Options de connectivité réseau activées AWS pour les offres SaaS

AWS Conseils prescriptifs

AWS Conseils prescriptifs: Options de connectivité réseau activées AWS pour les offres SaaS

Table of Contents

Options de connectivité réseau activées AWS pour les offres SaaS

Tomas Sykora et Luca Schumann, Amazon Web Services

Septembre 2025 (historique du document)

Ce guide explore des scénarios courants pour connecter des applications grand public à des fournisseurs de logiciels en tant que service (SaaS). Il explique comment se connecter à des ressources qui se trouvent sur site, dans le AWS Cloud cloud d'autres fournisseurs de services cloud (CSP) ou dans des architectures hybrides. Ces scénarios incluent les suivants :

- Exposition de services Web via HTTPS
- Exposer les services basés sur le protocole TCP
- Utilisation AWS AppSyncpour implémenter publish-subscribe (Pub/Sub) et GraphQL APIs
- Utilisation AWS des ressources WebSockets pour exposer des applications en temps réel
- Permettre un accès bidirectionnel pour une communication de service interactive

En s'alignant sur les meilleures pratiques décrites dans ce guide, les fournisseurs de SaaS peuvent renforcer la confiance des clients et garantir un accès évolutif, sécurisé et résilient aux offres SaaS.

Ce guide inclut également des critères d'auto-évaluation pour vous aider à évaluer dans quelle mesure vous répondez aux exigences de mise en réseau des consommateurs pour votre offre SaaS. Au-delà des modèles de connectivité, vous trouverez des comparaisons complètes des services AWS réseau, des diagrammes architecturaux de haut niveau pour différents scénarios de déploiement, ainsi que des conseils pratiques sur la manière de sélectionner la bonne approche en fonction de votre contexte commercial spécifique. Le guide explore les considérations de sécurité pour chaque option réseau, décrit les pièges courants à éviter et fournit des recommandations de mise en œuvre qui équilibrent les exigences techniques avec l'efficacité opérationnelle. En outre, vous trouverez des cadres stratégiques pour aligner vos décisions de mise en réseau sur votre modèle commercial, vos objectifs de croissance et vos besoins en matière de conformité réglementaire.

Public visé

Ce guide est destiné aux fournisseurs de SaaS. Il aide les architectes cloud, les chefs de produit et les ingénieurs réseau qui conçoivent, mettent en œuvre et optimisent la connectivité réseau pour les offres SaaS dans le AWS Cloud. Pour comprendre les concepts et les recommandations de ce guide, vous devez connaître les principes fondamentaux, les concepts AWS fondamentaux du SaaS et les principes de mise en réseau de haut niveau.

Objectifs

Ce guide décrit les options d'architecture réseau et les meilleures pratiques testées sur le terrain qui aident les consommateurs à optimiser l'accès aux offres SaaS. La mise en œuvre des recommandations de ce guide permet de réaliser les objectifs suivants :

- Facilité d'intégration : offrez un parcours client simple, de l'intégration à la production, afin d'accélérer le délai de rentabilisation de vos clients et de raccourcir leur cycle de comptabilisation des revenus.
- Adaptabilité — Intégrez parfaitement les infrastructures réseau existantes de vos clients en vous adaptant à l'évolution de leurs besoins. Cela améliore la proposition de valeur de votre produit.
- Coût total de propriété : standardisez l'accès au réseau afin de réduire les coûts de modification et les coûts par locataire. En améliorant la cohérence du déploiement, vous pouvez également réduire le temps nécessaire pour effectuer une analyse ou une réparation des causes premières.
- Gestion des dépendances — Comprenez les dépendances, les implications à long terme et les compromis des différentes options d'accès au réseau. Cela permet aux responsables de produits de prendre des décisions éclairées en matière de produits.
- Composabilité et extensibilité : dissociez le développement des fonctionnalités de base de l'infrastructure opérationnelle. Cela permet aux équipes de développement d'agir plus rapidement et de se concentrer sur la création de valeur pour vos clients.
- Favorisez la confiance : en fournissant un accès résilient, tolérant aux pannes, sécurisé et évolutif aux offres SaaS, vous pouvez réduire les risques réglementaires et gagner la confiance en votre capacité à soutenir la croissance de vos clients.

Évaluation des décisions relatives à l'accès au réseau pour les offres SaaS

Comprendre votre marché

Les décisions que vous prenez actuellement en matière de mise en réseau déterminent si la proposition de valeur de votre produit SaaS peut être fournie à vos clients. Malgré l'importance stratégique de ces décisions, l'accès à votre offre SaaS est souvent perçu comme un sujet purement technologique. Cette perception comporte des risques tels que des cycles prolongés de comptabilisation des revenus, des inefficiences opérationnelles et un décalage avec la stratégie commerciale. Par exemple, si l'expansion rapide est un objectif commercial stratégique, votre processus décisionnel doit être guidé par la question de savoir si les solutions que vous envisagez sont suffisamment évolutives et flexibles pour soutenir l'expansion. Même si vous parvenez à développer votre activité, les frais d'exploitation ne doivent pas constituer un obstacle à la croissance future, et une structure de coûts mal alignée pourrait anéantir tous vos bénéfices.

Par exemple, considérez comment les considérations de marché suivantes affectent les aspects techniques du produit, tels que la mise en réseau :

- Si votre modèle commercial est basé sur les abonnements, vos clients préféreront probablement des solutions avec des coûts récurrents prévisibles plutôt que des investissements initiaux importants.
- Si votre stratégie commerciale cible des clients d'entreprise à forte valeur ajoutée, les critères de sécurité, de gouvernance et de conformité réglementaire détermineront si votre offre SaaS sera même prise en compte.
- Si votre marché cible est principalement constitué de startups, la facilité d'intégration, le délai de rentabilisation et l'adaptabilité sont probablement des facteurs importants. Les startups privilégient généralement la rapidité et l'agilité. Parce qu'ils ont besoin de créer une marque et de générer des profits rapidement, ils sont susceptibles de préférer des solutions rapides et faciles à intégrer, capables d'évoluer de manière rentable, réduisant la dépendance à l'égard des experts et ne bloquant pas de précieux cycles.
- Certaines entreprises ont besoin d'un accès stable, à haut débit et à faible latence. Cela inclut l'industrie du divertissement et des médias, la fabrication et le traitement des transactions financières. S'il s'agit de vos clients cibles, la fiabilité est leur principale préoccupation.

Dans tous ces cas, les clients peuvent percevoir une offre SaaS par ailleurs saine si l'accès au réseau n'est pas fluide. Si le réseautage devient un obstacle, cela ne soutient pas votre analyse de rentabilisation. Si vos clients ne peuvent pas accéder de manière fiable aux services que vous proposez, la proposition de valeur de vos offres SaaS est nulle.

Comprendre votre rôle

Votre rôle dans la réalisation des objectifs commerciaux dépend de qui vous êtes, de vos objectifs individuels et de ceux de votre équipe, de l'identité de vos clients et de ce qui est important pour eux. Même si vous ne faites pas partie d'une équipe qui interagit généralement avec les clients, vous devez vous préoccuper de leur identité et de leurs besoins. Les équipes d'ingénierie et de développement doivent également se préoccuper de leurs clients internes, en particulier de ceux avec lesquels elles interagissent régulièrement. Il s'agit généralement des équipes chargées des opérations et de la réussite client.

Si vous faites partie d'une organisation commerciale, il est essentiel que vous communiquiez avec les équipes de produits et d'ingénierie au sujet du réseautage, même s'il s'agit apparemment d'un sujet purement technologique. Partagez des informations sur la structure du marché cible. Communiquez les points faibles et les besoins de vos clients et partenaires existants et potentiels. Partagez des données et des anecdotes sur les opportunités manquées, la croissance prévue par segment et les événements. Posez des questions qui remettent en question la capacité de votre organisation à soutenir la croissance de l'entreprise. Cela augmente le nombre d'opportunités et améliore la rentabilité à long terme de votre entreprise. En fin de compte, cela aide votre organisation à financer son expansion et son développement futurs.

Si vous faites partie d'une organisation d'ingénierie, comprenez la stratégie commerciale de votre organisation avant de tenter de rédiger une solution. L'alignement sur la stratégie commerciale vous aide à choisir les bons indicateurs pour évaluer les différentes options d'accès au réseau. Cela peut également empêcher une refonte coûteuse et à grande échelle du réseau au fur et à mesure que votre organisation se développe. L'alignement des activités permet à votre équipe de sécuriser et de conserver les ressources nécessaires pour relever les défis futurs. Les effectifs de votre équipe, le budget consacré au développement professionnel ou l'accès à une technologie de pointe dépendront de votre capacité à démontrer l'alignement de vos activités. Idéalement, vous pouvez montrer comment vos décisions ont contribué au succès commercial de l'organisation. Par conséquent, nous vous suggérons de saisir le processus décisionnel, y compris les critères de sélection des métriques. Passez régulièrement en revue vos indicateurs pour vous assurer qu'ils correspondent aux objectifs commerciaux. Cela peut aider votre équipe à obtenir le crédit qu'elle

mérite. Les examens périodiques permettent également de vérifier que votre équipe ne prend pas de décisions basées sur des hypothèses ou des raisons historiques obsolètes.

La liste des mesures figurant dans les sections suivantes concerne l'accès au réseau :

- Indicateurs commerciaux et relatifs aux produits
- Mesures d'ingénierie qui influencent les décisions relatives au réseau

Ce guide utilise un sous-ensemble de ces indicateurs pour vous aider à identifier les approches d'accès réseau optimales pour vos offres SaaS. Choisissez les indicateurs les plus importants et les plus pertinents pour votre entreprise, puis évaluez les approches en fonction de ces indicateurs.

Indicateurs commerciaux et relatifs aux produits qui influencent les décisions relatives au réseau

Les équipes chargées des produits et des ventes utilisent des critères de réussite pour évaluer si elles atteignent leurs objectifs commerciaux. Cette section décrit les indicateurs commerciaux ou relatifs aux produits qui peuvent être influencés positivement ou négativement par les décisions prises par votre organisation en matière d'accès au réseau.

Utilisez ces indicateurs et ces questions d'auto-évaluation pour évaluer dans quelle mesure votre approche de l'accès au réseau correspond au positionnement de votre entreprise et à votre stratégie de marché. Cette évaluation vous aide à déterminer si vos décisions actuelles en matière de mise en réseau soutiennent la différenciation de votre entreprise sur le marché, ses avantages concurrentiels et les besoins de son public cible.

Cette section contient des métriques et des questions d'auto-évaluation pour les sujets suivants :

- Modèle commercial et positionnement sur le marché
- Marché adressable total, taux d'acquisition de nouveaux clients, croissance et évolutivité
- Expérience client et fidélisation
- Efficacité et performance financière
- Conformité réglementaire et gestion des risques
- Stratégie des partenaires

Modèle commercial et positionnement sur le marché

Ces indicateurs concernent la position de votre entreprise sur le marché, notamment la différenciation concurrentielle, la portée du marché et la perception de la marque. Il est essentiel que vous évaluiez l'alignement entre l'approche d'accès au réseau et le modèle commercial. Réalisez une évaluation, qu'elle soit basée sur un abonnement, basée sur l'utilisation, freemium, hiérarchisée, sur le marché, axée sur les API ou en marque blanche. Assurez-vous que le modèle soutient les objectifs de l'organisation et ceux des clients.

Critères de score élevé

L'approche de l'accès au réseau s'aligne parfaitement sur le modèle commercial. Cela facilite l'adoption et la prestation du service. Il soutient la viabilité financière à long terme du modèle commercial et la structure des coûts est compatible avec la croissance attendue. Cela minimise les frictions pour les clients ou les partenaires lors de l'adoption de l'offre. Cela améliore l'expérience utilisateur et encourage une plus large adoption du service.

Indicateurs à faible score

L'approche d'accès au réseau sélectionnée n'est pas conforme au modèle commercial qu'elle devrait prendre en charge. La structure des coûts et le délai de déploiement constituent un obstacle à l'adoption sur le marché cible. Les coûts d'infrastructure et d'exploitation permanents inhibent tout profit potentiel. Cela empêche la croissance de l'entreprise et rend difficile le fonctionnement à l'échelle prévue. Par ailleurs, les propriétés de l'approche d'accès au réseau peuvent empêcher les clients d'envisager le service pour des raisons réglementaires.

Questions d'auto-évaluation

- Quelles sont les implications financières de l'approche d'accès au réseau sélectionnée pour le déploiement initial et la livraison continue ? Quels sont les coûts fixes et variables de l'approche ?
- L'approche de l'accès au réseau peut-elle évoluer de manière efficace et efficiente pour répondre aux exigences croissantes du modèle commercial ? Tenez compte de la taille de chaque locataire et du nombre de locataires intégrés.
- L'approche de l'accès au réseau impose-t-elle des limites techniques ou opérationnelles susceptibles de limiter la flexibilité ou l'adaptabilité du modèle commercial ?
- En ce qui concerne l'approche d'accès au réseau, comment le délai de déploiement correspond-il à la vitesse de mise sur le marché requise par le modèle commercial ?

Marché adressable total, taux d'acquisition de nouveaux clients, croissance et évolutivité

Il est essentiel que vous évaluiez l'impact des décisions relatives au réseau sur la capacité de l'entreprise à se développer sur de nouveaux marchés, à acquérir des clients efficacement et à maintenir l'évolutivité opérationnelle. Ces facteurs influent sur les taux de conversion. Ils déterminent également si l'approche d'accès au réseau favorise l'expansion dans des segments de marché importants ou vous limite à ne servir que des types de clients spécifiques.

Critères de score élevé

L'approche d'accès au réseau aide l'organisation à atteindre une partie importante du marché cible, ou elle peut être combinée efficacement avec d'autres approches réseau pour étendre la portée du marché. Cette approche ne devrait nécessiter qu'un effort d'intégration supplémentaire minimal. Cette approche prend en charge des délais de déploiement courts, une entrée rapide sur le marché et une expansion. Il permet un grand nombre de déploiements parallèles. L'intégration est simple pour les clients, ce qui réduit les obstacles à l'adoption et améliore l'expérience client. L'approche minimise les frais d'exploitation, préserve la capacité opérationnelle et soutient les prévisions de croissance.

Indicateurs à faible score

L'approche d'accès au réseau ne prend en charge qu'une petite partie du marché cible ou convient principalement aux segments de niche qui ne sont pas prioritaires dans la stratégie commerciale. Elle ne complète pas efficacement les autres approches d'accès au réseau déjà prises en charge. Les délais de déploiement sont décalés par rapport aux demandes du marché, ce qui limite l'expansion du marché et l'acquisition de nouveaux clients. Le modèle de déploiement est séquentiel, ce qui augmente les risques d'engorgement des services à mesure que la demande augmente. Les processus d'intégration complexes dissuadent les clients potentiels, ce qui a une incidence négative sur le taux d'acquisition et les taux de conversion. Des frais d'exploitation importants diminuent la capacité opérationnelle de l'organisation. Cela devient un obstacle à la croissance prévue.

Pour ces indicateurs, évaluez si l'introduction d'une nouvelle approche d'accès au réseau peut aider l'organisation à atteindre ses objectifs commerciaux stratégiques. Déterminez si la nouvelle approche d'accès au réseau risque de créer de nouvelles dépendances entre les produits ou de consommer des ressources opérationnelles sans obtenir les résultats souhaités.

Questions d'auto-évaluation

- Y a-t-il des lacunes dans l'approche actuelle qui vous empêchent d'atteindre de plus grands segments du marché cible ?
- Quelle est la liste minimale d'approches d'accès au réseau normalisées et non chevauchantes que vous devriez prendre en charge pour couvrir 70 à 90 % du marché cible ?
- Quelle est la portée de chaque approche d'accès au réseau et quelles sont les augmentations connexes des indicateurs importants, tels que les coûts d'infrastructure, les cycles opérationnels et la dépendance à l'égard des experts ?
- Comment les capacités de déploiement et les limites de service de l'infrastructure réseau correspondent-elles aux attentes de croissance de vos marchés cibles ?
- L'intégration du réseau crée-t-elle des obstacles à l'entrée de nouveaux clients ? Comment y remédier pour améliorer les taux de conversion ?
- Comment les frais opérationnels liés à la gestion du réseau affectent-ils votre capacité de croissance et d'évolutivité ?
- Quelles stratégies pouvez-vous mettre en œuvre pour réduire les délais de déploiement du réseau et améliorer l'expansion du marché et l'acquisition de clients ?
- Existe-t-il des dépendances à l'égard des ressources d'experts qui retarderaient le déploiement ou l'intégration dans les écosystèmes des clients ?

Expérience client et fidélisation

Les indicateurs présentés dans cette section vous aident à comprendre la capacité de votre entreprise à acquérir des clients et, surtout, à les fidéliser. Comprendre la relation entre les approches d'accès au réseau et la satisfaction des clients peut aider les équipes de produits et d'ingénierie à prendre des décisions éclairées par les données.

Critères de score élevé

L'approche d'accès au réseau est fiable et facile à gérer. Cela contribue à un niveau élevé de satisfaction client (CSAT) et à des résultats de score net de promotion (NPS) élevés. Ces scores sont révélateurs d'une solide réputation de marque et de la fidélité de la clientèle. Grâce à une intégration parfaite avec les écosystèmes existants de vos clients, les frictions en matière d'adoption sont faibles et la dépendance à l'égard des experts est faible. Votre organisation respecte systématiquement les accords de niveau de service (SLAs), ce qui renforce la confiance des clients et les obligations

contractuelles. Parce que les clients bénéficient de services stables et fiables, vous avez un taux de fidélisation élevé.

Indicateurs à faible score

Une intégration difficile et un accès irrégulier aux services sont souvent source de frustration chez les clients et de commentaires négatifs. Cela nuit à la réputation de la marque. Les nouveaux clients ne parviennent pas à passer des forfaits gratuits ou d'essai aux services payants en raison de leur dépendance à l'égard d'experts ou en raison de délais d'intégration et d'intégration prolongés. Les manquements fréquents SLAs entraînent des pénalités financières et une perte de crédibilité, ce qui peut réduire le taux de fidélisation des clients.

Questions d'auto-évaluation

- Comment les performances du réseau (telles que la vitesse, le temps de disponibilité et la latence) affectent-elles directement les résultats CSAT et NPS ? Quelles améliorations spécifiques du réseau pourraient entraîner une hausse de ces scores ?
- Comment les indicateurs actuels de latence et de disponibilité du réseau affectent-ils l'expérience utilisateur initiale et les taux d'adoption ? Quelles améliorations spécifiques des performances du réseau sont nécessaires pour optimiser ces indicateurs ?
- Existe-t-il des problèmes récurrents liés aux configurations réseau ou aux paramètres de sécurité qui compliquent l'intégration pour les nouveaux clients ? Comment pouvez-vous rationaliser ces processus ?
- Comment la facilité de configuration de l'accès au réseau affecte-t-elle l'expérience d'intégration des nouveaux utilisateurs ? Existe-t-il des points d'accès au réseau ou des délais spécifiques qui peuvent être optimisés pour améliorer les impressions initiales des utilisateurs ?
- Quels sont les défis liés à l'automatisation de la fourniture de services réseau pour les nouveaux clients ? Comment ajuster ce processus pour améliorer l'évolutivité et la fiabilité ?
- Analysez les causes profondes des récentes violations des SLA. Étaient-ils liés à la configuration du réseau, à la planification des capacités ou à des problèmes liés à des fournisseurs externes ?
- À quelle fréquence les problèmes de réseau vous empêchent-ils de respecter vos engagements en matière de SLA ? Quelles sont les pannes réseau les plus fréquentes ?
- Quelles améliorations des performances du réseau ont eu l'impact positif le plus significatif sur la satisfaction des clients par le passé ?

Efficacité et performance financière

Cette catégorie évalue les aspects liés à la santé financière et à la rentabilité de votre entreprise, tels que la rentabilité, la viabilité à long terme, la rentabilité, le retour sur investissement (ROI) et le coût total de possession (TCO). En rationalisant les opérations du réseau grâce à la standardisation, vous pouvez réduire les frais d'exploitation et les coûts de maintenance. Cela soutient les objectifs de croissance de votre organisation.

Critères de score élevé

La structure des coûts de l'approche d'accès au réseau est bien alignée sur le modèle commercial. Il soutient une croissance durable et les importantes économies de coûts que vous réalisez augmentent la rentabilité. L'accès efficace au réseau permet une intégration rapide des clients, ce qui réduit le délai nécessaire pour créer de la valeur et accélère la pénétration du marché. Cela raccourcit directement le cycle de comptabilisation des recettes.

Indicateurs à faible score

Les clients se tournent vers vos concurrents pour accélérer la livraison de leurs applications et services. Votre organisation a augmenté les coûts d'exploitation associés à des configurations réseau complexes et variées et à des délais prolongés. La structure des coûts et le modèle commercial ne sont pas alignés, ce qui peut entraîner des coûts initiaux élevés pour les services par abonnement. Les processus d'intégration fastidieux réduisent la pénétration du marché et retardent la comptabilisation des revenus.

