Ketersediaan tinggi dan pemulihan bencana adalah pertimbangan umum ketika merencanakan ketahanan di. AWS Cloud Untuk informasi lebih lanjut, lihat AWS Cloud Ketahanan.

kebijakan berbasis sumber daya

Kebijakan yang dilampirkan ke sumber daya, seperti bucket Amazon S3, titik akhir, atau kunci enkripsi. Jenis kebijakan ini menentukan prinsipal mana yang diizinkan mengakses, tindakan yang didukung, dan kondisi lain yang harus dipenuhi.

matriks yang bertanggung jawab, akuntabel, dikonsultasikan, diinformasikan (RACI)

Matriks yang mendefinisikan peran dan tanggung jawab untuk semua pihak yang terlibat dalam kegiatan migrasi dan operasi cloud. Nama matriks berasal dari jenis tanggung jawab yang didefinisikan dalam matriks: bertanggung jawab (R), akuntabel (A), dikonsultasikan (C), dan diinformasikan (I). Jenis dukungan (S) adalah opsional. Jika Anda menyertakan dukungan, matriks disebut matriks RASCI, dan jika Anda mengecualikannya, itu disebut matriks RACI.

kontrol responsif

Kontrol keamanan yang dirancang untuk mendorong remediasi efek samping atau penyimpangan dari garis dasar keamanan Anda. Untuk informasi selengkapnya, lihat Kontrol responsif dalam Menerapkan kontrol keamanan pada AWS.

melestarikan

Lihat 7 Rs.

pensiun

Lihat 7 Rs.

Retrieval Augmented Generation (RAG)

Teknologi AI generatif di mana LLM mereferensikan sumber data otoritatif yang berada di luar sumber data pelatihannya sebelum menghasilkan respons. Misalnya, model RAG mungkin melakukan pencarian semantik dari basis pengetahuan organisasi atau data kustom. Untuk informasi lebih lanjut, lihat Apa itu RAG.

rotasi

Proses memperbarui rahasia secara berkala untuk membuatnya lebih sulit bagi penyerang untuk mengakses kredensil.

kontrol akses baris dan kolom (RCAC)

Penggunaan ekspresi SQL dasar dan fleksibel yang telah menetapkan aturan akses. RCAC terdiri dari izin baris dan topeng kolom.

RPO

Lihat tujuan titik pemulihan.

RTO

Lihat tujuan waktu pemulihan.

buku runbook

Satu set prosedur manual atau otomatis yang diperlukan untuk melakukan tugas tertentu. Ini biasanya dibangun untuk merampingkan operasi berulang atau prosedur dengan tingkat kesalahan yang tinggi.

D

SAML 2.0

Standar terbuka yang digunakan oleh banyak penyedia identitas (IdPs). Fitur ini memungkinkan sistem masuk tunggal gabungan (SSO), sehingga pengguna dapat masuk ke Konsol Manajemen AWS atau memanggil operasi AWS API tanpa Anda harus membuat pengguna di IAM untuk semua orang di organisasi Anda. Untuk informasi lebih lanjut tentang federasi berbasis SAMP 2.0, lihat Tentang federasi berbasis SAMP 2.0 dalam dokumentasi IAM.

SCADA

Lihat kontrol pengawasan dan akuisisi data.

SCP

Lihat kebijakan kontrol layanan.

Rahasia

Dalam AWS Secrets Manager, informasi rahasia atau terbatas, seperti kata sandi atau kredensil pengguna, yang Anda simpan dalam bentuk terenkripsi. Ini terdiri dari nilai rahasia dan metadata-nya. Nilai rahasia dapat berupa biner, string tunggal, atau beberapa string. Untuk informasi selengkapnya, lihat Apa yang ada di rahasia Secrets Manager? dalam dokumentasi Secrets Manager.

keamanan dengan desain

Pendekatan rekayasa sistem yang memperhitungkan keamanan melalui seluruh proses pengembangan.

kontrol keamanan

Pagar pembatas teknis atau administratif yang mencegah, mendeteksi, atau mengurangi kemampuan pelaku ancaman untuk mengeksploitasi kerentanan keamanan. Ada empat jenis kontrol keamanan utama: preventif, detektif, responsif, dan proaktif.

pengerasan keamanan

Proses mengurangi permukaan serangan untuk membuatnya lebih tahan terhadap serangan. Ini dapat mencakup tindakan seperti menghapus sumber daya yang tidak lagi diperlukan, menerapkan praktik keamanan terbaik untuk memberikan hak istimewa paling sedikit, atau menonaktifkan fitur yang tidak perlu dalam file konfigurasi.