Questions d'auto-évaluation

- Quels sont les délais actuels pour le déploiement de nouveaux services, et comment influent-ils sur le délai de commercialisation et la comptabilisation des revenus ?
- Dans quelle mesure les opérations réseau normalisées réduisent-elles efficacement les frais généraux et les coûts de maintenance ?
- Des ressources spécialisées sont-elles nécessaires pour mener à bien l'intégration initiale, opérer au quotidien, résoudre les problèmes ou mettre en œuvre des modifications ?
- Dans quelle mesure les investissements actuels dans les réseaux sont-ils durables en termes d'avancées technologiques ? Investissez-vous dans des technologies pérennes qui correspondent aux évolutions prévues du marché ?
- Dans quelle mesure répartissez-vous et suivez-vous efficacement les coûts liés au trafic réseau et à l'utilisation par les locataires individuels ?

Conformité réglementaire et gestion des risques

Il est fondamental de valider la conformité aux réglementations relatives au réseau. Cela confirme que vous opérez légalement et que vous pouvez conserver la confiance de vos clients. La normalisation des opérations du réseau simplifie le processus de conformité et favorise la cohérence entre les différentes juridictions et zones géographiques. Ces mesures vous aident à étendre vos services.

Critères de score élevé

Les opérations du réseau respectent systématiquement les normes légales sans complications, ce qui contribue à l'expansion du marché, réduit les difficultés d'adoption et renforce la confiance des clients. Une conformité démontrée aux cadres réglementaires critiques, tels que la loi DORA (Digital Operational Resilience Act) et le National Institute of Standards and Technology (NIST), vous aide à gagner des clients sensibles à la conformité réglementaire. La visibilité continue de votre statut de conformité réduit le temps nécessaire à la réalisation d'un audit.

Indicateurs à faible score

Les lacunes en matière de conformité du réseau entraînent de fortes frictions en matière d'adoption, des retards dans le lancement des services, des contestations juridiques et des amendes potentielles. Ces défis entraînent des retards ou des annulations de plans d'expansion sur de nouveaux marchés. Il est difficile de maintenir des pratiques de conformité standard dans différentes juridictions, ce qui affecte l'efficacité opérationnelle et la réputation du marché.

Questions d'auto-évaluation

- Dans quelle mesure les opérations de votre réseau sont-elles conformes aux directives réglementaires ou sectorielles applicables ? Qu'a révélé votre dernier audit de conformité ?
- Comment maintenez-vous la conformité aux nouvelles réglementations dans les domaines de la sécurité numérique et des réseaux ?
- Dans quelle mesure votre processus de documentation et de reporting répond-il efficacement aux exigences des différents organismes de réglementation ?
- Quelles stratégies de gestion des risques avez-vous mises en place pour identifier et traiter les risques de conformité potentiels avant qu'ils n'entraînent des contestations juridiques ?
- De quel niveau de formation et de sensibilisation à la conformité vos équipes de gestion de réseau ont-elles besoin pour soutenir vos approches en matière d'accès au réseau ?

Stratégie des partenaires

Évaluez dans quelle mesure l'approche d'accès au réseau s'aligne sur un écosystème de partenaires, de plateformes et de places de marché reconnus. Cela est essentiel, en particulier si votre stratégie de croissance dépend de la mise à l'échelle par le biais de partenaires.

Critères de score élevé

L'approche d'accès au réseau est intégrée à l'ensemble de votre écosystème de partenaires. Sa structure de coûts correspond parfaitement aux modèles commerciaux de vos principaux partenaires. Les partenaires possèdent les compétences nécessaires en matière de mise en réseau pour une intégration fluide de vos offres SaaS, et ils peuvent fournir un accès et des fonctionnalités durables.

Indicateurs à faible score

L'approche d'accès au réseau choisie exige des compétences, des ressources ou des équipements spécialisés rares ou difficiles à acquérir. Il diffère des protocoles d'accès réseau standard couramment utilisés par les plateformes et les places de marché. Il en résulte une structure de coûts imprévisible qu'il est difficile de concilier. L'approche d'accès au réseau n'est pas alignée sur les modèles commerciaux de vos principaux partenaires.

Questions d'auto-évaluation

- Quelles sont les implications financières de l'approche d'accès au réseau pour les partenaires ? Comment ces coûts s'alignent-ils sur leurs modèles commerciaux ? Quel aspect de l'intégration supporte l'essentiel des coûts, et combien de cycles opérationnels faut-il investir ?
- En ce qui concerne l'approche d'accès au réseau, existe-t-il des obstacles à l'intégration ou à la maintenance susceptibles d'affecter les relations avec les partenaires ou l'évolutivité de l'écosystème ?
- Comment optimiser l'approche d'accès au réseau pour améliorer la compatibilité et faciliter l'intégration au sein de l'écosystème ?

Mesures d'ingénierie qui influencent les décisions relatives au réseau

À l'instar des équipes de produits et commerciales, les équipes d'ingénierie utilisent également des critères de réussite pour évaluer si elles atteignent leurs objectifs commerciaux. Cependant,

ces indicateurs diffèrent et se concentrent sur la capacité de l'équipe à développer, à exploiter et à respecter les exigences de sécurité et de conformité. Cette section décrit les mesures d'ingénierie qui peuvent être influencées positivement ou négativement par les décisions prises par votre organisation en matière d'accès au réseau.

Utilisez ces indicateurs et ces questions d'auto-évaluation pour évaluer votre approche actuelle en matière d'accès au réseau par rapport aux exigences de votre entreprise et à vos capacités techniques. Cette évaluation vous aide à identifier les lacunes de votre architecture et à prioriser les améliorations conformes à vos objectifs stratégiques. En revoyant régulièrement ces critères, vous pouvez vous assurer que votre stratégie d'accès au réseau continue de répondre aux besoins de vos clients et aux plans de croissance de votre entreprise.

Cette section contient des métriques et des questions d'auto-évaluation pour les catégories et sujets suivants :

Indicateurs de développement liés à l'accès au réseau pour les offres SaaS

Cette section contient les métriques suivantes :

- Qualité du code et performance de l'équipe d'ingénierie
- Réduction de la dette technique
- Évolutivité, capacité et performances

Fréquence de déploiement, délai de déploiement et vitesse du sprint

Pour optimiser l'efficacité du cycle de développement, il est essentiel que vous compreniez l'influence du provisionnement du stack réseau sur la vitesse des sprints.

Critères de score élevé

Le provisionnement de la pile réseau est rationalisé et automatisé, et ne nécessite qu'une intervention manuelle minimale. Cela n'a pas d'impact significatif sur la vitesse du sprint. Le provisionnement et le redéploiement de la pile réseau peuvent être effectués par n'importe quel membre de l'équipe. Cela réduit les goulets d'étranglement et les dépendances à l'égard de ressources spécialisées.

Indicateurs à faible score

Un grand nombre de points d'histoire sont nécessaires pour approvisionner la pile réseau. Cela suggère un processus complexe et long qui nuit au développement de nouvelles fonctionnalités. Le redéploiement fréquent de la pile réseau entraîne des coûts importants en termes de temps et d'argent. Les tâches de provisionnement du réseau nécessitent une expertise technique spécialisée, ce qui crée des goulots d'étranglement et ralentit le cycle de développement.

Questions d'auto-évaluation

- Quelles étapes manuelles, le cas échéant, sont impliquées dans le processus de déploiement. Quel est leur impact sur la fréquence et le temps de déploiement ?
- Comment les annulations sont-elles gérées en cas d'échec du déploiement ? Quel est leur impact sur la fréquence de déploiement et le temps de reprise ?
- Combien de points d'histoire sont nécessaires pour approvisionner la pile réseau lorsque vous configurez de nouveaux environnements ?
- Quels sont les coûts supplémentaires et les délais associés au redéploiement fréquent de la pile réseau au cours du processus de développement ?
- Le provisionnement de la pile réseau dépend-il d'une expertise technique spécialisée ou s'agit-il d'une tâche qui peut être gérée par n'importe quel membre de l'équipe ?

Flexibilité et fourniture de fonctionnalités

L'approche d'accès au réseau peut influencer la capacité de l'équipe d'ingénierie à innover et à déployer efficacement de nouvelles fonctionnalités.

Critères de score élevé

L'approche d'accès au réseau offre la flexibilité nécessaire pour un déploiement rapide et fluide des fonctionnalités. Il prend en charge un large éventail de protocoles de communication, de communications unidirectionnelles et bidirectionnelles et de tailles de messages. Il n'impose pas de contraintes importantes aux processus de développement ou à l'innovation.

Indicateurs à faible score

L'approche d'accès au réseau limite la capacité de l'équipe à déployer de nouvelles fonctionnalités en raison de l'absence de protocoles de communication pris en charge, de la rigidité de la taille des messages ou de la dépendance à l'égard de technologies spécifiques et de ressources spécialisées associées. Cela peut ralentir les cycles de développement et entraver l'évolution du service.

Questions d'auto-évaluation

- Quel est l'impact de l'approche d'accès au réseau sur l'agilité de l'équipe dans le développement et le déploiement de nouvelles fonctionnalités ?
- L'approche d'accès au réseau comporte-t-elle des limites qui limitent la prise en charge de certains protocoles ou technologies de communication ?
- Comment l'approche facilite-t-elle ou limite-t-elle l'intégration des nouvelles technologies et innovations dans le service ?
- Comment l'approche d'accès au réseau affecte-t-elle les délais de développement et la feuille de route du produit ?

Modifier le taux d'échec

L'approche d'accès réseau que vous choisissez peut affecter le taux d'échec des modifications lors du déploiement de nouveaux services ou fonctionnalités. Un meilleur contrôle signifie souvent une plus grande flexibilité, mais cela augmente également le risque de mauvaises configurations, par exemple lors de la gestion d'une configuration de routage complexe.

Critères de score élevé

Vous pouvez apporter des modifications à la pile réseau avec un risque de défaillance minimal. Des mécanismes de test suffisants sont présents, des mécanismes de restauration efficaces existent et une surveillance efficace vous aide à identifier et à résoudre rapidement les problèmes.

Indicateurs à faible score

L'approche d'accès au réseau est sujette à des défaillances lors des modifications. Les options de test sont limitées, les stratégies de déploiement sont complexes ou les capacités de surveillance et de dépannage sont insuffisantes. Plusieurs parties sont tenues de participer aux sessions de résolution des problèmes. Cela peut entraîner une augmentation des temps d'arrêt et une diminution de la disponibilité de l'offre SaaS.

Questions d'auto-évaluation

- Quelles sont les mesures mises en place pour atténuer le risque d'échec des modifications lors de la mise à jour de la pile réseau ?
- Existe-t-il des processus de test et de validation approfondis ?
- À quelle vitesse le système peut-il se rétablir après un échec de modification ? Existe-t-il un processus de rétrogradation efficace ?
- Existe-t-il des systèmes de surveillance et d'alerte proactifs permettant de détecter et de résoudre les problèmes rapidement pendant et après les modifications du stack réseau ?
- Quel est le taux d'échec historique des modifications pour les déploiements de stack réseau ? Quelles leçons ont été tirées des incidents passés ?
- Comment l'approche d'accès au réseau facilite-t-elle ou limite-t-elle la mise en œuvre des changements ? L'approche minimise-t-elle les interruptions de service ?
- Quel est le risque d'impact sur la disponibilité de l'offre SaaS dans l'environnement de production lorsque vous déployez des modifications impliquant l'approche d'accès au réseau ?

Qualité du code et performance de l'équipe d'ingénierie

Les approches d'accès au réseau peuvent affecter indirectement la qualité du code des offres SaaS. Le manque de standardisation de l'accès au réseau peut obliger l'équipe d'ingénierie à prendre en charge plusieurs approches d'intégration, ce qui peut entraîner une base de code gonflée. Cela peut à son tour entraver la capacité de l'équipe à développer la profondeur et le contrôle de la qualité du code nécessaires pour maintenir des équipes d'ingénierie performantes.

Critères de score élevé

L'équipe d'ingénierie reste concentrée grâce à la modularité du code et à sa réutilisabilité dans le cadre des approches d'accès réseau prises en charge. Les approches d'accès au réseau sont compatibles avec les pipelines de déploiement existants et les stratégies de test automatisées.

Indicateurs à faible score

Les performances de l'équipe d'ingénierie sont réduites en raison des frais généraux associés à l'intégration et à la maintenance d'un trop grand nombre d'approches d'accès au réseau. Certaines approches augmentent considérablement la complexité, génèrent des dettes technologiques ou nécessitent le développement de solutions de contournement pour remédier aux capacités manquantes ou insuffisantes.

Questions d'auto-évaluation

- Comment l'approche d'accès au réseau gère-t-elle la variabilité du réseau ?
- Avez-vous besoin de développer un code supplémentaire pour gérer les perturbations de connectivité ?
- Une nouvelle approche d'accès au réseau s'intègre-t-elle parfaitement aux approches existantes ou nécessite-t-elle un développement personnalisé important ?
- Quelle est l'ampleur du changement nécessaire pour adopter une nouvelle approche d'accès au réseau ? La base de code existante et les tests automatisés peuvent-ils être utilisés efficacement ?
- Est-il facile ou difficile de déployer ou de redéployer le service avec l'approche d'accès au réseau sélectionnée ? Est-ce que cela peut être fait fréquemment ? Existe-t-il des dépendances à l'égard des ressources d'experts ?
- L'approche d'accès au réseau facilite-t-elle ou complique-t-elle le respect des normes de codage et des meilleures pratiques ?
- Comment cette approche affecte-t-elle time-to-market les nouvelles fonctionnalités ou les correctifs ?

Réduction de la dette technique

Une évaluation de l'impact d'une approche d'accès au réseau sur la dette technique doit prendre en compte ses capacités d'évolutivité, d'observabilité et de sécurité.

Critères de score élevé

Cette approche rationalise efficacement la gestion de l'infrastructure à mesure que la clientèle s'élargit. Il offre de solides capacités d'observabilité. out-of-the-box Cela favorise une surveillance et une maintenance efficaces.

Indicateurs à faible score

L'approche d'accès au réseau ne sécurise pas suffisamment les canaux de communication et ne dispose pas d'outils suffisants pour une observation métrique qualitative. Cela peut également nécessiter un développement supplémentaire pour la gestion de l'infrastructure à mesure que la clientèle augmente, ou des solutions de contournement des problèmes de fiabilité.

Questions d'auto-évaluation

- Comment l'approche d'accès au réseau influence-t-elle l'évolutivité à long terme de l'infrastructure ? Facilite-t-il une croissance fluide avec un investissement supplémentaire minimal ?
- Dans quelle mesure les outils d'observabilité inclus sont-ils complets ? Permettent-ils une surveillance et une résolution proactives des problèmes ?
- Quel est l'impact prévu de l'approche d'accès au réseau sur la maintenance et l'évolution de la base de code au fil du temps ?
- L'approche s'intègre-t-elle bien à l'infrastructure existante et prévue ? Cela nécessite-t-il des modifications ou des ajouts importants ?

Évolutivité, capacité et performances

Pour déterminer la pertinence d'une approche d'accès réseau pour une offre SaaS, il est essentiel d'analyser comment elle maintient des performances optimales à mesure que la demande augmente.

Critères de score élevé

L'approche de l'accès au réseau facilite parfaitement l'expansion. Il maintient une faible latence pendant le traitement des demandes et gère efficacement les pics de trafic. Il fournit des performances constantes indépendamment de l'augmentation du trafic et n'impose aucune limite opérationnelle à la croissance.

Indicateurs à faible score

L'approche d'accès au réseau ne s'adapte pas efficacement, probablement en raison de limites inhérentes à la bande passante ou d'une capacité d'infrastructure insuffisante. Le provisionnement et

la gestion des ressources augmentent la complexité ou créent des dépendances. Les performances du service sont dégradées en raison de l'augmentation de la latence, de l'instabilité et de la variabilité du débit, en particulier dans des conditions de congestion du réseau.

Questions d'auto-évaluation

- Comment l'approche d'accès au réseau s'adapte-t-elle à un nombre croissant de locataires et à leurs volumes de données ?
- Est-il intrinsèquement évolutif pour répondre aux demandes futures ?
- Quelles sont les mesures mises en place pour garantir la cohérence des performances, même en période de pic de trafic ou lors d'événements de mise à l'échelle rapide ?
- Comment cette approche gère-t-elle la latence et l'instabilité du réseau ? Existe-t-il des mécanismes permettant d'optimiser le débit des données et de minimiser les délais ?
- L'approche d'accès au réseau peut-elle s'adapter aux différentes conditions du réseau ? Peut-il offrir une expérience à locataire unique à chaque client ?
- Quel est l'impact de l'approche d'accès au réseau sur l'infrastructure sous-jacente ? Cela nécessite-t-il des mises à niveau ou des modifications importantes des systèmes existants ?

Indicateurs d'excellence opérationnelle liés à l'accès au réseau pour les offres SaaS

Cette section contient les métriques suivantes :

- Résilience opérationnelle et reprise après sinistre
- Surveillance des performances des services et des applications

Résilience opérationnelle et reprise après sinistre

L'approche d'accès au réseau devrait aider l'offre SaaS à résister à différents types de perturbations et à se remettre rapidement en cas de sinistre.

Critères de score élevé

Les plans de reprise après sinistre établis et testés montrent systématiquement que l'approche d'accès au réseau répond aux exigences de reprise après sinistre. L'approche d'accès au réseau prend en charge les configurations à haute disponibilité et prend en charge les mécanismes de basculement automatiques, rapides et fiables.

Indicateurs à faible score

L'approche d'accès au réseau rend difficile l'élaboration d'une stratégie cohérente de reprise après sinistre. Vous observez des temps de récupération prolongés après des interruptions. Les défaillances opérationnelles fréquentes de l'infrastructure réseau ont un impact sur la prestation de services.

Questions d'auto-évaluation

- À quand remonte le dernier exercice de reprise après sinistre et quels en ont été les résultats ?
- Combien de temps faut-il pour rétablir les services critiques après une interruption ? Quelle partie de l'infrastructure réseau doit être redéployée ?
- Quelles améliorations peuvent être apportées à l'infrastructure réseau pour rationaliser vos plans de reprise après sinistre ?
- Des redondances sont-elles en place pour les composants réseau les plus critiques ?
- Avez-vous automatisé le redéploiement potentiel de l'infrastructure réseau après une panne critique ?
- Comment l'approche d'accès au réseau favorise-t-elle la tolérance aux pannes et la fiabilité ? Existe-t-il des mécanismes intégrés pour gérer les interruptions du réseau et préserver l'intégrité des données ?

Surveillance des performances des services et des applications

L'approche d'accès réseau peut affecter les outils de surveillance des performances utilisés pour valider le fonctionnement optimal et la disponibilité des services. Selon le service, vous pouvez avoir accès à des métriques de bas niveau (telles que les taux de perte de paquets) ou à des métriques de niveau supérieur (telles que la durée de session). Les indicateurs de bas niveau fournissent des informations techniques détaillées sur le comportement du réseau, mais leur interprétation peut être complexe. En revanche, les indicateurs de niveau supérieur constituent souvent un moyen plus direct et plus simple d'évaluer l'expérience utilisateur globale. Cela s'explique par le fait qu'ils regroupent l'impact des conditions sous-jacentes du réseau sous forme d'indicateurs clairs de qualité de service.

Critères de score élevé

Des outils de surveillance complets fournissant des informations en temps quasi réel sont facilement disponibles. Vous disposez de systèmes d'alerte et de réponse automatisés qui résolvent les problèmes de performance. Vous pouvez prévoir les blocages ou défaillances potentiels des services avant qu'ils n'affectent les utilisateurs.

Indicateurs à faible score

Des interruptions de service ou des problèmes de performance fréquents se produisent sans être observés ni pris en compte. Le manque de visibilité sur les performances des services entraîne une lenteur à réagir aux problèmes de performance. Des équipes multipartites sont nécessaires pour résoudre les problèmes d'infrastructure réseau.

Questions d'auto-évaluation

- Quels outils de surveillance et indicateurs d'infrastructure réseau sont actuellement disponibles ? Dans quelle mesure sont-ils efficaces pour détecter les anomalies de service ?
- À quelle vitesse pouvez-vous identifier et résoudre les problèmes de performance ?
- Disposez-vous de mécanismes permettant de prévoir les problèmes de performance potentiels ?
- Quelles améliorations pouvez-vous apporter pour améliorer les capacités d'observabilité ?

Mesures de sécurité et de gouvernance liées à l'accès au réseau pour les offres SaaS

Cette section contient les métriques suivantes :

- Gestion de la sécurité, de la conformité et des vulnérabilités

Gestion de la sécurité, de la conformité et des vulnérabilités

Il est essentiel que vous évaluiez les aspects de sécurité de l'approche d'accès au réseau, notamment le respect des normes de sécurité et la gestion des vulnérabilités.

Critères de score élevé

L'approche de l'accès au réseau aide votre équipe à respecter les cadres de sécurité, tels que l'Organisation internationale de normalisation (ISO) 27001, le System and Organization Controls 2 (SOC 2) ou le NIST. Cela facilite la réalisation d'audits de sécurité réguliers. De solides mécanismes de cryptage et d'authentification sont en place. Les réseaux sont isolés et seules les ressources nécessaires sont exposées à l'infrastructure du client. Vous pouvez détecter les anomalies du réseau en temps quasi réel, sans surcharge excessive.

Indicateurs à faible score

L'approche d'accès au réseau est sujette à des failles de sécurité ou à des vulnérabilités récurrentes, et elle n'est pas conforme aux principales normes de sécurité. Vous observez fréquemment des retards dans la détection et les réponses aux incidents de sécurité.