sistem informasi keamanan dan manajemen acara (SIEM)

Alat dan layanan yang menggabungkan sistem manajemen informasi keamanan (SIM) dan manajemen acara keamanan (SEM). Sistem SIEM mengumpulkan, memantau, dan menganalisis data dari server, jaringan, perangkat, dan sumber lain untuk mendeteksi ancaman dan pelanggaran keamanan, dan untuk menghasilkan peringatan.

otomatisasi respons keamanan

Tindakan yang telah ditentukan dan diprogram yang dirancang untuk secara otomatis merespons atau memulihkan peristiwa keamanan. Otomatisasi ini berfungsi sebagai kontrol keamanan detektif atau responsif yang membantu Anda menerapkan praktik terbaik AWS keamanan. Contoh tindakan respons otomatis termasuk memodifikasi grup keamanan VPC, menambal instans Amazon EC2, atau memutar kredensil.

enkripsi sisi server

Enkripsi data di tujuannya, oleh Layanan AWS yang menerimanya.

kebijakan kontrol layanan (SCP)

Kebijakan yang menyediakan kontrol terpusat atas izin untuk semua akun di organisasi. AWS Organizations SCP menentukan pagar pembatas atau menetapkan batasan pada tindakan yang dapat didelegasikan oleh administrator kepada pengguna atau peran. Anda dapat menggunakan SCP sebagai daftar izin atau daftar penolakan, untuk menentukan layanan atau tindakan mana

yang diizinkan atau dilarang. Untuk informasi selengkapnya, lihat Kebijakan kontrol layanan dalam AWS Organizations dokumentasi.

titik akhir layanan

URL titik masuk untuk file Layanan AWS. Anda dapat menggunakan endpoint untuk terhubung secara terprogram ke layanan target. Untuk informasi selengkapnya, lihat Layanan AWS titik akhir di Referensi Umum AWS.

perjanjian tingkat layanan (SLA)

Perjanjian yang menjelaskan apa yang dijanjikan oleh tim TI untuk diberikan kepada pelanggan mereka, seperti uptime dan kinerja layanan.

indikator tingkat layanan (SLI)

Pengukuran aspek kinerja layanan, seperti tingkat kesalahan, ketersediaan, atau throughputnya.

tujuan tingkat layanan (SLO)

Metrik target yang mewakili kesehatan layanan, yang diukur dengan indikator tingkat layanan.

model tanggung jawab bersama

Model yang menjelaskan tanggung jawab yang Anda bagikan AWS untuk keamanan dan kepatuhan cloud. AWS bertanggung jawab atas keamanan cloud, sedangkan Anda bertanggung jawab atas keamanan di cloud. Untuk informasi selengkapnya, lihat Model tanggung jawab bersama.

Bayangan AI

Aplikasi AI yang tidak sah dibuat atau digunakan di luar saluran yang diatur dalam suatu organisasi.

SIEM

Lihat informasi keamanan dan sistem manajemen acara.

titik kegagalan tunggal (SPOF)

Kegagalan dalam satu komponen penting dari aplikasi yang dapat mengganggu sistem.

SLA

Lihat perjanjian tingkat layanan.

SLI

Lihat indikator tingkat layanan.

SLO

Lihat tujuan tingkat layanan.

model split-and-seed

Pola untuk menskalakan dan mempercepat proyek modernisasi. Ketika fitur baru dan rilis produk didefinisikan, tim inti berpisah untuk membuat tim produk baru. Ini membantu meningkatkan kemampuan dan layanan organisasi Anda, meningkatkan produktivitas pengembang, dan mendukung inovasi yang cepat. Untuk informasi lebih lanjut, lihat Pendekatan bertahap untuk memodernisasi aplikasi di. AWS Cloud

SPOF

Lihat satu titik kegagalan.

skema bintang

Struktur organisasi database yang menggunakan satu tabel fakta besar untuk menyimpan data transaksional atau terukur dan menggunakan satu atau lebih tabel dimensi yang lebih kecil untuk menyimpan atribut data. Struktur ini dirancang untuk digunakan dalam gudang data atau untuk tujuan intelijen bisnis.

pola ara pencekik

Pendekatan untuk memodernisasi sistem monolitik dengan menulis ulang secara bertahap dan mengganti fungsionalitas sistem sampai sistem warisan dapat dinonaktifkan. Pola ini menggunakan analogi pohon ara yang tumbuh menjadi pohon yang sudah mapan dan akhirnya mengatasi dan menggantikan inangnya. Pola ini diperkenalkan oleh Martin Fowler sebagai cara untuk mengelola risiko saat menulis ulang sistem monolitik. Untuk contoh cara menerapkan pola ini, lihat Memodernisasi layanan web ASP.NET Microsoft (ASMX) lama secara bertahap menggunakan container dan Amazon API Gateway.

subnet

Rentang alamat IP dalam VPC Anda. Subnet harus berada di Availability Zone tunggal.

kontrol pengawasan dan akuisisi data (SCADA)

Di bidang manufaktur, sistem yang menggunakan perangkat keras dan perangkat lunak untuk memantau aset fisik dan operasi produksi.

enkripsi simetris

Algoritma enkripsi yang menggunakan kunci yang sama untuk mengenkripsi dan mendekripsi data.

pengujian sintetis

Menguji sistem dengan cara yang mensimulasikan interaksi pengguna untuk mendeteksi potensi masalah atau untuk memantau kinerja. Anda dapat menggunakan Amazon CloudWatch Synthetics untuk membuat tes ini.

sistem prompt

Teknik untuk memberikan konteks, instruksi, atau pedoman ke LLM untuk mengarahkan perilakunya. Permintaan sistem membantu mengatur konteks dan menetapkan aturan untuk interaksi dengan pengguna.

T

tag

Key-value pasangan yang bertindak sebagai metadata untuk mengatur sumber daya Anda AWS . Tag membantu Anda mengelola, mengidentifikasi, mengatur, dan memfilter sumber daya. Untuk informasi selengkapnya, lihat Menandai sumber daya AWS.

variabel target

Nilai yang Anda coba prediksi dalam ML yang diawasi. Ini juga disebut sebagai variabel hasil. Misalnya, dalam pengaturan manufaktur, variabel target bisa menjadi cacat produk.

daftar tugas

Alat yang digunakan untuk melacak kemajuan melalui runbook. Daftar tugas berisi ikhtisar runbook dan daftar tugas umum yang harus diselesaikan. Untuk setiap tugas umum, itu termasuk perkiraan jumlah waktu yang dibutuhkan, pemilik, dan kemajuan.

lingkungan uji

Lihat lingkungan.

pelatihan

Untuk menyediakan data bagi model ML Anda untuk dipelajari. Data pelatihan harus berisi jawaban yang benar. Algoritma pembelajaran menemukan pola dalam data pelatihan yang memetakan atribut data input ke target (jawaban yang ingin Anda prediksi). Ini menghasilkan model ML yang menangkap pola-pola ini. Anda kemudian dapat menggunakan model ML untuk membuat prediksi pada data baru yang Anda tidak tahu targetnya.

alat

Fungsi atau API yang dapat dipanggil agen untuk melakukan operasi di sistem eksternal.

gerbang transit

Hub transit jaringan yang dapat Anda gunakan untuk menghubungkan VPC dan jaringan lokal Anda. Untuk informasi selengkapnya, lihat Apa itu gateway transit dalam AWS Transit Gateway dokumentasi.

alur kerja berbasis batang

Pendekatan di mana pengembang membangun dan menguji fitur secara lokal di cabang fitur dan kemudian menggabungkan perubahan tersebut ke cabang utama. Cabang utama kemudian dibangun untuk pengembangan, praproduksi, dan lingkungan produksi, secara berurutan.

akses tepercaya

Memberikan izin ke layanan yang Anda tentukan untuk melakukan tugas di organisasi Anda di dalam AWS Organizations dan di akunnya atas nama Anda. Layanan tepercaya menciptakan peran terkait layanan di setiap akun, ketika peran itu diperlukan, untuk melakukan tugas manajemen untuk Anda. Untuk informasi selengkapnya, lihat Menggunakan AWS Organizations dengan AWS layanan lain dalam AWS Organizations dokumentasi.

penyetelan

Untuk mengubah aspek proses pelatihan Anda untuk meningkatkan akurasi model ML. Misalnya, Anda dapat melatih model ML dengan membuat set pelabelan, menambahkan label, dan kemudian mengulangi langkah-langkah ini beberapa kali di bawah pengaturan yang berbeda untuk mengoptimalkan model.

tim dua pizza

Sebuah DevOps tim kecil yang bisa Anda beri makan dengan dua pizza. Ukuran tim dua pizza memastikan peluang terbaik untuk berkolaborasi dalam pengembangan perangkat lunak.