Questions d'auto-évaluation

- Y a-t-il eu récemment des failles de sécurité liées à l'approche d'accès au réseau sélectionnée, et qu'en avons-nous appris ?
- Dans quelle mesure votre approche de l'accès au réseau est-elle conforme aux normes de sécurité internationales ?
- Combien de temps faut-il pour détecter les menaces de sécurité et y répondre ? Comment l'accès au réseau aide-t-il ou limite-t-il cette capacité ?
- À quelle fréquence les évaluations de sécurité sont-elles effectuées sur les approches d'accès au réseau ? Pouvez-vous utiliser des outils courants pour évaluer la sécurité de l'approche d'accès au réseau, ou un logiciel spécialisé est-il requis ?
- Quel est le niveau de sécurité inhérent à l'approche d'accès au réseau, et comment est-il conforme aux meilleures pratiques du secteur et aux exigences réglementaires ?

Vue d'ensemble des services AWS réseau pour les offres SaaS

Cette section décrit les services AWS réseau référencés dans ce guide. Il compare également leurs capacités et décrit les considérations de sécurité pour chaque service.

Cette section contient les rubriques suivantes :

- AWS services de mise en réseau
- Comparaison des capacités des services
- Caractéristiques et considérations relatives à la sécurité

AWS services de mise en réseau

Les Services AWS points suivants sont abordés de manière cohérente dans ce guide.

AWS PrivateLink

AWS PrivateLinkest un service cloud natif qui peut donner accès à votre offre SaaS si vos clients opèrent déjà dans le AWS Cloud. Votre client se connecte à l'offre SaaS via un point de terminaison VPC d'interface. Il s'agit d'une interface réseau de point de terminaison qui est provisionnée dans un ou plusieurs sous-réseaux du client. Compte AWS Dans les scénarios présentés dans ce guide, le trafic passe par le point de terminaison VPC de l'interface et arrive à un Network Load Balancer de votre compte. Le Network Load Balancer transmet le trafic vers l'application SaaS, que vous avez enregistrée en tant que service de point de terminaison. Les points de terminaison VPC AWS PrivateLink peuvent également vous aider à accéder à d'autres ressources, telles que des bases de données.

Amazon VPC Lattice

Amazon VPC Lattice est un service de mise en réseau d'applications qui aide les fournisseurs de SaaS à proposer leurs services de manière sécurisée et efficace aux clients qui opèrent sur plusieurs et. VPCs Comptes AWS Les clients accèdent à votre offre SaaS via VPC Lattice, qui fournit une connectivité réseau cohérente, des contrôles d'accès robustes et une gestion avancée du trafic. Dans ces scénarios, le trafic circule via VPC Lattice vers vos services d'application enregistrés. Il fournit des communications évolutives et sécurisées, quel que soit le service informatique que vous utilisez.

Appairage de VPC

Le peering VPC est une connexion réseau entre deux clouds privés virtuels (VPCs) qui achemine le trafic entre eux à l'aide d'adresses ou d' IPv4 adresses privées. IPv6 Le peering VPC est généralement utilisé entre des entités de confiance, comme celles d'une même organisation. Votre client crée une demande de peering pour l'un de vos VPCs clients. Lorsque vous l'acceptez, le trafic peut circuler entre les deux VPCs dans les deux sens. Cette approche de connexion se distingue par son caractère unique car elle implique une communication directe entre deux personnes VPCs sans aucun service intermédiaire ni infrastructure à gérer.

AWS Transit Gateway

AWS Transit Gatewayest un hub de transit réseau centralisé qui peut connecter des connexions à un réseau privé virtuel (VPN) VPCs, des AWS Direct Connect passerelles, des dispositifs virtuels tiers dans un VPC et d'autres passerelles de transit. Une passerelle de transit peut avoir une table de routage différente pour chaque pièce jointe. Cela fournit une flexibilité maximale pour le routage et vous aide à isoler les réseaux. Il est souvent utilisé pour connecter plusieurs personnes entre VPCs elles ou pour une inspection centralisée.

AWS Site-to-Site VPN

AWS Site-to-Site VPNpeut utiliser la technologie Internet Protocol Security (IPsec) pour établir des connexions entre les réseaux locaux, les bureaux distants, les usines, les autres fournisseurs de cloud et le réseau AWS mondial. La connexion est établie à partir d'une passerelle privée virtuelle ou d'une passerelle de transit dans un VPC AWS Cloud vers une passerelle client physique ou logicielle, qui peut se trouver dans le cloud, sur site ou dans le AWS Cloud cloud d'un autre CSP. La connexion peut se faire via Internet ou via une AWS Direct Connect connexion physique. Il est également possible d'avoir une connexion Site-to-Site VPN accélérée en utilisant AWS Global Accelerator. Une connexion accélérée achemine le trafic vers un emplacement AWS périphérique, tout en réduisant le temps de latence et en améliorant les performances.

AWS Direct Connect

AWS Direct Connectétablit une connexion privée haut débit entre un centre de données sur site et le AWS Cloud. En contournant l'Internet public, Direct Connect fournit une connexion à faible latence plus fiable, sécurisée et cohérente au AWS Cloud. Les clients se connectent à l'un Direct Connect des sites, puis choisissent une connexion hébergée ou dédiée à AWS. Bien qu'il s'agisse d'un choix

d'architecture peu courant pour les offres SaaS, il peut convenir aux fournisseurs de solutions SaaS qui comptent peu d'entreprises, mais de grandes entreprises.

Comparaison des capacités des services

Le tableau suivant décrit les fonctionnalités prises en Services AWS charge décrites dans ce guide. Les fonctionnalités incluses dans ce tableau sont décrites ci-dessous :

- Plages d'adresses CIDR qui se chevauchent : permet de connecter deux réseaux ou plus avec des plages d'adresses CIDR identiques ou se chevauchant
- Communication bidirectionnelle — Peut prendre en charge un canal de communication bidirectionnel afin que le consommateur de SaaS puisse exposer des ressources internes, telles qu'une base de données, au fournisseur de SaaS
- IPv6— Peut prendre IPv6 en charge une ou deux piles
- Trame jumbo — Peut prendre en charge les trames jumbo d'une taille d'image maximale de 8 500 octets
- Cloud hybride : peut prendre en charge une connexion avec un réseau sur site
- Multicloud — Peut prendre en charge une connexion entre les réseaux de différents fournisseurs de services cloud

Service ou approche	Plages CIDR qui se chevauchent	Communication bidirectionnelle	IPv6	cadre Jumbo	Cloud hybride	Multicloud
Appairage de VPC	Non	Oui	Oui	Oui [5]	Non	Non
AWS PrivateLink	Oui	Oui [1]	Oui	Oui	Numéro [6]	Numéro [6]
Amazon VPC Lattice	Oui	Oui [1]	Oui	Oui	Numéro [6]	Numéro [6]

AWS Transit Gateway	Non	Oui	Oui	Oui	Oui [3]	Oui [3]
AWS Site-to-Site VPN	Non	Oui	Oui	Non	Oui	Oui
AWS Direct Connect	Non	Oui	Oui	Oui [2]	Oui	Oui
Accès public à Internet [4]	Non applicable	Non	Oui	Oui	Oui	Oui

1. Avec des ressources VPC dans Amazon VPC Lattice
2. Uniquement pour les interfaces virtuelles privées et de transit
3. Avec Site-to-Site VPN ou AWS Direct Connect pièces jointes
4. En tant que terme général désignant les AWS ressources qui rendent une application accessible au public, telles qu'un Application Load Balancer
5. Uniquement pour les connexions de peering au sein d'une Région AWS
6. Possible grâce à une connexion de couche 3 préexistante entre les environnements

Caractéristiques et considérations relatives à la sécurité

Le tableau suivant décrit les fonctionnalités de Services AWS sécurité décrites dans ce guide.

- Moyens d'authentification — Comment vous assurer que seuls vos clients peuvent se connecter à votre service. Un autre niveau d'authentification pour les demandes entrantes est généralement toujours requis, en particulier dans les environnements mutualisés.
- Chiffrement en transit : indique si le chiffrement en transit est fourni par défaut. Le chiffrement natif décrit le chiffrement qui AWS couvre l'ensemble du trafic au sein VPCs VPCs, entre ou entre les centres de données. Le chiffrement supplémentaire décrit le chiffrement que vous contrôlez et qui peut être arrêté par le service concerné.

Service ou approche	Moyens d'authentification	Chiffrement en transit
Appairage de VPC	Vous lancez une demande de peering auprès du Compte AWS VPC de votre client ou vous acceptez une demande qu'il initie. Consultez Accepter ou rejeter une connexion d'appairage VPC.	Chiffrement natif uniquement
AWS PrivateLink	Vous choisissez ceux qui Comptes AWS sont autorisés à créer des points de terminaison pour votre service. Ces comptes sont appelés comptes principaux autorisés . Voir Accepter ou rejeter les demandes de connexion.	Chiffrement natif uniquement
Amazon VPC Lattice	Vous partagez un service ou un réseau de services VPC Lattice avec vos clients. Comptes AWS Voir Partager vos entités VPC Lattice.	Chiffrement natif et chiffrement TLS supplémentaire
AWS Transit Gateway	Votre client crée une demande de pièce jointe de peering à partir de lui Compte AWS, ou vous êtes à l'origine de la demande. Consultez les pièces jointes de peering de Transit Gateway dans Amazon VPC Transit Gateways.	Chiffrement natif et IPsec chiffrement supplémentaire avec une pièce jointe VPN
AWS Site-to-Site VPN	Vous utilisez des clés IPsec pré-partagées ou un certificat privé sur l'appareil du client. Voir les options d'authent	IPsec Chiffrement supplémentaire

	ification du AWS Site-to-Site VPN tunnel.	
AWS Direct Connect	Votre client crée une demande d'interface virtuelle à partir de son Compte AWS. Voir interfaces Direct Connect virtuelles et interfaces virtuelles hébergées.	Chiffrement supplémentaire de couche 2 possible sur certains sites. Voir Direct Connect Emplacements.
Accès public à Internet [1]	Une authentification personnalisée est requise.	Chiffrement TLS supplémentaire possible

1. En tant que terme général désignant les AWS ressources qui rendent une application accessible au public, telles qu'un Application Load Balancer

Évaluation des options d'accès au réseau pour les offres SaaS

Les indicateurs importants pour votre organisation dépendront de l'identité de vos clients, de votre stratégie commerciale et de vos objectifs organisationnels. Ce guide présente des indicateurs que vous pouvez utiliser pour choisir une approche d'accès réseau, mais vous devez donner la priorité à ceux qui répondent aux exigences uniques de votre cas d'utilisation.

Cette section contient les rubriques suivantes :

- Métriques d'évaluation
- Coût total de propriété
- Carte des valeurs du réseau

Métriques d'évaluation

Certains indicateurs sont cohérents entre les organisations et les cas d'utilisation, et nous pouvons vous aider à les évaluer. Les indicateurs suivants sont les suivants :

- Facilité d'intégration — Avec quelle rapidité et quelle facilité pouvez-vous intégrer de nouveaux clients ?
- Coût total de possession (TCO) — Quelle est la structure des coûts ? Au-delà des coûts d'infrastructure fixes et variables, il existe d'importantes considérations liées aux coûts supplémentaires liés aux frais d'exploitation, à la dépendance à l'égard des experts, au coût de mise en œuvre des modifications et à la conformité. Pour plus d’informations, consultez la section Coût total de propriété.
- Évolutivité — Votre approche de l'accès au réseau est-elle capable d'évoluer afin de soutenir la croissance de votre entreprise ? Le développement de votre clientèle implique des considérations architecturales et organisationnelles importantes. Réfléchissez à la manière dont vous pourriez vous adapter pour accueillir 5 à 100 fois plus de clients que ce que vous soutenez aujourd'hui.
- Adaptabilité — Pouvez-vous facilement mettre en œuvre des changements ? Les modifications peuvent inclure une nouvelle application, une nouvelle fonctionnalité, une plate-forme différente ou un réseau différent.
- Isolation du réseau : quelle part de l'infrastructure réseau exposez-vous à vos clients ? Fournissez-vous le bon degré d'accès ou exposez-vous à des réseaux entiers ? Si vous isolez les ressources

du réseau à un stade précoce, il sera plus facile de fournir des garanties de sécurité, de confidentialité et de conformité ultérieurement.

- Observabilité — Quelle est votre capacité à détecter les défaillances ou les dégradations des services ? Est-il facile et rapide d'identifier le problème ? Avec quelle rapidité (et avec quels frais généraux) pouvez-vous aider vos clients à comprendre leurs points de défaillance et à les résoudre ?
- Délai de réparation : quel est le délai entre la détection d'une panne ou d'une dégradation du service et la reprise des opérations ? Quels sont les facteurs qui influent sur cette capacité ?

D'autres indicateurs sont propres à votre organisation ou à votre offre car ils concernent vos opérations, votre stratégie ou vos objectifs commerciaux. Vous êtes le seul à pouvoir évaluer ces indicateurs. Les indicateurs suivants sont les suivants :

- Harmonisation du modèle commercial — Quel est votre modèle commercial et dans quelle mesure les approches d'accès individuel s'y alignent-elles ?
- Marché adressable total (TAM) — Quel est votre marché actuel et futur, et dans quelle mesure est-il couvert par l'approche d'accès au réseau ?
- Retour sur investissement (ROI) — Quelles améliorations attendez-vous en termes de rentabilité et de marges ? Les avantages financiers attendus sont-ils suffisants pour répondre à vos besoins en matière d'accès souple et adaptable aux services ?
- Conformité réglementaire — Quels types d'exigences réglementaires s'appliquent et sur quel marché ?
- Contrats de niveau de service (SLAs) — Les clients ont-ils besoin que votre offre SaaS soit hautement disponible ? Quels types d'engagements êtes-vous contractuellement tenus de respecter ?

Coût total de propriété

Cette section explore le coût total de possession (TCO), qui est l'une des mesures d'évaluation utilisées pour comparer les approches d'accès au réseau. Le TCO est un indicateur composite comprenant les coûts d'infrastructure fixes et variables, les frais d'exploitation, la dépendance à des spécialistes, le coût du changement et les coûts de conformité.

Le coût total de possession pour chaque approche d'accès au réseau peut varier en fonction de votre cas d'utilisation. Par exemple, le coût du changement pour un fournisseur de SaaS proposant

un simple service Web et cinq locataires est différent de celui d'un fournisseur de services SaaS doté d'un portefeuille de produits complexe et interconnecté comptant des centaines ou des milliers de locataires. De plus, tous les composants n'ont pas le même poids. Par exemple, l'embauche d'un spécialiste des réseaux est souvent plus coûteuse que les coûts d'infrastructure nécessaires au déploiement individuel de votre service. Utilisez les valeurs du tableau suivant pour l'orientation initiale et comme point de référence pour une discussion plus approfondie.

Approche d'accès	Coûts d'infrastructure fixes	Coûts d'infrastructure variables	Frais généraux d'exploitation	Dépendance spécialisée	Coût du changement	Coûts de conformité
Appairage de VPC	Aucune	Aucune	Élevé	Faible	Élevé	Moyen
AWS PrivateLink	Faible	Faible	Faible	Aucune	Faible	Faible
Amazon VPC Lattice	Moyen	Moyen	Faible	Faible	Faible	Faible
AWS Transit Gateway	Moyen	Moyen	Faible	Faible	Faible	Moyen
AWS Site-to-Site VPN	Medium	Élevé	Élevé	Moyen	Moyen	Faible
AWS Direct Connect	Élevé	Moyen	Medium	Élevé	Élevé	Faible
Accès public à Internet	Faible	Élevé	Moyen	Faible	Faible	Élevé

Coûts de peering VPC

Aucun coût d'infrastructure direct n'est associé à une connexion d'appairage VPC. Lorsque le trafic reste dans la même zone de disponibilité, aucun frais de transfert de données n'est facturé. Cependant, les frais d'exploitation peuvent être importants car la gestion et la complexité augmentent de façon exponentielle à chaque connexion d'appairage supplémentaire. Des connaissances de base du réseau sont suffisantes pour configurer une connexion d'appairage, mais les modifications du réseau sont difficiles à mettre en œuvre avec plus d'une poignée de connexions d'appairage. Les coûts de mise en conformité sont légèrement plus élevés car les deux parties exposent mutuellement un VPC complet, plutôt que des services individuels.

AWS PrivateLink coûts

AWS PrivateLink est souvent une solution rentable avec de faibles frais d'exploitation. En effet, le fournisseur de SaaS doit gérer uniquement un Network Load Balancer et le consommateur doit gérer uniquement les points de terminaison VPC. Vous pouvez apporter des modifications des deux côtés de manière transparente, ce qui réduit la collaboration interorganisationnelle coûteuse et gourmande en ressources. Les coûts de mise en conformité ont tendance à être faibles car le fournisseur de SaaS n'expose que les services qu'il souhaite et non l'ensemble du réseau.

Coûts d'Amazon VPC Lattice

Amazon VPC Lattice propose une structure de coûts équilibrée avec des coûts d'infrastructure fixes et variables modérés. En tant que réseau de services entièrement géré, il réduit considérablement les frais opérationnels en automatisant la découverte des services, la gestion du trafic et les contrôles d'accès sur plusieurs VPCs réseaux. Cela simplifie à la fois le déploiement initial et la gestion continue par rapport aux configurations réseau manuelles. Vous pouvez mettre en œuvre des modifications par le biais de contrôles basés sur des politiques sans mettre à jour de routage complexes, ce qui réduit la dépendance à l'égard des spécialistes réseau. Les coûts de mise en conformité ont tendance à être inférieurs à ceux des approches réseau traditionnelles, car VPC Lattice fournit des contrôles d'accès précis et une visibilité complète grâce à des fonctionnalités de surveillance et de journalisation intégrées. Cela peut faciliter la démonstration de la conformité réglementaire.

AWS Transit Gateway coûts

AWS Transit Gateway a des frais horaires et de traitement des données plus AWS PrivateLinkélevés que, mais ses frais opérationnels sont similaires. Vous devez avoir une connaissance approfondie du

AWS Transit Gateway service et du routage AWS afin de configurer correctement toutes les tables de routage. Les modifications de l'infrastructure peuvent nécessiter des mises à jour du routage ou du DNS. Les coûts de mise en conformité sont similaires à ceux du peering VPC, car les deux parties peuvent potentiellement exposer des sous-réseaux ou des réseaux complets l' VPCs un à l'autre. AWS Transit Gateway les tables de routage doivent également être manipulées avec précaution, car elles sont partagées par plusieurs consommateurs et vous ne devez autoriser aucun trafic entre eux.

AWS Site-to-Site VPN coûts

Comme le Site-to-Site VPN envoie essentiellement du trafic vers Internet, le coût variable est le plus élevé par rapport aux frais de transfert de données. Bien qu'il s'agisse d'un service de réseau privé virtuel (VPN) géré, il entraîne des frais opérationnels importants, en particulier au niveau de la passerelle client. Le provisionnement et les opérations nécessitent des connaissances avancées en matière de réseau, et les modifications nécessitent souvent l'intervention des deux parties. Les coûts de mise en conformité sont généralement faibles car les équipes de sécurité préapprouvent souvent IPsec les tunnels sans examen supplémentaire.

AWS Direct Connect coûts

AWS Direct Connect entraîne le coût d'infrastructure fixe le plus élevé car il s'agit d'une connexion physique privée directement au AWS Cloud. Des connaissances spécialisées sont nécessaires pour configurer et exploiter une session BGP (Border Gateway Protocol) (si nécessaire), pour exploiter une connexion VPN et pour effectuer l'ingénierie du trafic. Ce service réduit les efforts des équipes de sécurité car il associe une connectivité privée à la possibilité de bénéficier en plus du contrôle d'accès aux médias, de la sécurité (MACsec) et du IPsec chiffrement.