U

waswas

Sebuah konsep yang mengacu pada informasi yang tidak tepat, tidak lengkap, atau tidak diketahui yang dapat merusak keandalan model ML prediktif. Ada dua jenis ketidakpastian:

ketidakpastian epistemik disebabkan oleh data yang terbatas dan tidak lengkap, sedangkan ketidakpastian aleatorik disebabkan oleh kebisingan dan keacakan yang melekat dalam data.

tugas yang tidak terdiferensiasi

Juga dikenal sebagai angkat berat, pekerjaan yang diperlukan untuk membuat dan mengoperasikan aplikasi tetapi itu tidak memberikan nilai langsung kepada pengguna akhir atau memberikan keunggulan kompetitif. Contoh tugas yang tidak terdiferensiasi termasuk pengadaan, pemeliharaan, dan perencanaan kapasitas.

lingkungan atas

Lihat lingkungan.

V

menyedot debu

Operasi pemeliharaan database yang melibatkan pembersihan setelah pembaruan tambahan untuk merebut kembali penyimpanan dan meningkatkan kinerja.

kendali versi

Proses dan alat yang melacak perubahan, seperti perubahan kode sumber dalam repositori.

Peering VPC

Koneksi antara dua VPC yang memungkinkan Anda merutekan lalu lintas dengan menggunakan alamat IP pribadi. Untuk informasi selengkapnya, lihat Apa itu peering VPC di dokumentasi VPC Amazon.

kerentanan

Kelemahan perangkat lunak atau perangkat keras yang membahayakan keamanan sistem.

W

cache hangat

Cache buffer yang berisi data terkini dan relevan yang sering diakses. Instance database dapat membaca dari cache buffer, yang lebih cepat daripada membaca dari memori utama atau disk.

data hangat

Data yang jarang diakses. Saat menanyakan jenis data ini, kueri yang cukup lambat biasanya dapat diterima.

fungsi jendela

Fungsi SQL yang melakukan perhitungan pada sekelompok baris yang berhubungan dengan catatan saat ini. Fungsi jendela berguna untuk memproses tugas, seperti menghitung rata-rata bergerak atau mengakses nilai baris berdasarkan posisi relatif dari baris saat ini.

beban kerja

Kumpulan sumber daya dan kode yang memberikan nilai bisnis, seperti aplikasi yang dihadapi pelanggan atau proses backend.

aliran kerja

Grup fungsional dalam proyek migrasi yang bertanggung jawab atas serangkaian tugas tertentu. Setiap alur kerja independen tetapi mendukung alur kerja lain dalam proyek. Misalnya, alur kerja portofolio bertanggung jawab untuk memprioritaskan aplikasi, perencanaan gelombang, dan mengumpulkan metadata migrasi. Alur kerja portofolio mengirimkan aset ini ke alur kerja migrasi, yang kemudian memigrasikan server dan aplikasi.

CACING

Lihat menulis sekali, baca banyak.

WQF

Lihat AWS Kerangka Kualifikasi Beban Kerja.

tulis sekali, baca banyak (WORM)

Model penyimpanan yang menulis data satu kali dan mencegah data dihapus atau dimodifikasi. Pengguna yang berwenang dapat membaca data sebanyak yang diperlukan, tetapi mereka tidak dapat mengubahnya. Infrastruktur penyimpanan data ini dianggap tidak dapat diubah.

Z

eksploitasi zero-day

Serangan, biasanya malware, yang memanfaatkan kerentanan zero-day.

kerentanan zero-day

Cacat atau kerentanan yang tak tanggung-tanggung dalam sistem produksi. Aktor ancaman dapat menggunakan jenis kerentanan ini untuk menyerang sistem. Pengembang sering menyadari kerentanan sebagai akibat dari serangan tersebut.

bisikan zero-shot

Memberikan LLM dengan instruksi untuk melakukan tugas tetapi tidak ada contoh (tembakan) yang dapat membantu membimbingnya. LLM harus menggunakan pengetahuan pra-terlatih untuk menangani tugas. Efektivitas bidikan nol tergantung pada kompleksitas tugas dan kualitas prompt. Lihat juga beberapa bidikan yang diminta.

aplikasi zombie

Aplikasi yang memiliki CPU rata-rata dan penggunaan memori di bawah 5 persen. Dalam proyek migrasi, adalah umum untuk menghentikan aplikasi ini.

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