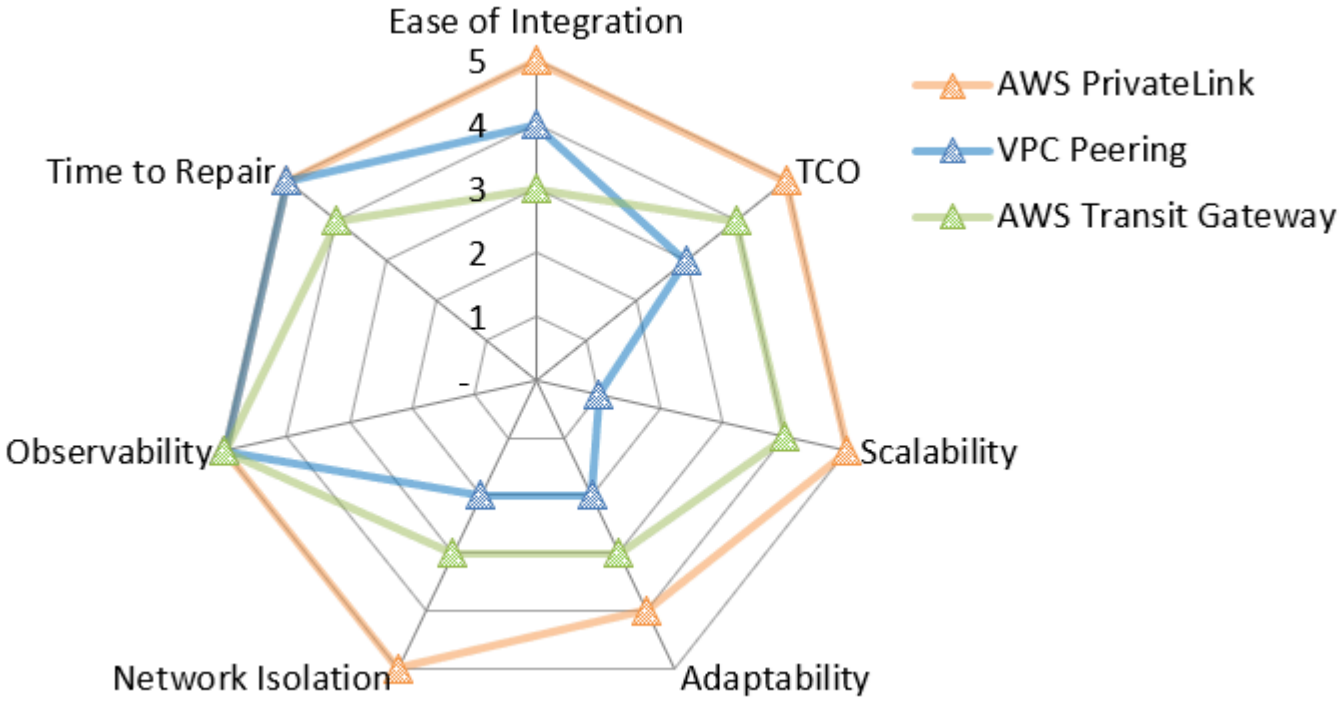
Coûts d'accès public à Internet

L'accès public à Internet fait référence aux AWS ressources que vous pouvez utiliser pour rendre une application accessible au public, telles qu'un Application Load Balancer. Cette approche implique des coûts variables liés à l'accès à vos services, notamment les frais de transfert de données vers Internet. Les frais d'exploitation et de conformité peuvent être importants car vous exposez le service à Internet et aurez besoin de mécanismes de sécurité et d'authentification supplémentaires. Cependant, aucun routage complexe n'est impliqué, et aucune des parties n'a besoin de connaître les détails de l'infrastructure de l'autre.

Carte des valeurs du réseau

Pour vous aider à avoir une vue d'ensemble et à prendre des décisions éclairées, ce guide inclut une carte des valeurs du réseau pour chaque scénario. Étant donné que les notes varient d'un scénario à l'autre, le même service peut obtenir des notes différentes pour deux scénarios. Les cartes de valeurs sont des cartes radar, où un score parfait hypothétique serait de cinq dans toutes les catégories.

Par exemple, l'image suivante montre un exemple de carte radar. Il inclut uniquement les indicateurs que nous pouvons aider à évaluer. Nous vous recommandons de créer votre propre carte de valeurs qui inclut les mesures supplémentaires que vous seul pouvez évaluer.



Scénarios d'accès au réseau pour les offres SaaS dans le AWS Cloud

Cette section couvre les différentes options d'accès au réseau pour vos offres SaaS dans le AWS Cloud. Il aborde les approches du point de vue de votre consommateur, qui peut avoir des besoins de connectivité au sein de AWS Cloud, à partir de centres de données sur site ou auprès d'autres fournisseurs de services cloud (CSPs). En outre, vous devrez peut-être prendre en charge l'accès depuis plusieurs types d'environnements grand public.

Comprendre les exigences de connectivité réseau dans ces divers environnements est essentiel pour créer une stratégie d'accès complète. Vos décisions architecturales doivent tenir compte de la diversité des modèles de sécurité, des attentes en matière de performances et des contraintes techniques, tout en préservant l'efficacité opérationnelle. La bonne approche fournit une connectivité sécurisée et fiable qui s'adapte à la croissance de votre entreprise et minimise à la fois la complexité de la mise en œuvre et les frais de gestion permanents.

Lorsque vous évaluez les options d'accès au réseau, considérez l'impact de chaque approche sur votre coût total de possession, y compris non seulement les coûts d'infrastructure, mais également les frais opérationnels et les exigences de conformité. Certaines approches excellent en termes d'évolutivité mais peuvent introduire de la complexité, tandis que d'autres privilégient la facilité d'intégration au détriment de l'isolation du réseau. Les capacités techniques et les ressources de vos clients jouent également un rôle important dans la détermination de la solution la plus appropriée.

Pour les consommateurs AWS Cloud, des services tels que ceux-ci AWS PrivateLink offrent des avantages significatifs en termes de sécurité et d'évolutivité. Les consommateurs sur site peuvent bénéficier de AWS Direct Connect performances constantes ou d'un Site-to-Site VPN pour une connectivité rentable. Les scénarios multicloud nécessitent souvent un examen attentif des problèmes d'interopérabilité, et vous pouvez utiliser des architectures VPC de transit pour standardiser les modèles d'accès. Dans tous les cas, votre conception doit anticiper la croissance future des consommateurs et du trafic afin que votre architecture réseau reste résiliente et adaptable au fur et à mesure de l'évolution de votre offre SaaS.

Cette section contient les scénarios suivants :

- Consommateurs de solutions SaaS opérant sur AWS
- Consommateurs de services opérant sur place
- Consommateurs de solutions SaaS faisant appel à d'autres fournisseurs de services cloud

- Prise en charge des environnements hybrides

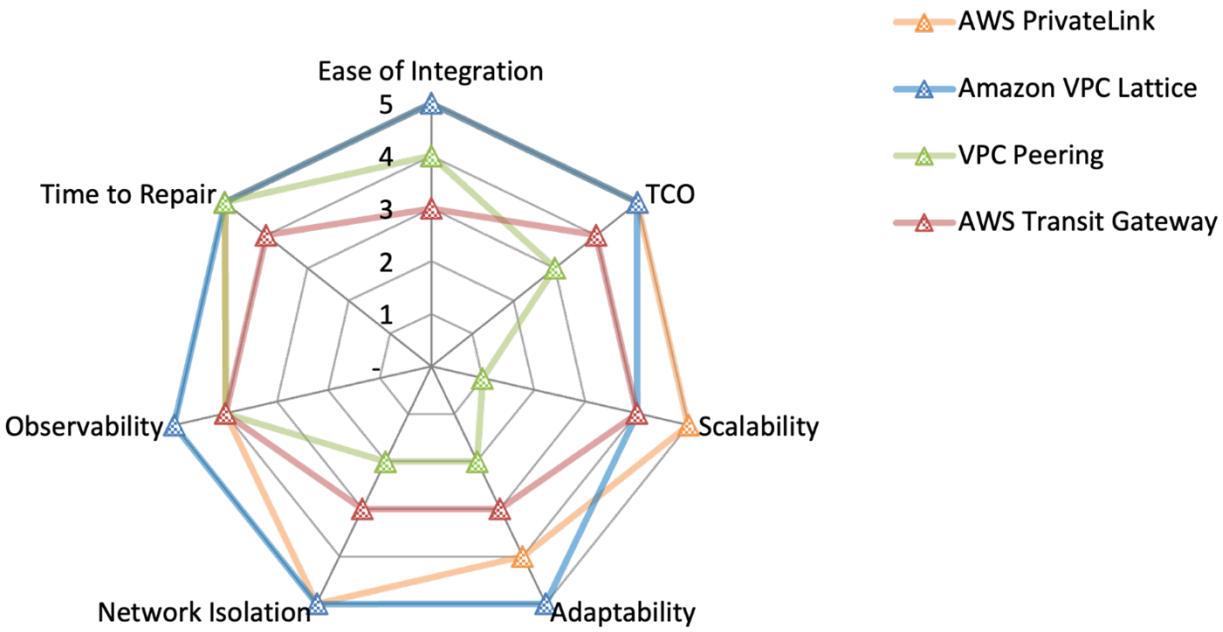
Consommateurs de solutions SaaS opérant sur AWS

Cette section décrit les options de connectivité si vous et vos clients opérez dans le AWS Cloud. Ce scénario offre la plus grande flexibilité, car bon nombre d'entre eux s'intègrent de Services AWS manière native et parce que les deux parties ont accès à l'ensemble du Service AWS portefeuille.

Cette section décrit les approches d'accès réseau suivantes :

- Intégration avec AWS PrivateLink
- Partage d'un service Amazon VPC Lattice
- Création de connexions d'appairage VPC
- Connexion VPCs avec AWS Transit Gateway

La carte des valeurs réseau suivante résume le score de chacune de ces options pour chaque métrique d'évaluation. Pour plus d'informations sur les mesures d'évaluation, consultez la section Mesures d'évaluation dans ce guide. Sur la carte, cinq représente le meilleur score, tel que le coût total de possession le plus faible, la meilleure isolation du réseau ou le délai de réparation le plus court. Pour plus d'informations sur la lecture de cette carte radar, consultez Carte des valeurs du réseau ce guide.



La carte radar montre les valeurs suivantes.

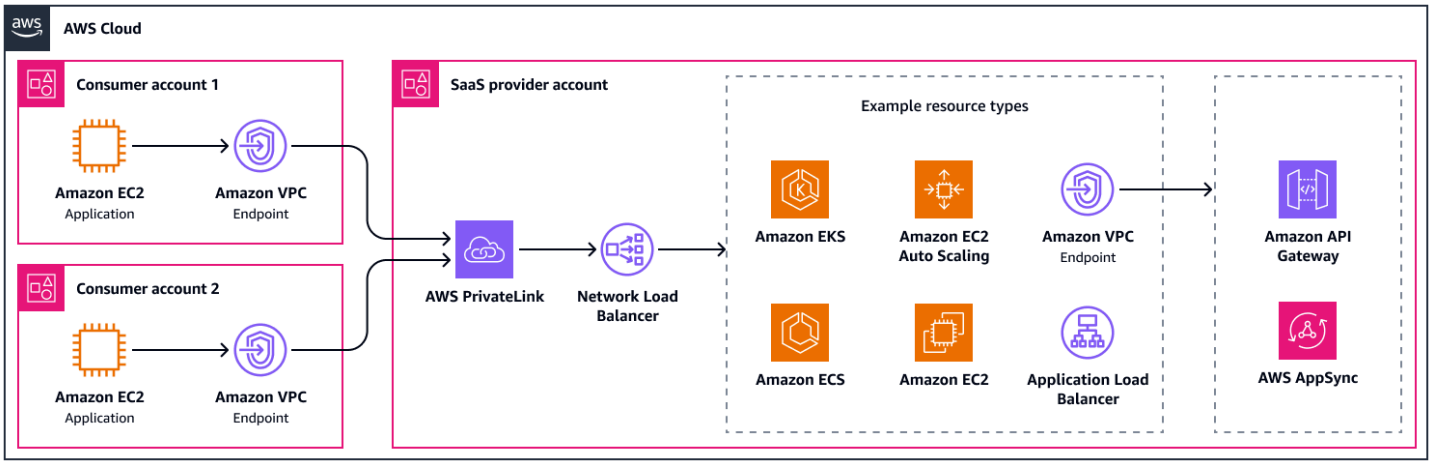
Métrique d'évaluation	AWS PrivateLink	Amazon VPC Lattice	Appairage de VPC	AWS Transit Gateway
Facilité d'intégration	5	5	4	3
TCO	5	5	3	4
Evolutivité	5	4	1	4
Adaptabilité	4	5	2	3
Isolation du réseau	5	5	2	3
Observabilité	4	5	4	4
Il est temps de réparer	5	5	5	4

Intégration avec AWS PrivateLink

AWS PrivateLinkest le moyen le plus natif du cloud pour intégrer une offre SaaS. Les fournisseurs de SaaS peuvent héberger leur application soit derrière un Network Load Balancer. Le Network Load Balancer s'intègre directement à un Application Load Balancer, à Amazon Elastic Container Service (Amazon ECS), à Amazon Elastic Kubernetes Service(Amazon EKS) et à des groupes Auto Scaling. Il est également possible d'acheminer le trafic depuis le Network Load Balancer vers les points de terminaison VPC de l'interface dans le compte du fournisseur SaaS. Cela vous permet d'utiliser une API pour accéder à des applications, par exemple via Amazon API Gateway ou AWS AppSync. Si votre application nécessite l'accès à des ressources de l'environnement client qui ne sont pas équilibrées en termes de charge, telles qu'une base de données, vous pouvez utiliser des points de terminaison VPC de ressources.

AWS PrivateLink prend en charge une bande passante allant jusqu'à 100 Gbit/s par zone de disponibilité. Le schéma suivant montre une configuration de base avec quelques intégrations possibles. Il connecte deux comptes consommateurs au compte du fournisseur de SaaS via AWS

PrivateLink. Il existe des points de terminaison de service dans les comptes des consommateurs et un Network Load Balancer dans le compte du fournisseur de services SaaS.



Les avantages de cette approche sont les suivants :

- Facilité d'intégration : aucune modification de la table de routage n'est requise
- Facilité d'intégration : vous pouvez proposer des services de point de terminaison via AWS Marketplace
- Facilité d'intégration : les points de terminaison VPC prennent en charge les noms DNS conviviaux
- Évolutivité : elle peut s'adapter à des milliers de consommateurs de solutions SaaS
- Adaptabilité : Support pour les plages CIDR qui se chevauchent
- Adaptabilité : Support pour IPv6
- Adaptabilité : soutien interrégional
- TCO : AWS PrivateLink est un service entièrement géré, il nécessite donc moins d'efforts opérationnels
- Isolation du réseau : avantage en termes de sécurité pour le consommateur du SaaS, car le trafic ne peut pas être initié par le fournisseur de SaaS
- Isolation du réseau : avantage en termes de sécurité pour le fournisseur de SaaS, car il n'expose pas l'intégralité d'un sous-réseau ou d'un VPC

Les inconvénients de cette approche sont les suivants :

- Adaptabilité : le fournisseur de SaaS doit utiliser les mêmes zones de disponibilité que le consommateur

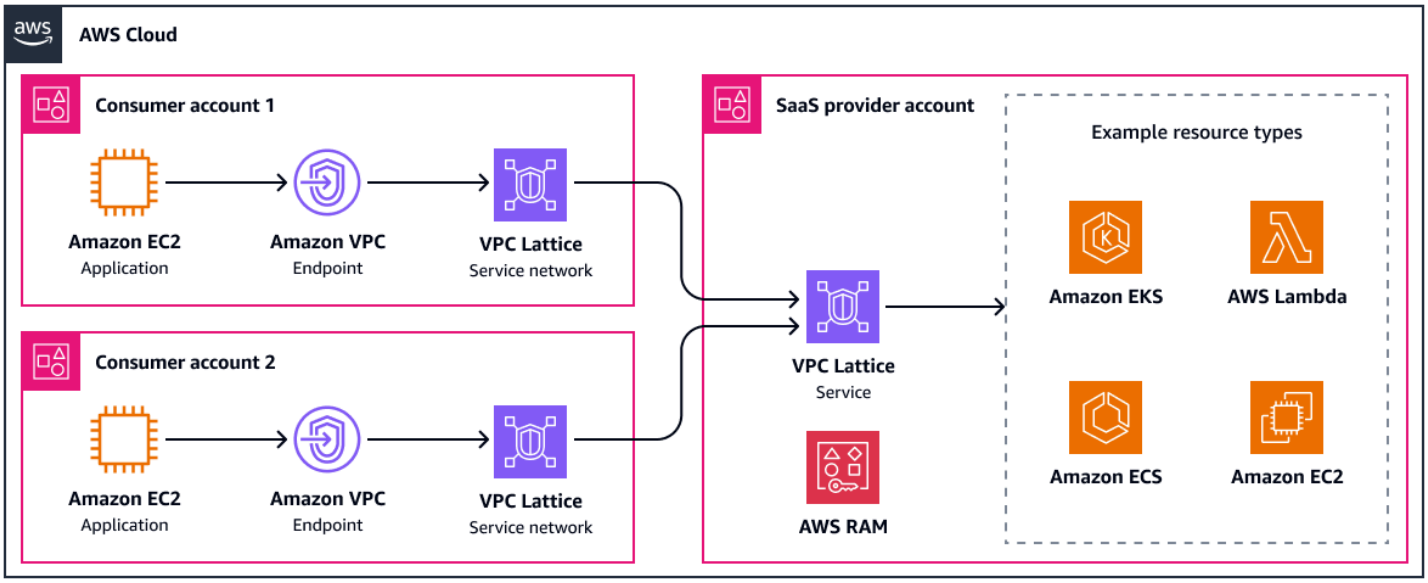
- Adaptabilité : Support uniquement pour les connexions initiées par le client, et des points de terminaison VPC de ressources sont nécessaires pour les communications initiées par le service
- Adaptabilité : Network Load Balancer est la seule intégration directe pour AWS PrivateLink

Partage d'un service Amazon VPC Lattice

Pour utiliser Amazon VPC Lattice comme option de connectivité pour votre application SaaS, vous devez d'abord créer un ou plusieurs services VPC Lattice qui représentent les composants de votre application SaaS. Vous configurez les écouteurs et les règles de routage pour diriger le trafic vers vos cibles principales, telles que les instances, les conteneurs ou les fonctions Amazon EC2. AWS Lambda Pour plus d'informations, consultez Connecter des services Saas au sein d'un réseau de services VPC Lattice (AWS article de blog). D'un point de vue conceptuel, cela revient presque à configurer un Application Load Balancer. Ensuite, vous partagez votre service SaaS en toute sécurité avec le client Comptes AWS ou les organisations en utilisant AWS Resource Access Manager (AWS RAM), en spécifiant les autorisations dont ils disposent. Une fois que les clients ont accepté le partage des ressources, ils peuvent associer votre service SaaS à leurs réseaux de services VPC Lattice existants ou nouvellement créés pour permettre la communication. service-to-service

Chaque service VPC Lattice peut prendre en charge jusqu'à 10 Gbit/s et 10 000 demandes par seconde par zone de disponibilité. En mettant en œuvre des politiques d'authentification, vos clients peuvent avoir un contrôle précis sur les services et les ressources qui peuvent accéder à l'application SaaS. Vous pouvez utiliser des passerelles de ressources pour accéder aux ressources qui nécessitent une connexion TCP. Par exemple, il peut s'agir d'un cluster Amazon EKS que vous gérez ou d'une ressource gérée par le client à laquelle votre application doit accéder. Pour plus d'informations sur l'utilisation de passerelles de ressources pour les offres SaaS, consultez Étendre les fonctionnalités SaaS à Comptes AWS l'aide de la AWS PrivateLink prise en charge des ressources VPCAWS (article de blog).

Le schéma suivant montre une configuration VPC Lattice de haut niveau avec quelques exemples d'intégrations. Il utilise des réseaux de services gérés par le client pour accéder à l'application SaaS.



Les avantages de cette approche sont les suivants :

- Facilité d'intégration : aucune modification de la table de routage n'est requise
- Facilité d'intégration : découverte de services prête à l'emploi
- Évolutivité : elle peut s'adapter à des milliers de consommateurs de solutions SaaS
- Adaptabilité : Support pour les plages CIDR qui se chevauchent
- Adaptabilité : Support pour IPv6
- Adaptabilité : s'intègre à n'importe quel service de AWS calcul en tant que service VPC Lattice
- TCO : VPC Lattice est un service entièrement géré, il nécessite donc moins d'efforts opérationnels
- TCO : équilibrage de charge intégré avec routage avancé du trafic
- Isolation du réseau : autorisation précise avec politiques d'authentification
- Isolation du réseau : avantage en termes de sécurité pour le consommateur du SaaS, car le trafic ne peut pas être initié par le fournisseur SaaS
- Isolation du réseau : avantage en termes de sécurité pour le fournisseur de SaaS, car vous n'exposez pas l'intégralité d'un sous-réseau ou d'un VPC

Les inconvénients de cette approche sont les suivants :

- Adaptabilité : Support uniquement pour les connexions initiées par le client, et des passerelles de ressources sont nécessaires pour les communications initiées par le service
- Adaptabilité : aucun soutien interrégional

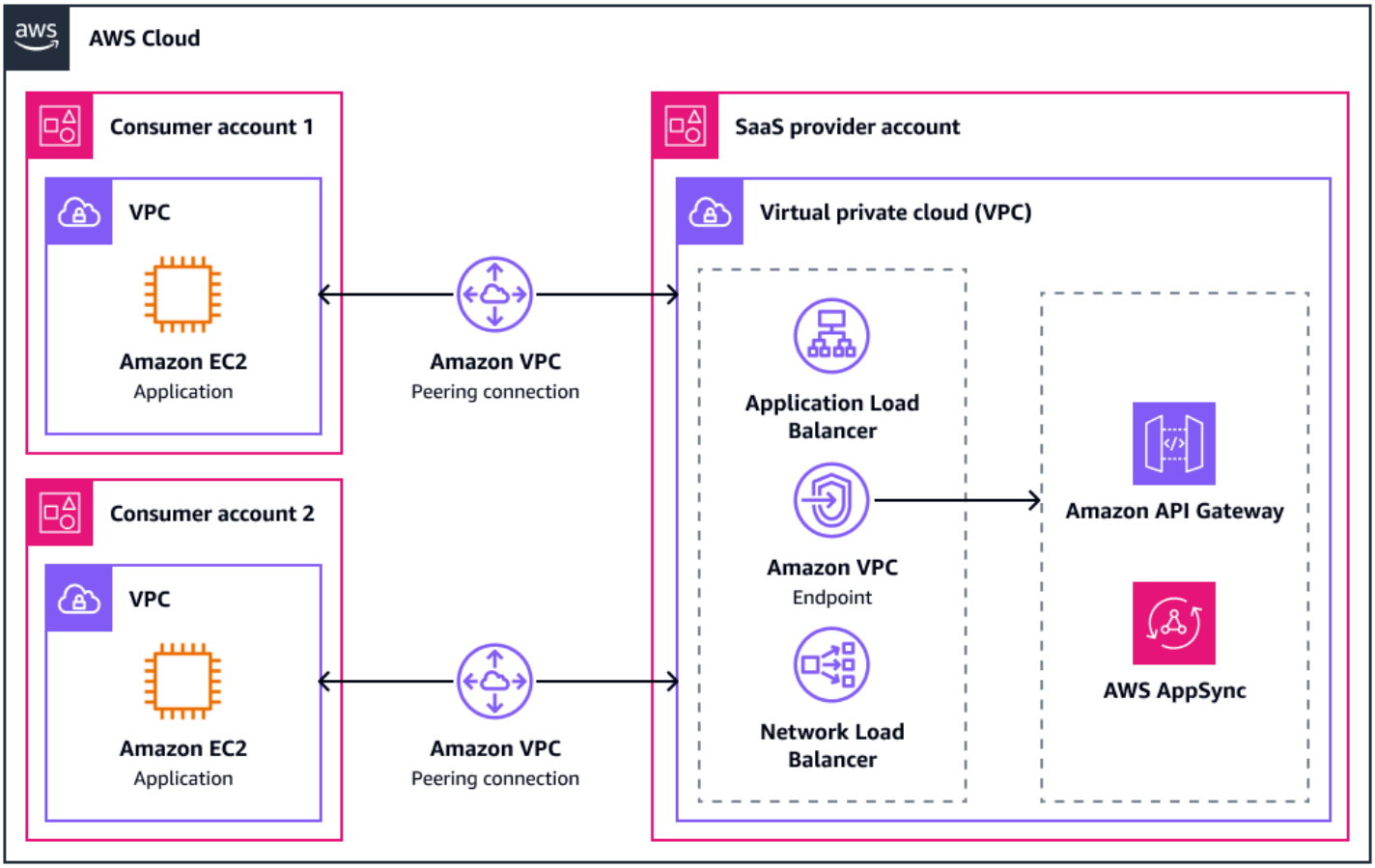
Création de connexions d'appairage VPC

Lorsque vous utilisez le peering VPC pour connecter le VPC du fournisseur de SaaS au VPC du consommateur, les deux parties peuvent établir des connexions. Cela nécessite une configuration appropriée des groupes de sécurité, des pare-feux et des listes de contrôle d'accès réseau (NACLs) dans les deux comptes. Dans le cas contraire, du trafic indésirable risque d'entrer sur le réseau via la connexion d'appairage. Vous pouvez utiliser des groupes de sécurité pour référencer des groupes de sécurité issus de peered VPCs. Cela peut vous aider à contrôler l'accès à votre application, car les groupes de sécurité autorisés fournissent un contrôle d'accès plus explicite et plus précis que les listes d'adresses IP autorisées.

Avec le peering VPC, l'offre SaaS est accessible via un service ou une ressource déployés dans le VPC. La plupart des applications SaaS reposent sur un Application Load Balancer ou un Network Load Balancer. AWS AppSync private APIs ou Amazon API Gateway private APIs sont d'autres points d'entrée courants pour les applications SaaS, car ils peuvent être la cible d'une connexion d'appairage via des points de terminaison VPC d'interface.

Après avoir établi une connexion d'appairage, vous devez mettre à jour les tables de routage pour les VPCs deux comptes afin de définir la connexion d'appairage comme le prochain saut pour la plage d'adresses CIDR correspondante. Cette solution est recommandée uniquement aux fournisseurs de SaaS qui ont peu de clients, car la gestion de plusieurs connexions de peering devient rapidement trop complexe.

Le schéma suivant montre une configuration de base avec quelques intégrations possibles. VPCs dans deux comptes consommateurs, disposez d'une connexion de peering avec un VPC sur le compte du fournisseur SaaS.



Les avantages de cette approche sont les suivants :

- Temps de réparation : aucun point de défaillance unique en matière de communication
- Évolutivité : aucune limite de bande passante grâce au peering VPC
- TCO : aucun coût pour la connexion d'appairage ou le trafic via la connexion d'appairage au sein de la même zone de disponibilité
- TCO : aucune infrastructure à gérer
- Adaptabilité : Support pour IPv6
- Adaptabilité : soutien au peering interrégional

Les inconvénients de cette approche sont les suivants :

- Adaptabilité : aucun support pour le routage transitif
- Adaptabilité : aucun support pour les plages CIDR qui se chevauchent
- Extensibilité : évolutivité limitée (maximum de 125 connexions d'appairage par VPC)

- TCO : la complexité augmente de façon exponentielle à chaque connexion d'appairage supplémentaire
- TCO : surcharge liée à la gestion des tables de routage, au peering des connexions elles-mêmes, aux règles des groupes de sécurité et à l'inspection du trafic
- Isolation du réseau : des contrôles de sécurité stricts sont nécessaires car VPCs l'ensemble des deux parties est exposé

Connexion VPCs avec AWS Transit Gateway

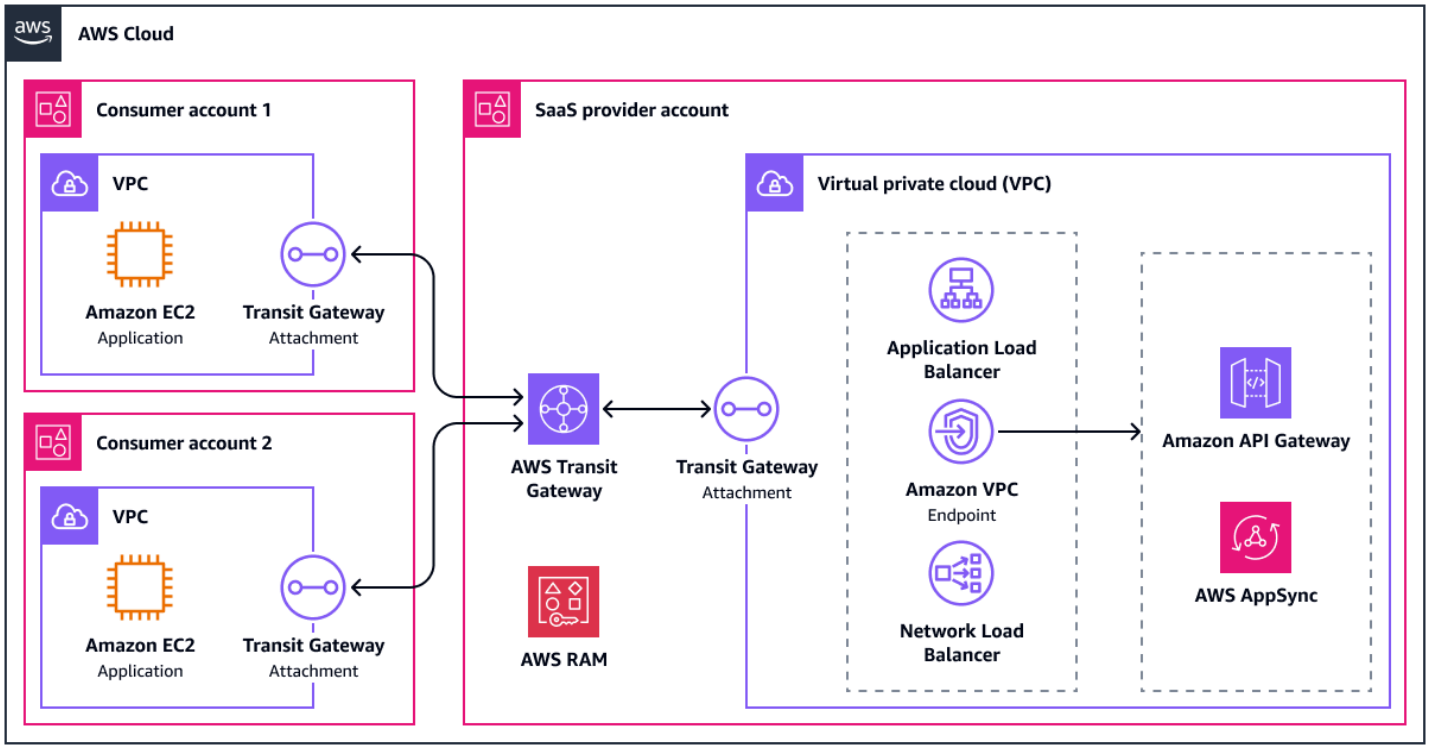
Lorsque vous vous connectez AWS Transit Gateway, il crée des pièces jointes VPCs au VPC et déploie des interfaces réseau dans les sous-réseaux de chaque zone de disponibilité qui doivent acheminer le trafic vers et depuis le VPC. Il est recommandé de disposer d'un `/28` sous-réseau dédié dans chaque zone de disponibilité pour l'attachement VPC. Pour plus d'informations, consultez les meilleures pratiques de conception d'Amazon VPC Transit Gateway. Une table de routage mise à jour est VPCs nécessaire pour envoyer le trafic via l'interface réseau déployée, et les tables de routage de Transit Gateway doivent être mises à jour en conséquence. Dans une configuration multi-locataires, vous souhaitez que le VPC du fournisseur de SaaS dispose d'un itinéraire vers tous les consommateurs. VPCs Le consommateur VPCs doit avoir un itinéraire uniquement vers le VPC du fournisseur de SaaS.

Transit Gateway est conçu pour être hautement disponible. Il prend en charge la surveillance à l'aide des journaux de flux VPC, et la bande passante maximale pour une connexion Transit Gateway est de 100 Gbit/s par zone de disponibilité. Tout comme le peering VPC, cette approche permet le référencement des groupes de sécurité inter-VPC, ce qui simplifie le contrôle d'accès entre les environnements.

Il existe deux options principales pour connecter les consommateurs à votre offre SaaS avec Transit Gateway.

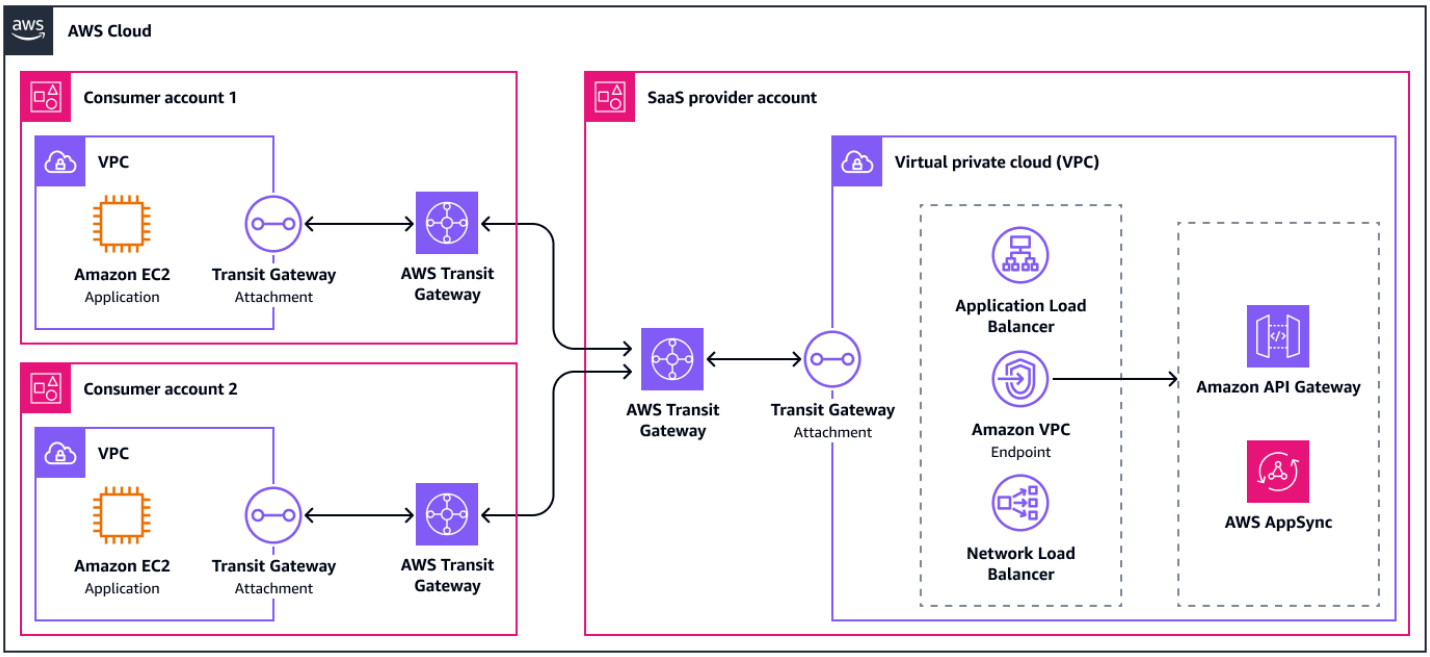
Option 1 : utilisation de la RAM

Dans la première option, le fournisseur de services partage le Transit Gateway avec les consommateurs en utilisant AWS Resource Access Manager (AWS RAM). Cela permet aux consommateurs de déployer les pièces jointes VPC dans leurs propres comptes. Le schéma suivant montre cette option à un niveau élevé.



Option 2 : passerelles de transport en commun jumelées

La deuxième option consiste à associer votre passerelle de transit à une passerelle de transit dans les comptes des consommateurs. Cela offre aux consommateurs une plus grande flexibilité, car ils peuvent désormais contrôler entièrement les tables de routage au sein de leur passerelle de transit. Par exemple, ils pourraient configurer une inspection centralisée entre le service et leurs charges de travail. L'inconvénient de cette option est que seul le routage statique entre les passerelles de transit est pris en charge. Le schéma suivant montre cette option à un niveau élevé.



Les avantages de cette approche sont les suivants :

- Évolutivité : Support pour un maximum de 5 000 pièces jointes
- Évolutivité : un seul endroit pour gérer et surveiller tous les appareils connectés VPCs
- Adaptabilité : Transit Gateway peut également se connecter à des appareils SD-WAN tiers VPNs, à des Direct Connect passerelles et à des appareils SD-WAN tiers
- Adaptabilité : architecture flexible, telle que l'ajout d'un VPC d'inspection
- Adaptabilité : Support pour le routage transitif
- Adaptabilité : Peut comparer les passerelles de transit intra-régionales et interrégionales
- Adaptabilité : Support pour IPv6
- TCO : AWS Transit Gateway est un service entièrement géré, il nécessite donc moins d'efforts opérationnels
- Coût total de possession : le coût total de possession augmente de façon linéaire à chaque connexion à une passerelle de transit supplémentaire

Les inconvénients de cette approche sont les suivants :

- Facilité d'intégration : la configuration du routage nécessite des connaissances réseau avancées
- Adaptabilité : aucun support pour les plages CIDR qui se chevauchent

- TCO : surcharge liée à la gestion des entrées des tables de routage, aux règles des groupes de sécurité et à l'inspection du trafic
- Sécurité : des contrôles de sécurité stricts sont nécessaires car VPCs l'ensemble des deux parties est exposé

Consommateurs de services opérant sur place

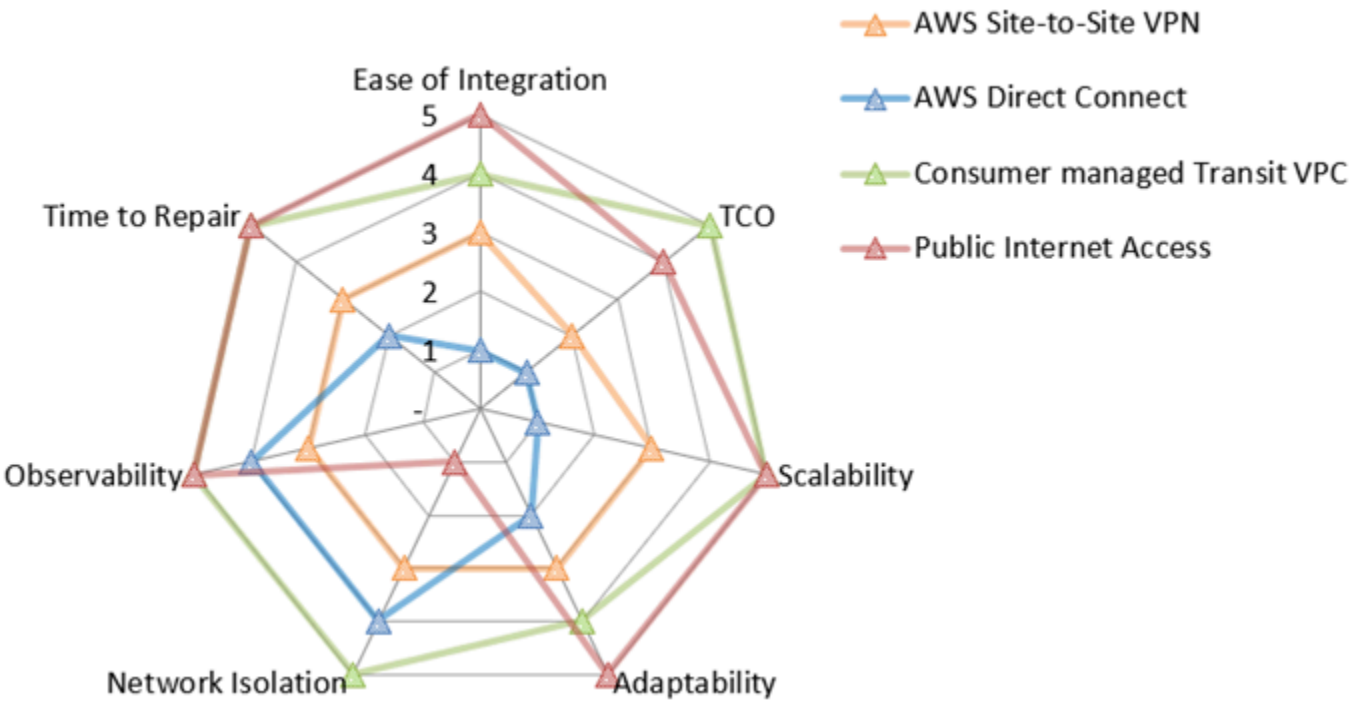
Cette section décrit les options de connectivité entre les charges de travail SaaS dans les AWS Cloud centres de données et sur site. De nombreux consommateurs ayant des exigences sur site, en particulier au niveau de l'entreprise, considèrent le cloud comme une extension de leur réseau physique, et ils souhaitent en tenir compte dans leur architecture. Cela signifie une connectivité privée à l'offre SaaS dans le cloud, soit via des tunnels logiques, soit même via une connexion physique privée. Les autres consommateurs accepteront la connectivité via l'Internet public, qui est également abordée dans cette section.

Cette section décrit les approches d'accès réseau suivantes :

- Connexion avec AWS Site-to-Site VPN
- Connexion avec AWS Direct Connect
- Connexion à une architecture VPC de transit
- Connexion via l'Internet public

La carte des valeurs réseau suivante résume le score de chacune de ces options pour chaque métrique d'évaluation. Pour plus d'informations sur les mesures d'évaluation, consultez la section Mesures d'évaluation dans ce guide. Sur la carte, cinq représente le meilleur score, tel que le coût total de possession le plus faible, la meilleure isolation du réseau ou le délai de réparation le plus court. Pour plus d'informations sur la lecture de cette carte radar, consultez Carte des valeurs du réseau ce guide.

> **Note**
>
> L'option VPC de transport géré par le fournisseur est exclue car les scores dépendent fortement des services exploités.



La carte radar montre les valeurs suivantes.

Métrique d'évaluation	AWS Site-to-Site VPN	AWS Direct Connect	VPC de transit géré par le consommateur	Accès public à Internet
Facilité d'intégration	3	1	4	5
TCO	2	1	5	4
Evolutivité	3	1	5	5
Adaptabilité	3	2	4	5
Isolation du réseau	3	4	5	1
Observabilité	3	4	5	5
Il est temps de réparer	3	2	5	5

Connexion avec AWS Site-to-Site VPN

AWS Site-to-Site VPNles connexions peuvent se terminer sur une passerelle privée virtuelle ou une passerelle de transit. Une passerelle privée virtuelle est le point de terminaison VPN situé du AWS côté de votre connexion Site-to-Site VPN qui peut être rattaché à un seul VPC. Une passerelle de transit est un hub de transit qui peut être utilisé pour interconnecter plusieurs VPCs réseaux locaux. Il peut également être utilisé comme point de terminaison VPN du AWS côté de la connexion Site-to-Site VPN. Cette section décrit les deux options.

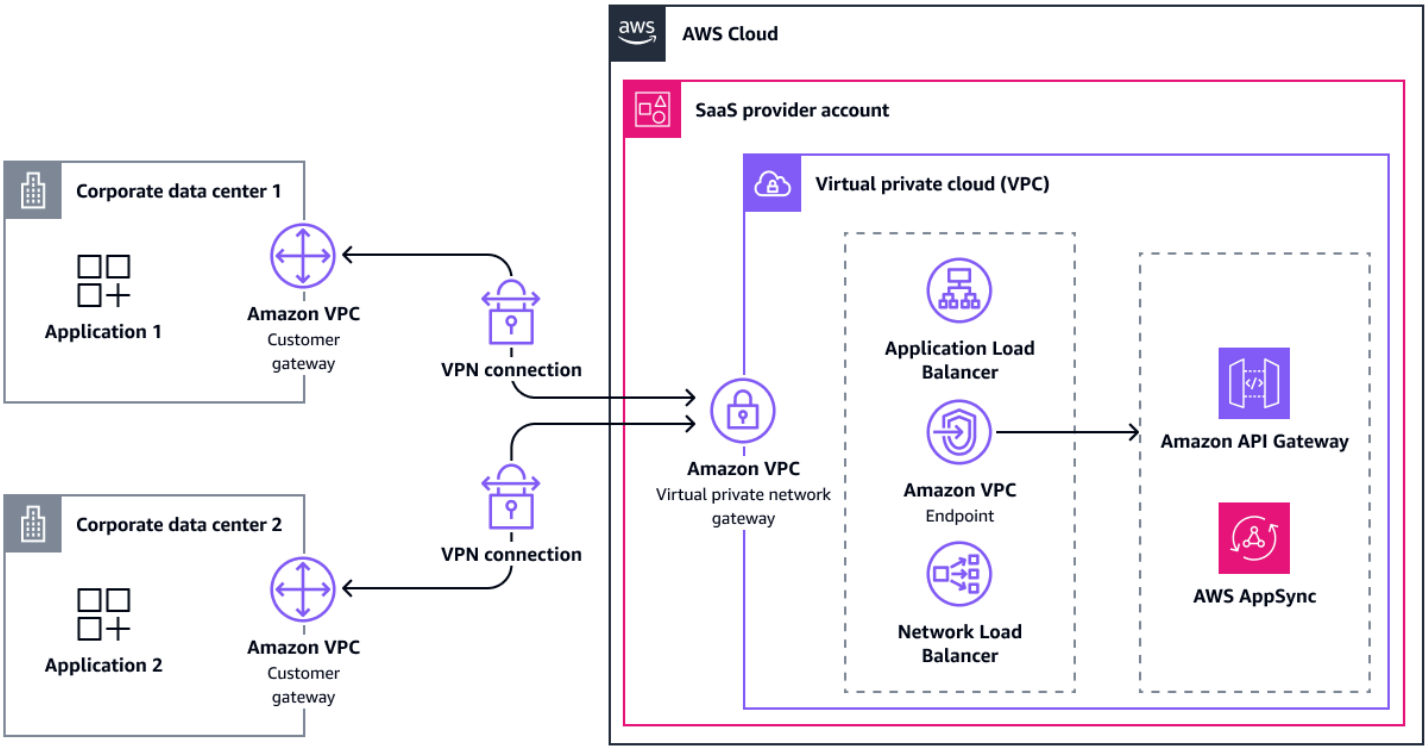
Connexion via une passerelle privée virtuelle

Après avoir créé une passerelle privée virtuelle, vous l'associez au VPC qui contient votre offre SaaS. Ensuite, vous activez la propagation des routes pour propager les routes VPN vers la table de routage du VPC. Ces routes peuvent être statiques ou dynamiques annoncées par le BGP.

Pour une haute disponibilité, une connexion Site-to-Site VPN possède deux tunnels VPN qui se terminent par deux zones de disponibilité sur le AWS côté. Si l'un d'entre eux n'est plus disponible, le second tunnel peut prendre le relais. Un seul tunnel permet une bande passante maximale de 1,25 Gbit/s. Les passerelles privées virtuelles ne prenant pas en charge le routage multichemin à coût égal (ECMP), vous ne pouvez utiliser qu'un seul tunnel à la fois.

Pour augmenter la tolérance aux pannes, vous pouvez configurer une deuxième connexion VPN vers une deuxième passerelle client physique. Une fois la connexion établie, le consommateur peut accéder aux ressources du VPC du fournisseur de SaaS.

Le schéma suivant illustre cette architecture.



Les avantages de cette approche sont les suivants :

- Temps de réparation : basculement géré vers le tunnel VPN secondaire
- Observabilité : intégration pour une surveillance active gérée à l'aide de Network Synthetic Monitor
- Facilité d'intégration : support de routage dynamique via BGP
- Adaptabilité : compatibilité avec la plupart des équipements réseau sur site
- Adaptabilité : soutien IPv6
- TCO : AWS Site-to-Site VPN est un service entièrement géré, il nécessite donc moins d'efforts opérationnels
- TCO : aucun coût pour les passerelles virtuelles, bien que des frais soient facturés pour les deux IPv4 adresses publiques associées à chacune
- Isolation du réseau : permet une communication privée sécurisée via Internet

Les inconvénients de cette approche sont les suivants :

- Facilité d'intégration : le consommateur doit configurer sa passerelle client
- Évolutivité : l'absence de support ECMP limite la bande passante à 1,25 Gbit/s par passerelle virtuelle

- Évolutivité : évolutivité limitée en raison de la complexité accrue du réseau et des frais opérationnels
- Adaptabilité : IPv6 prise en charge uniquement des adresses IP internes des tunnels VPN
- Adaptabilité : pas de routage transitif
- TCO : frais opérationnels liés à la maintenance, à la gestion et à la configuration de nombreuses connexions VPN pour le fournisseur de SaaS

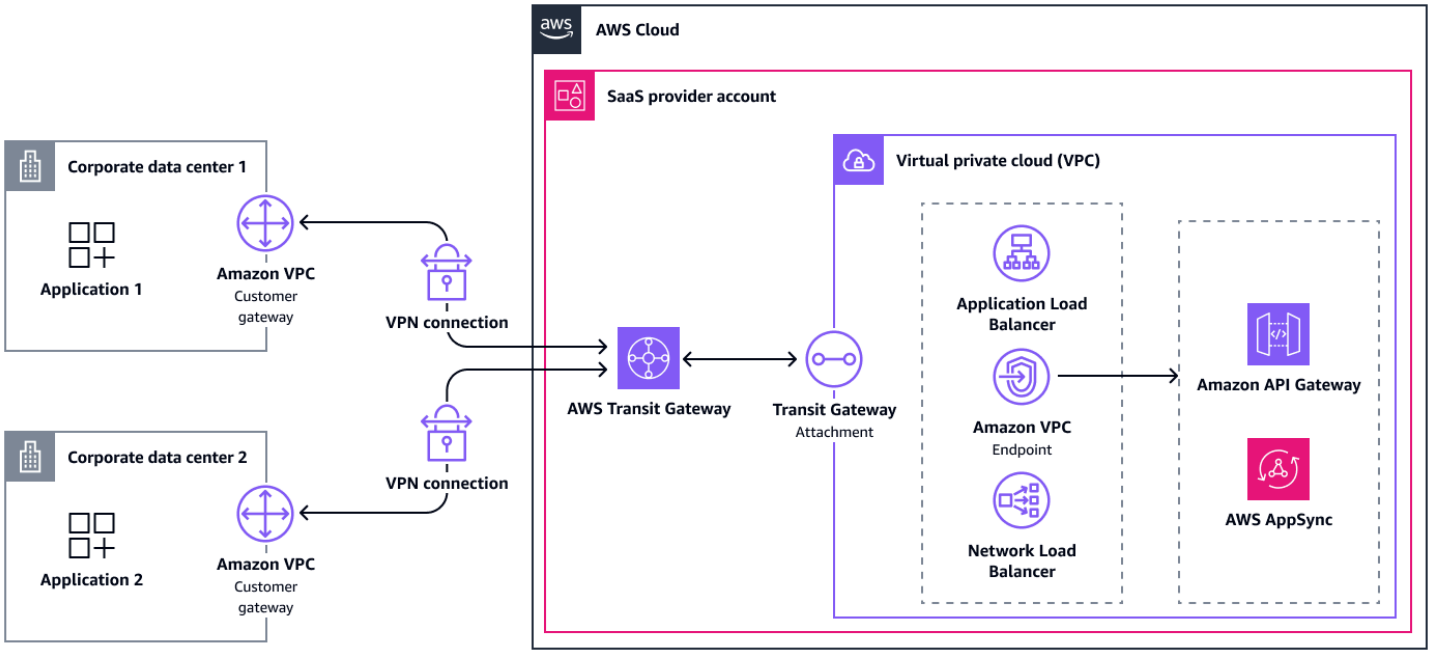
Connexion via une passerelle de transit

Les connexions via les passerelles de transit sont similaires aux passerelles virtuelles. Cependant, il y a quelques différences à garder à l'esprit.

Tout d'abord, les itinéraires de la pièce jointe VPN peuvent être automatiquement propagés dans la table de routage de la passerelle de transit, mais vous devez ajouter manuellement les itinéraires à la pièce jointe VPCs.

Comparé à une passerelle virtuelle, Transit Gateway prend en charge l'ECMP. Si la passerelle client prend en charge l'ECMP, elle peut utiliser les deux tunnels pour atteindre un débit maximal total de 2,5 Gbit/s. Vous pouvez établir plusieurs connexions entre le même réseau local et la passerelle de transit. Grâce à cette approche, vous pouvez augmenter la bande passante maximale jusqu'à 2,5 Gbit/s par connexion.

Le schéma suivant illustre cette architecture.



Les avantages de cette approche sont les suivants :

- Temps de réparation : basculement géré vers le tunnel VPN secondaire
- Observabilité : intégration pour une surveillance active gérée à l'aide de Network Synthetic Monitor
- Facilité d'intégration : support de routage dynamique via BGP
- Évolutivité : la prise en charge de l'ECMP permet de faire évoluer le débit du VPN pour répondre aux exigences de bande passante importantes
- Évolutivité : grand nombre de connexions VPN prises en charge par une seule passerelle de transit (jusqu'à près de 5 000)
- Évolutivité : un seul endroit pour gérer et surveiller toutes les connexions VPN
- Adaptabilité : compatibilité avec la plupart des équipements réseau sur site
- Adaptabilité : soutien IPv6
- Adaptabilité : héritez de la flexibilité de AWS Transit Gateway
- TCO : AWS Transit Gateway est un service entièrement géré, il nécessite donc moins d'efforts opérationnels
- TCO : aucun coût pour les passerelles virtuelles, bien que des frais soient facturés pour les deux IPv4 adresses publiques associées à chacune
- Isolation du réseau : permet une communication privée sécurisée via Internet

Les inconvénients de cette approche sont les suivants :

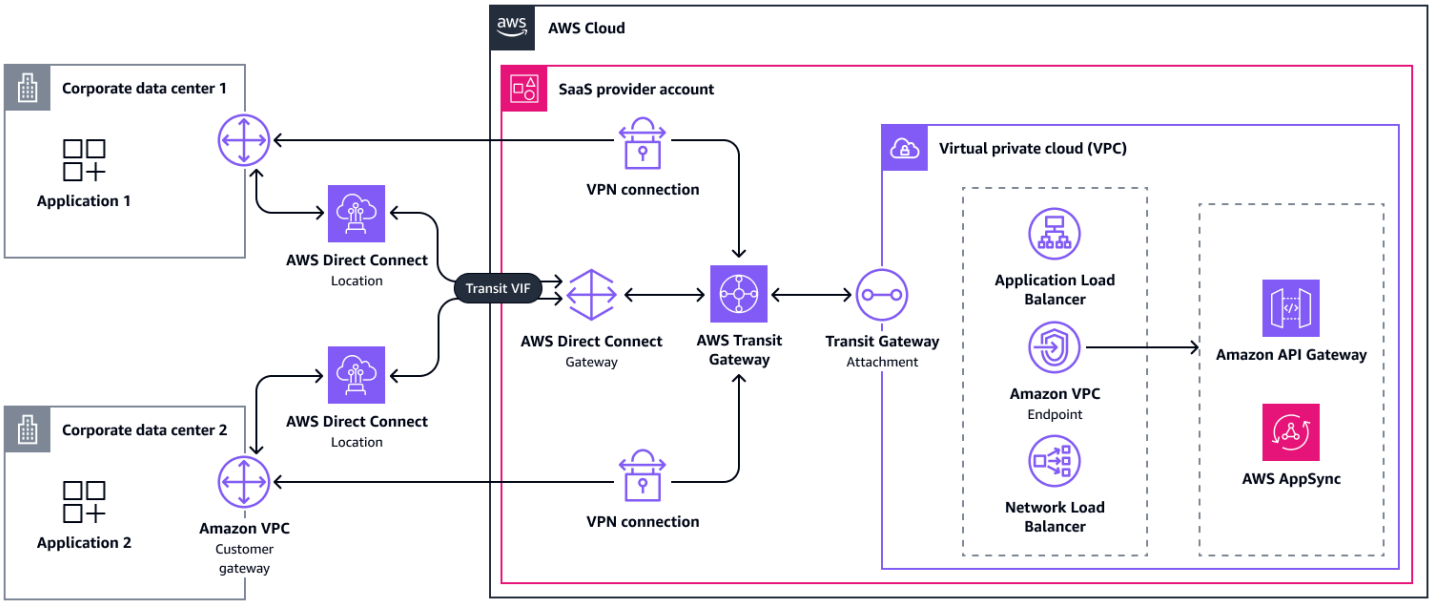
- Facilité d'intégration : le consommateur doit configurer sa passerelle client
- Évolutivité : évolutivité limitée en raison de la complexité accrue du réseau et des frais opérationnels
- Adaptabilité : IPv6 prise en charge uniquement des adresses IP internes des tunnels VPN
- TCO : frais opérationnels liés à la maintenance, à la gestion et à la configuration de nombreuses connexions VPN pour le fournisseur de SaaS
- TCO : frais supplémentaires pour l'utilisation de AWS Transit Gateway
- TCO : complexité accrue de la gestion des tables de routage des passerelles de transit

Connexion avec AWS Direct Connect

AWS Direct Connectrelie votre réseau interne à un Direct Connect emplacement via un câble à fibre optique Ethernet standard. Contrairement aux autres options d'architecture, une connexion dédiée ne peut pas être établie en quelques minutes. Au lieu de cela, ce processus peut prendre plusieurs jours si toutes les exigences sont satisfaites. Si ce n'est pas le cas, cela peut prendre plus de temps. Nous vous suggérons donc de contacter l'équipe chargée de votre AWS compte ou d' AWS Support obtenir de l'aide concernant cette approche. Vous pouvez éventuellement choisir une connexion hébergée fournie par un AWS partenaire et partagée avec d'autres clients. L'architecture est toujours la même. Vous pouvez choisir Direct Connect parce qu'il réduit le temps de latence, améliore la bande passante ou est conforme aux exigences réglementaires.

Pour utiliser la Direct Connect connexion, les consommateurs doivent créer une interface virtuelle publique, privée ou de transit. Différentes options d'architecture sont disponibles. La solution la plus flexible pour connecter plusieurs sites sur site AWS Cloud est une interface virtuelle de transit connectée à une Direct Connect passerelle. Une Direct Connect passerelle est un composant logique global qui permet au fournisseur de services d'y connecter jusqu'à six passerelles de transit. En outre, vous pouvez connecter jusqu'à 30 interfaces virtuelles à la passerelle. Pour des raisons d'échelle, vous pouvez créer des Direct Connect passerelles supplémentaires. Dans le compte du fournisseur SaaS, les passerelles de transit se rattachent ensuite au VPCs, comme décrit précédemment.

Les consommateurs peuvent se connecter en utilisant une à quatre Direct Connect connexions à partir d'un ou deux Direct Connect sites au total, selon le niveau de résilience souhaité. Pour plus d'informations, voir Configurer Direct Connect pour une résilience maximale. Une AWS Site-to-Site VPN connexion via Internet peut également servir de chemin de sauvegarde à moindre coût pour une Direct Connect connexion. Les connexions Direct Connect dédiées prises en charge peuvent être utilisées MACsecpour chiffrer le lien sur la couche 2 entre l' Direct Connect emplacement et le centre de données. Il est courant d'avoir une connexion Site-to-Site VPN pour renforcer la confidentialité des données. La connexion Site-to-Site VPN peut se terminer sur la passerelle de transit en utilisant une connexion VPN normale. Le schéma suivant illustre cette architecture.



Les avantages de cette approche sont les suivants :

- Observabilité : intégration pour une surveillance active gérée à l'aide de Network Synthetic Monitor
- Évolutivité : Support pour un débit de bande passante accru
- Adaptabilité : soutien IPv6
- TCO : potentiel de réduction du transfert de données
- TCO : expérience réseau cohérente
- Isolation du réseau : connectivité privée capable de répondre aux exigences réglementaires

Les inconvénients de cette approche sont les suivants :

- Facilité d'intégration : temps et effort manuel nécessaires à la configuration
- Évolutivité : évolutivité limitée au-delà des dizaines de Direct Connect connexions, car il existe plusieurs quotas à suivre
- Adaptabilité : les options de configuration dépendent des emplacements disponibles Direct Connect
- TCO : la Direct Connect maintenance planifiée peut entraîner des temps d'arrêt nécessitant une intervention

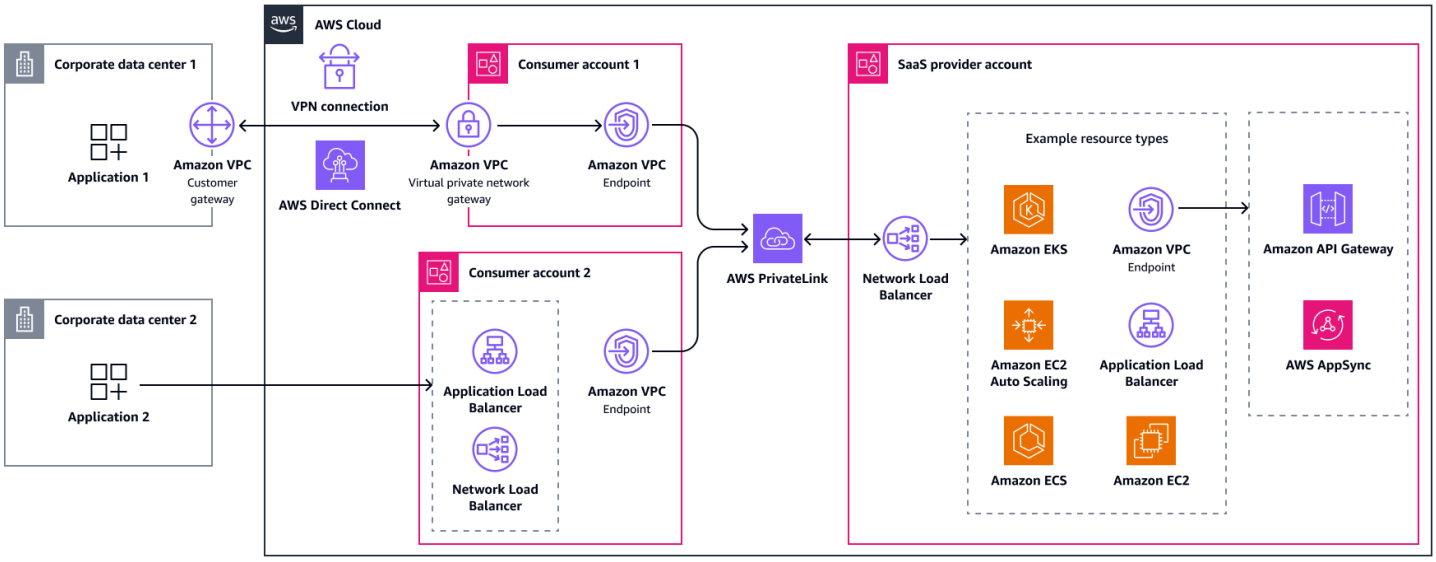
Connexion à une architecture VPC de transit

Transit VPC est une option d'architecture qui donne de la flexibilité aux consommateurs quant à la manière de se connecter AWS, et qui permet aux fournisseurs de SaaS de bénéficier d'un accès unifié à leur service via. AWS PrivateLink Le consommateur se connecte sur site à un VPC de transit qui ne contient qu'un point d'entrée (tel qu'une passerelle privée virtuelle) et un point de terminaison VPC d'interface, qui est une ressource. AWS PrivateLink Le transit VPCs doit appartenir soit au fournisseur de SaaS, soit aux consommateurs. Cette section décrit les deux options.

Vous pouvez créer le VPC de transit et les sous-réseaux avec des plages d'adresses CIDR compatibles avec le centre de données sur site. S'ils ont besoin d'une connectivité privée, les consommateurs peuvent se connecter à ce VPC via AWS Direct Connect ou. AWS Site-to-Site VPN Vous pouvez également configurer l'accès au compte de transit depuis l'Internet public à l'aide d'un Application Load Balancer ou d'un Network Load Balancer pointant vers le point de terminaison du VPC.

VPC de transit géré par le consommateur

Dans cette approche, le fournisseur de SaaS laisse la gestion du transit VPCs aux consommateurs. D'un point de vue technique, l'architecture du fournisseur de SaaS est la même que lors de la connexion directe aux AWS Cloud consommateurs AWS PrivateLink. Du point de vue des ventes et du produit, cela représente un effort supplémentaire, car certains consommateurs ne l'ont pas Comptes AWS encore fait. Ils peuvent hésiter à ouvrir et à gérer un compte. Le fournisseur de SaaS doit fournir des conseils à ses clients sur la manière de créer Comptes AWS et de connecter leur centre de données sur site. Le schéma suivant montre une combinaison d'accès public et privé, où les consommateurs sont propriétaires du transport en commun VPCs.



Les avantages de cette approche sont les suivants :

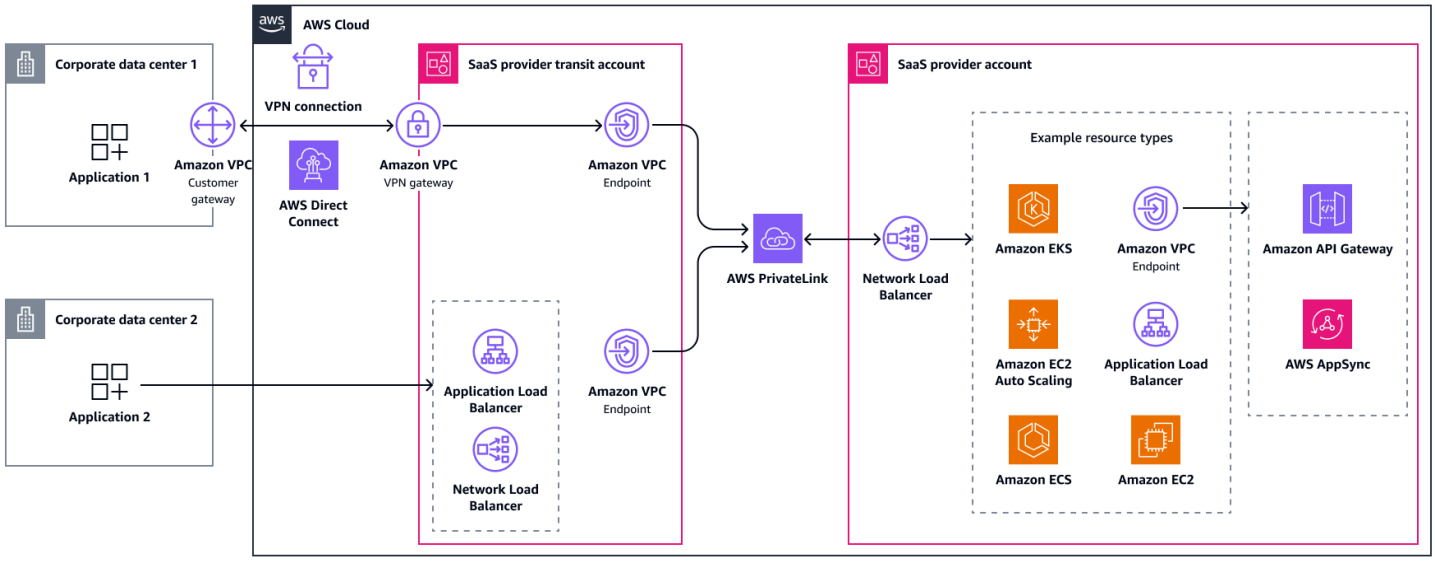
- Il est temps de réparer : les frais d'exploitation sont largement dévolus aux consommateurs du SaaS
- Adaptabilité : les utilisateurs du SaaS peuvent choisir parmi différentes options d'accès
- Adaptabilité : aucun conflit de plage CIDR, même en cas d'utilisation Site-to-Site d'un VPN ou Direct Connect
- Tous les indicateurs : le fournisseur de services hérite des avantages AWS PrivateLink

Les inconvénients de cette approche sont les suivants :

- Facilité d'intégration : les consommateurs de solutions SaaS ont besoin d'au moins un Compte AWS
- TCO : un VPC de transit est une architecture et non un service entièrement géré. Il nécessite donc un effort opérationnel accru

VPC de transit géré par le fournisseur

Cette approche utilise les mêmes technologies, mais les limites des comptes et les responsabilités changent. Ici, le fournisseur SaaS est propriétaire du transit VPCs, de préférence sur un compte distinct de celui de l'offre SaaS. Ce découplage réduit les coûts, réduit les risques et permet au compte de transport d'évoluer de manière indépendante. Pour les environnements nécessitant un degré élevé d'isolation, vous pouvez créer une séparation supplémentaire entre les locataires en utilisant un sous-réseau ou en créant un VPC de transit distinct pour chaque consommateur. Les consommateurs peuvent ensuite choisir le mode de connexion au VPC de transit. Cette approche offre davantage d'options pour étendre l'ensemble du marché adressable, mais elle entraîne un coût total de possession plus élevé pour le fournisseur de SaaS en raison de la nécessité d'exploiter et de surveiller des composants architecturaux supplémentaires.



Les avantages de cette approche sont les suivants :

- Adaptabilité : les utilisateurs du SaaS peuvent choisir parmi différentes options d'accès
- Adaptabilité : les consommateurs de SaaS n'ont pas besoin de Compte AWS
- Adaptabilité : aucun conflit de plage CIDR, même en cas d'utilisation Site-to-Site d'un VPN ou Direct Connect

Les inconvénients de cette approche sont les suivants :

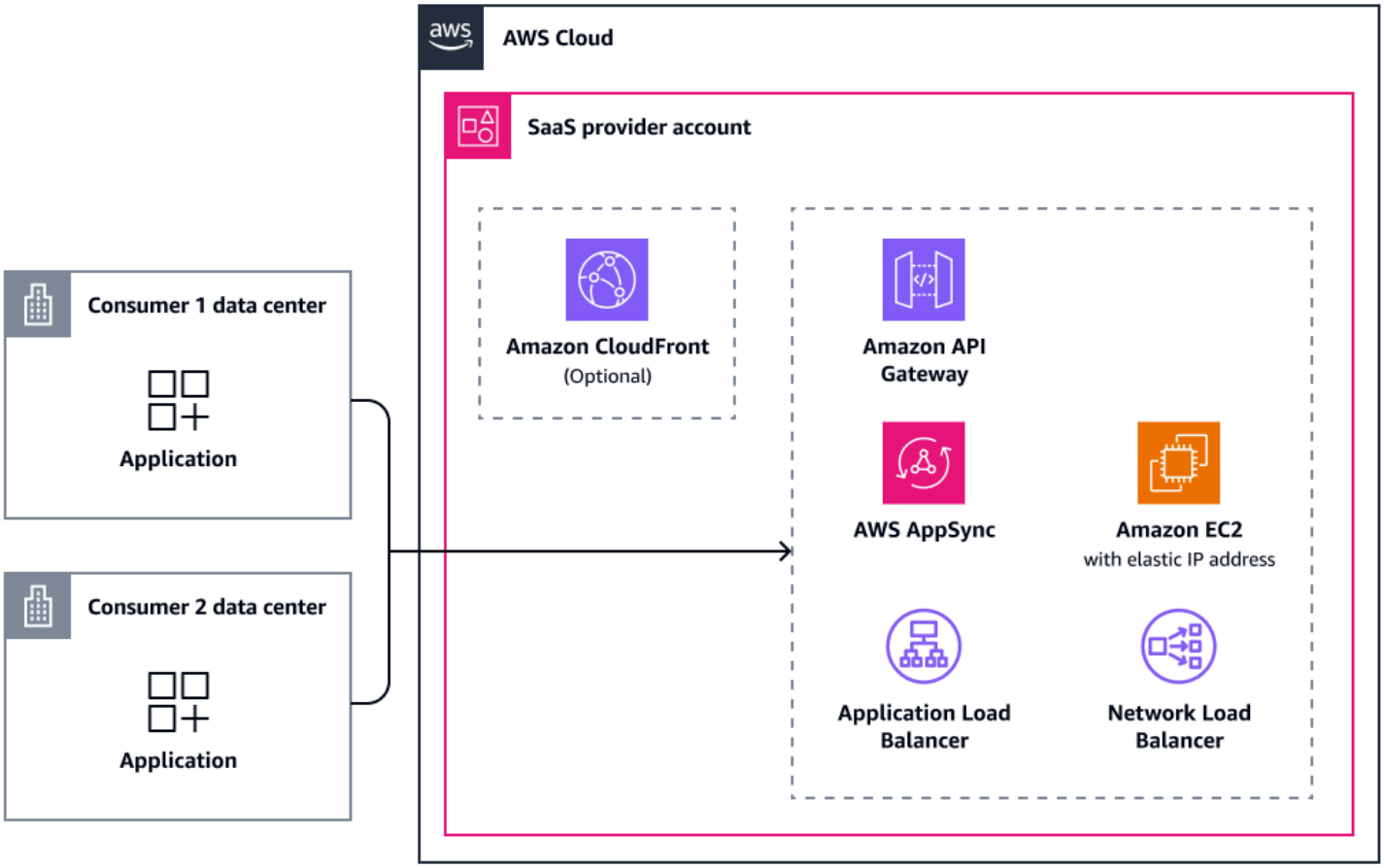
- TCO : un VPC de transit est une architecture et non un service entièrement géré. Il nécessite donc un effort opérationnel accru
- TCO : le fournisseur de SaaS doit exploiter et surveiller des composants architecturaux supplémentaires

Connexion via l'Internet public

L'accès public à Internet est également une option valable pour accéder à une offre SaaS, bien qu'il n'offre pas de connectivité privée au sens traditionnel du terme. Certains consommateurs peuvent toujours préférer une approche d'accès public, car elle ne nécessite aucune infrastructure réseau supplémentaire entre eux et le fournisseur de SaaS. Il réduit la complexité, les coûts et le temps d'intégration en échange d'une surface d'attaque accrue. Des mécanismes d'authentification et d'autorisation puissants peuvent contribuer à atténuer l'augmentation du niveau de menace, et vous

devez toujours chiffrer le trafic. Il est toujours recommandé de disposer d'un niveau de sécurité supplémentaire dans ce scénario, par exemple en utilisant AWS WAF.

L'architecture de ce scénario est simple. Le consommateur se connecte à un hébergeur public (le fournisseur de SaaS) via Internet. L'application peut être hébergée directement sur une instance publique Amazon Elastic Compute Cloud (Amazon EC2) dotée d'une adresse IP Elastic. L'option préférée est de l'héberger derrière un Application Load Balancer ou un service similaire. Pour améliorer les performances et la mise en cache des actifs statiques, vous pouvez utiliser un réseau de diffusion de contenu tel qu'Amazon CloudFront. Pour desservir une application avec une latence minimale sur deux adresses IP Anycast statiques globales, vous pouvez la placer AWS Global Acceleratordevant une instance Amazon EC2, Network Load Balancer ou Application Load Balancer. En outre CloudFront, les équilibreurs de charge d'application et Amazon API Gateway s'intègrent tous à AWS WAF. AWS AppSync Le schéma suivant donne un aperçu des options de connectivité pour l'accès public à Internet.



Le tableau suivant décrit les protocoles et les intégrations pris en charge pour ce scénario.

Service ou ressource	IPv6	AWS WAF intégration	Peut être un point de terminaison d'un accélérateur mondial
Amazon CloudFront	Pris en charge	Pris en charge	Non pris en charge
Amazon API Gateway	Pris en charge	Pris en charge	Non pris en charge
AWS AppSync	Partiellement pris en charge	Pris en charge	Non pris en charge
Amazon EC2 avec une adresse IP élastique	Pris en charge	Non pris en charge	Pris en charge
Application Load Balancer	Pris en charge	Pris en charge	Pris en charge
Network Load Balancer	Pris en charge	Non pris en charge	Pris en charge

Les avantages de cette approche sont les suivants :

- Facilité d'intégration : simplicité et accessibilité
- Évolutivité : échelle illimitée
- Adaptabilité : aucun conflit de plage CIDR possible
- Adaptabilité : soutien CloudFront

Les inconvénients de cette approche sont les suivants :

- Isolation du réseau : pas de connectivité privée
- Isolation du réseau : des mesures de sécurité strictes sont requises

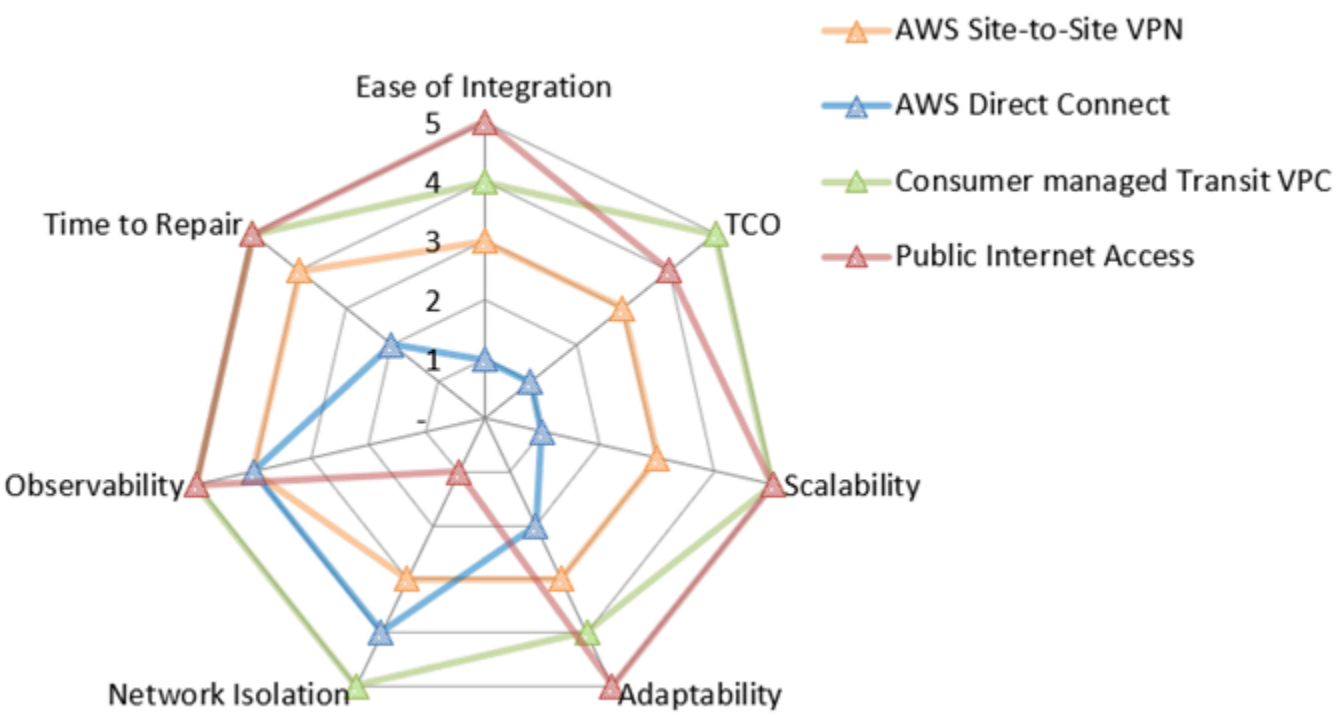
D'autres avantages et inconvénients s'appliquent, selon les services que vous choisissez.

Consommateurs de solutions SaaS faisant appel à d'autres fournisseurs de services cloud

Ce scénario décrit les solutions destinées aux consommateurs auprès d'autres fournisseurs de services cloud (CSPs). Ce scénario présente certains points communs avec les connexions aux centres de données locaux. En fait, toutes les options de connectivité pour les environnements sur site sont également valables pour les consommateurs CSPs, mais il AWS Direct Connect est même possible d'établir une connexion privée avec certaines CSPs d'entre elles. La plupart CSPs proposent de la documentation et une assistance sur la manière de se connecter au AWS Cloud via AWS Site-to-Site VPN ou AWS Direct Connect.

Lorsqu'ils choisissent Site-to-Site un VPN, les consommateurs peuvent bénéficier de passerelles gérées ou de ressources similaires provenant de leur fournisseur de services client respectif. Les consommateurs n'ont pas nécessairement à les configurer eux-mêmes, comme dans le scénario sur site. Cela influence certains paramètres du Site-to-Site VPN, tels que l'amélioration du délai de réparation et de l'observabilité. Cela est dû au fait que les deux extrémités de la connexion sont désormais gérées.

La carte des valeurs réseau suivante résume le score de chacune de ces options pour chaque métrique d'évaluation. Elle est très similaire à la carte des valeurs réseau pour les connexions sur site, bien que les valeurs pour le Site-to-Site VPN soient différentes. Pour plus d'informations sur les métriques d'évaluation, consultez Métriques d'évaluation ce guide. Sur la carte, cinq représente le meilleur score, tel que le coût total de possession le plus faible, la meilleure isolation du réseau ou le délai de réparation le plus court. Pour plus d'informations sur la lecture de cette carte radar, consultez Carte des valeurs du réseau ce guide.



La carte radar montre les valeurs suivantes.

Métrique d'évaluation	AWS Site-to-Site VPN	AWS Direct Connect	VPC de transit géré par le consommateur	Accès public à Internet
Facilité d'intégration	3	1	4	5
TCO	3	1	5	4
Evolutivité	3	1	5	5
Adaptabilité	3	2	4	5
Isolation du réseau	3	4	5	1
Observabilité	4	4	5	5

Il est temps de réparer	4	2	5	5

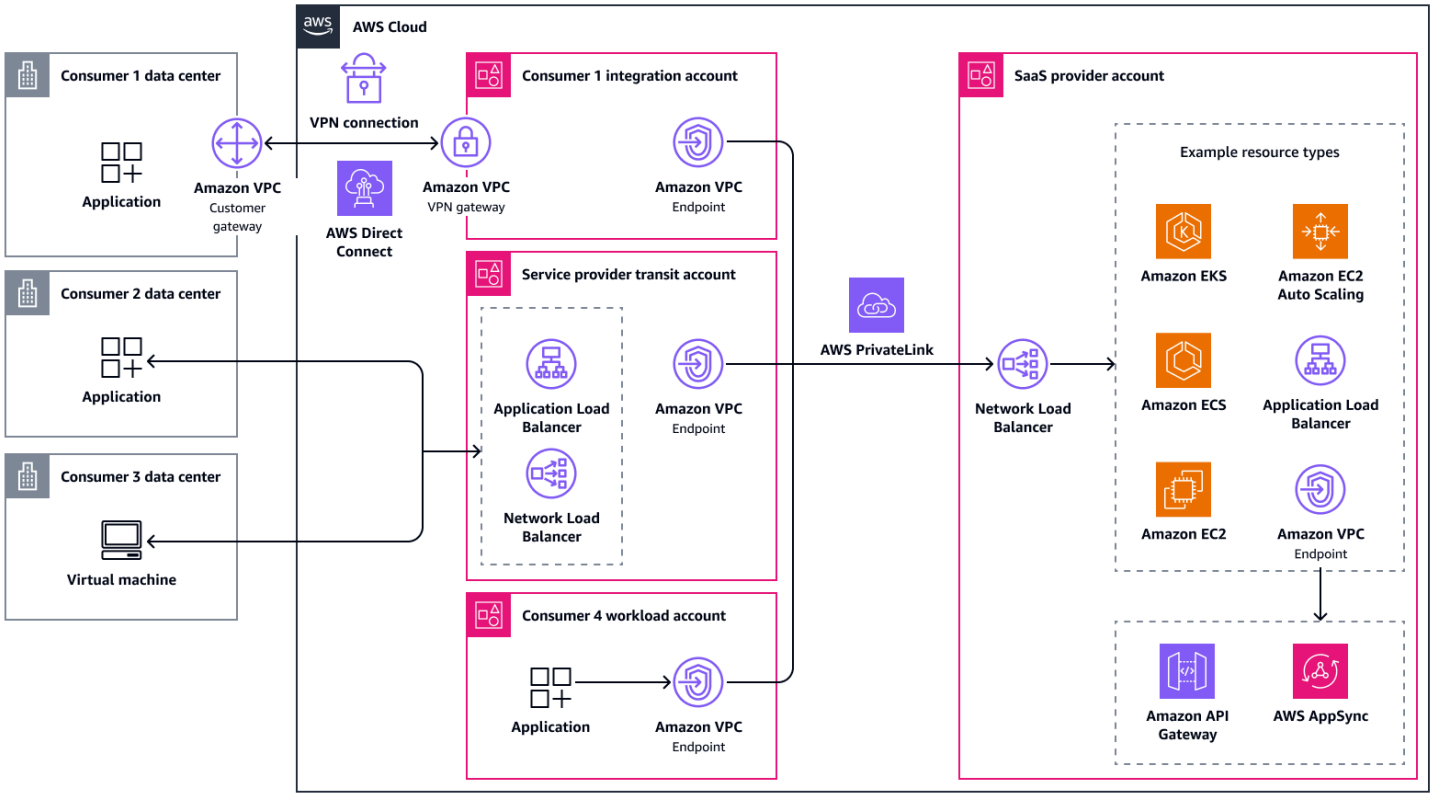
Prise en charge des environnements hybrides

Il est courant que les consommateurs proviennent d'environnements différents, chacun ayant ses propres contraintes techniques et de sécurité. Certains clients peuvent opérer entièrement à partir de centres de données sur site qui nécessitent une connectivité sécurisée via Internet ou via des liens réseau dédiés. D'autres exécutent peut-être déjà des charges de travail au sein de chemins réseau privés à faible latence AWS et s'attendent à une faible latence. Un troisième groupe peut s'appuyer sur un autre CSPs, où la connectivité doit relier différents réseaux cloud.

Quoi qu'il en soit, vous devez viser un accès réseau standardisé à votre application SaaS afin de simplifier votre architecture et de réduire la complexité opérationnelle. Deux des approches présentées précédemment, à savoir l'accès public à Internet et le transport en commun VPCs, fonctionnent bien dans tous ces scénarios. L'accès public à Internet constitue le moyen d'intégration le plus rapide avec une configuration minimale pour vos clients. VPCs Les transports en commun offrent un accès plus contrôlé et privé, souvent en utilisant AWS PrivateLink.

Lors de la conception de votre offre SaaS, vous pouvez adopter un modèle d'accès réseau unique ou combiner plusieurs approches dans une offre à plusieurs niveaux. Par exemple, vous pouvez proposer un niveau de déploiement d'accès public aux clients qui privilégient la facilité de connexion et une intégration rapide, et vous pouvez proposer un niveau de déploiement d'accès privé aux clients qui ont des exigences strictes en matière de conformité ou de contrôle de sécurité. Ces niveaux sont assortis de différents profils de coûts, de performances et de risques. Il est également possible de combiner les deux approches dans une architecture unique. Dans ce cas, assurez-vous de disposer de mesures de sécurité strictes afin que les chemins publics et privés restent isolés.

Le schéma suivant montre une approche d'accès hybride, dans laquelle les consommateurs ont la possibilité de se connecter en privé depuis leur centre de données ou leur CSP, publiquement ou directement via AWS PrivateLink (s'ils ont des charges de travail dans le AWS Cloud).


AWS Cloud
Consumer 1 data center
Application
Amazon VPC
Customer gateway
VPN connection
AWS Direct Connect
Amazon VPC
VPN gateway
Consumer 1 integration account
Amazon VPC
Endpoint
Consumer 2 data center
Application
Consumer 3 data center
Virtual machine
Service provider transit account
Application Load Balancer
Network Load Balancer
Amazon VPC
Endpoint
Consumer 4 workload account
Application
Amazon VPC
Endpoint
AWS PrivateLink
SaaS provider account
Network Load Balancer
Example resource types
Amazon EKS
Amazon EC2 Auto Scaling
Amazon ECS
Application Load Balancer
Amazon EC2
Amazon VPC
Endpoint
Amazon API Gateway
AWS AppSync

Scénarios d'accès réseau avancés pour les offres SaaS dans le AWS Cloud

Les architectures décrites dans Scénarios d'accès au réseau pour les offres SaaS dans le AWS Cloud cette section devraient vous aider à trouver une solution pour la majorité des cas d'utilisation. Cependant, certains scénarios présentent des exigences techniques spécifiques. Nombre d'entre eux dépassent le cadre de ce guide.

Cette section décrit les exigences et considérations techniques avancées suivantes :

- Communication bidirectionnelle
- TCP, UDP et protocoles propriétaires

Communication bidirectionnelle

Dans certains cas, les applications nécessitent un trafic bidirectionnel pour fonctionner comme prévu. Les cas d'utilisation courants sont les webhooks ou les services de notification. En général, vous pouvez y parvenir en établissant une WebSocket connexion entre le serveur et le client. Cette connexion maintient la session TCP ouverte et permet aux deux participants d'envoyer du trafic via la connexion. La plupart des services décrits dans ce guide sont pris en charge de manière native WebSocket, notamment les équilibreurs de charge réseau, les équilibreurs de charge d'application, Amazon API Gateway et AWS AppSync (via des points de AWS PrivateLink terminaison privés en temps réel).

Dans d'autres cas, une application du côté du fournisseur de SaaS peut avoir besoin d'accéder à des ressources côté consommateur, telles qu'une base de données. Lorsque vous vous connectez via des canaux bidirectionnels, tels qu'une AWS Site-to-Site VPN connexion, cela ne pose aucun problème.

D'autre part, AWS PrivateLink Elastic Load Balancing ne prend en charge que le trafic unidirectionnel. Si vous utilisez ces services, vous devez configurer un autre chemin réseau pour le trafic provenant de votre offre SaaS. Par exemple, il peut s'agir d'une AWS PrivateLink connexion supplémentaire qui va dans le sens inverse.

TCP, UDP et protocoles propriétaires

De nombreuses applications sont servies via HTTP ou HTTPS, mais pas toutes. Certains peuvent utiliser d'autres protocoles de couche 7 en plus du protocole TCP, tels que Message Queuing Telemetry Support (MQTT). D'autres pourraient même utiliser le protocole UDP pour servir les consommateurs. Dans de rares cas, les services utilisent des protocoles propriétaires qui doivent être transmis dans des paquets (couche 3). Pour ces scénarios, il est important de comprendre quels services prennent en charge votre offre SaaS.

Pour les services de couche 3, vous pouvez utiliser AWS PrivateLink des équilibreurs de charge réseau, qui prennent tous deux en charge l'ensemble du trafic TCP et UDP.

Pour les services de couche 7, les équilibreurs de charge d'application et Amazon CloudFront prennent en charge les protocoles HTTP WebSocket, HTTPS et Google Remote Procedure Calls (gRPC). De même, Amazon API Gateway et AWS AppSync chacun d'entre eux prennent en charge les protocoles HTTP, HTTPS et WebSocket. Amazon CloudFront est le seul service qui supporte actuellement le HTTP/3.

Vous pouvez utiliser Amazon VPC Lattice pour connecter des applications de couche 7 et des ressources de couche 3. Il prend en charge le transfert HTTP, HTTPS, gRPC, TCP et TLS.

Si l'application ne peut desservir le trafic que sur la couche 3, il est essentiel que vous utilisiez les services AWS réseau principaux, tels que AWS Transit Gateway, AWS Direct Connect AWS Site-to-Site VPN, et le peering VPC. Le trafic doit ensuite être acheminé directement du consommateur SaaS vers la couche de calcul de l'offre SaaS.

Anti-patterns pour l'accès au réseau dans AWS Cloud

Un anti-modèle est une solution fréquemment utilisée pour un problème récurrent où la solution est contre-productive, inefficace ou moins efficace qu'une alternative. Les options de conception mentionnées dans cette section fonctionnent généralement, mais elles présentent des inconvénients importants. Dans la mesure du possible, elles doivent être évitées car de meilleures alternatives sont disponibles.

Cette section aborde les anti-modèles et les défis suivants :

- Incompatibilité de la zone de disponibilité avec AWS PrivateLink
- AWS Site-to-Site VPN connexions entre Comptes AWS

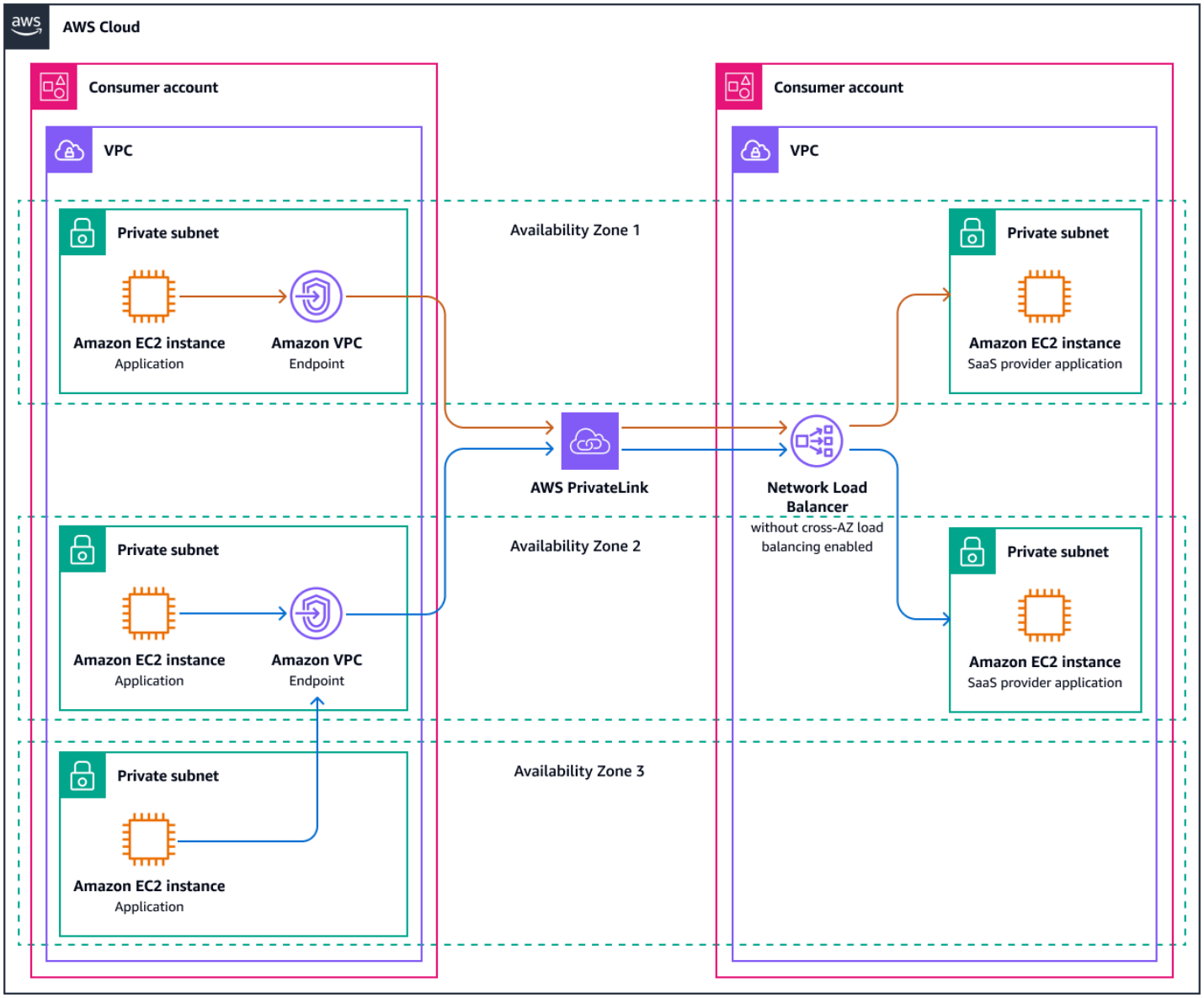
Incompatibilité de la zone de disponibilité avec AWS PrivateLink

Lorsqu'ils fournissent un accès à une application via AWS PrivateLink, les utilisateurs du SaaS peuvent créer des points de terminaison VPC d'interface uniquement dans les zones de disponibilité où l'application est déployée. Par exemple, si l'application est déployée dans `use1-az1` et`use1-az2`, le consommateur ne peut pas déployer de point de terminaison VPC dans. `use1-az3` Nous vous recommandons de déployer l'offre SaaS dans chaque zone de disponibilité. La majorité d'entre eux Régions AWS ont trois zones de disponibilité, bien que certains en aient plus. Pour une liste complète, voir Régions et zones de disponibilité. Tenez compte du nombre de zones de disponibilité lorsque vous choisissez un Région AWS.

> **Note**
>
> Les noms des zones de disponibilité sont différents de ceux des zones de disponibilité IDs. Pour plus d'informations, consultez la section Zone IDs de disponibilité de vos AWS ressources.

Si un fournisseur de SaaS choisit de ne pas le déployer dans toutes les zones de disponibilité, cela a des conséquences. Supposons que l'offre SaaS soit déployée dans `use1-az1` et`use1-az2`, mais que le consommateur utilise les trois zones de disponibilité, y compris`use1-az3`. Les points de terminaison VPC de l'interface sont déployés côté consommateur dans `use1-az1` et`use1-az2`, à présent, l'application `use1-az3` doit accéder à l'un de ces points de terminaison. Tout d'abord, le trafic doit être autorisé depuis les sous-réseaux des zones de disponibilité sans correspondance

vers les points de terminaison VPC respectifs. Le consommateur peut décider d'utiliser le nom AWS PrivateLink DNS régional, qui peut être attribué à l'un ou l'autre des points de terminaison VPC et qui répartit le trafic de manière égale entre les deux. Le consommateur peut également choisir d'envoyer le trafic directement vers un point de terminaison, par exemple`use1-az2`. Cela signifie que 67 % du trafic arrive du côté du fournisseur `use1-az2` et 33 % du trafic entrant`use1-az1`. La figure suivante illustre ce scénario.



Compte tenu du nombre important de consommateurs et de la répartition inégale du trafic, une charge de travail peut rencontrer des problèmes de capacité dans une zone de disponibilité et être insuffisante dans une autre. Pour résoudre ce problème, le fournisseur de SaaS peut décider d'équilibrer la charge du trafic de son côté de manière uniforme en activant l'équilibrage de charge entre zones sur le Network Load Balancer. Cela entraîne des frais supplémentaires.

Si le fournisseur de services ne correspond qu'à une seule zone de disponibilité, l'ensemble du trafic transitera par un seul point de terminaison. Cela crée un déséquilibre encore plus important. Par conséquent, l'offre SaaS n'est plus très disponible pour le consommateur. Peu importe pour le consommateur que l'application soit desservie par des zones de disponibilité supplémentaires qu'il n'utilise pas lui-même. Dans le pire des cas, un fournisseur de SaaS pourrait ne pas être en mesure de servir un consommateur qui n'utilise aucune des mêmes zones de disponibilité.

Dans les rares cas où le fournisseur SaaS ne dispose pas d'une option viable pour approvisionner son application sur toutes les zones de disponibilité, il est également possible de créer un sous-réseau uniquement dans les zones de disponibilité manquantes, puis d'étendre le service à ces zones de disponibilité vides. L'équilibrage de charge entre zones peut ensuite répartir le trafic entrant sur les points de terminaison de l'application réels dans les autres zones de disponibilité.

AWS Site-to-Site VPN connexions entre Comptes AWS

Les entreprises qui migrent d'environnements sur site vers le cloud tentent parfois de déplacer l'ensemble du réseau. Cela peut entraîner des problèmes car il existe des différences importantes entre les pratiques de mise en réseau sur site et dans le cloud. Si ce changement de mentalité ne se produit pas, des choses comme les AWS Site-to-Site VPN connexions d'un VPC à un autre VPC peuvent se produire. Cette approche ne permet pas de tirer parti des services réseau spécialement conçus dans le AWS Cloud, qui simplifient la gestion et améliorent les performances. L'adaptation aux conceptions natives du cloud permet de réduire les frais opérationnels et d'améliorer la fiabilité et l'évolutivité de la connectivité entre les deux VPCs.

Si vous envisagez de proposer cette option de connectivité en tant que fournisseur de SaaS, demandez-vous, ou demandez au consommateur, pourquoi elle AWS Site-to-Site VPN devrait être utilisée. Ensuite, revenez en arrière à partir de ces exigences pour trouver une meilleure option de connectivité. La section Comparaison des capacités de service de ce guide contient une matrice que vous pouvez utiliser pour identifier les options. Ensuite, vous pouvez parcourir les sections pertinentes de ce guide pour trouver une approche architecturale adaptée à votre cas d'utilisation.

Étapes suivantes

Ce guide décrit différentes approches d'accès au réseau dans différents scénarios, et décrit les avantages et les inconvénients de chaque architecture. Vous devez comprendre pourquoi le choix d'une approche d'accès au réseau ne doit pas être une simple discussion technologique. L'alignement entre le business et la technologie est essentiel. Les étapes et recommandations suivantes peuvent vous aider à évaluer et à normaliser votre stratégie d'architecture réseau en évaluant les capacités actuelles, en analysant les besoins du marché et en mettant en œuvre des contrôles de gouvernance.

Cette section contient les rubriques suivantes :

- Évaluation de l'architecture et des capacités actuelles
- Analyse du marché et de la clientèle
- Alignement stratégique
- Normalisation
- Gouvernance
- Répétition

Évaluation de l'architecture et des capacités actuelles

Passez en revue l'architecture réseau actuelle par rapport aux sources de données pertinentes, telles que le cadre d'auto-évaluation décrit dans ce guide, les exigences réglementaires en vigueur et l'état actuel du marché (à la fois en termes de clients et d'analyse concurrentielle). Par exemple, pensez à utiliser le AWS Well-Architected Framework, qui repose sur des décennies d'expérience dans l'exécution de systèmes de production à grande échelle dans le. AWS Cloud

Passez en revue les exceptions potentielles, les cas ponctuels et les décisions historiques relatives aux produits. Soyez curieux, mettez-les au défi et ne présumez pas automatiquement leur validité. Les exigences des clients d'il y a des années peuvent ne plus être valides. La remise en question des hypothèses crée l'opportunité de simplifier et de réduire la complexité de votre architecture.

En termes simples, documentez les observations afin qu'elles soient accessibles et comprises par les différents rôles de votre organisation. Capturez les points dans lesquels l'état actuel diffère de l'état cible, l'état cible, l'impact et le moment où les observations ont été effectuées. L'enregistrement de ces informations aide vos organisations à prendre des décisions basées sur de nouvelles données.

Analyse du marché et de la clientèle

Recueillez des informations sur les tendances du marché. Quel est actuellement le moyen préféré des consommateurs pour accéder à des offres SaaS comme la vôtre ? Est-ce que vous rencontrez toujours vos clients là où ils se trouvent ? Les cohortes de clients ou leur comportement ont-ils changé ? Vos dirigeants ont-ils orienté le navire vers un nouveau marché, une zone géographique soumise à des exigences réglementaires spécifiques ou un nouveau niveau de clientèle ? Votre modèle d'entreprise ou d'exploitation a-t-il changé ? Par exemple, envisagez-vous de mettre vos services en marque blanche ? Votre plan de croissance inclut-il la collaboration avec des partenaires afin que votre service soit disponible pour les clients lorsqu'ils entrent en contact avec ces partenaires ?

Alignement stratégique

Lorsque vous aurez compris vos capacités, votre architecture, votre marché et vos clients actuels, organisez une réunion d'alignement stratégique. Avec les acteurs concernés du produit, de l'entreprise et de la technologie, demandez-vous quelles exigences sont toujours valables et quelles nouvelles exigences doivent être prises en compte. Trouvez des opportunités pour réduire la complexité en supprimant les exigences qui ne sont plus nécessaires. Il ne s'agit pas d'une conception par un comité ; l'équipe d'ingénierie doit préparer et s'approprier l'architecture réelle et les détails de mise en œuvre. Cependant, cette réunion devrait expliquer pourquoi il s'agit de l'ensemble d'exigences qui maximise les avantages pour vos clients et votre organisation.

Normalisation

Pour attirer les clients, il peut être tentant de laisser chacun choisir librement comment se connecter à votre service. Après tout, n'importe quelle solution peut fonctionner techniquement, et vous pouvez également disposer du savoir-faire et des ressources nécessaires pour toutes les gérer et les exploiter. Cela peut bien fonctionner jusqu'à un certain point, mais à mesure que votre entreprise évolue, cela devient difficile à gérer. Votre infrastructure d'observabilité doit prendre en charge les indicateurs issus de plusieurs solutions, et les ingénieurs de fiabilité de votre site doivent également être en mesure de les comprendre. Vous avez besoin de up-to-date documentation pour chaque approche de connectivité. Les modifications majeures apportées à votre application doivent être évaluées par rapport à chaque approche d'accès que vous proposez. Vous devez écrire et gérer des automatisations et une infrastructure sous forme de code (IaC) pour chaque approche d'accès. Les frais supplémentaires liés à la non-standardisation de l'accès à votre service doivent être mis en balance avec la flexibilité que vous souhaitez offrir à vos clients.

Si vous avez besoin d'une étoile polaire pour guider votre prise de décision, nous vous suggérons la standardisation. La standardisation de la manière dont vos clients interagissent avec les services que vous fournissez est généralement la mesure la plus efficace que vous puissiez prendre pour améliorer de nombreux indicateurs de réussite au sein de votre organisation. La normalisation permet aux équipes produit de comprendre plus facilement la structure des coûts de vos services et de prendre des décisions relatives aux produits en fonction des données. Il est plus facile pour les équipes opérationnelles de résoudre les problèmes et d'automatiser certaines parties du processus de dépannage dans un environnement développé, déployé et exploité conformément à des normes prédéfinies. Il peut vous aider à détecter des anomalies, des comportements inattendus ou des actions d'un acteur malveillant. La normalisation réduit également la dette technique. Les équipes d'ingénierie mettent moins de cycles à tester et à mettre en œuvre les modifications apportées à la production. Cela peut également accélérer votre mise sur le marché, améliorer le succès de l'intégration en libre-service et réduire les risques réglementaires.

Par conséquent, nous vous suggérons de passer également en revue les mesures ponctuelles qui pourraient être en place aujourd'hui. Quantifiez le nombre de cycles opérationnels que vous consacrez à soutenir les clients existants. Comparez vos résultats avec les données historiques et déterminez si votre approche actuelle est évolutive pour les années à venir. Chaque fois qu'il est nécessaire de s'éloigner des normes, remettez en question les exigences qui sous-tendent ces demandes. Évaluez l'impact et équilibrez les avantages immédiats avec les engagements à long terme.

Dans les cas où la personnalisation est inévitable mais en conflit avec vos normes, envisagez un modèle de responsabilité partagée. Dans ce modèle, vos produits sont largement protégés des modifications demandées, et la personnalisation s'effectue dans un environnement minimaliste et dédié. Pour un exemple, consultez la Connexion à une architecture VPC de transit section.

Gouvernance

Pour se conformer aux exigences réglementaires et à vos propres normes internes, la gouvernance est essentielle. Une bonne gouvernance étant en place, vous pouvez contrôler où et comment appliquer les normes. Vous établissez également des contrôles pour détecter les divergences par rapport aux normes et informer les propriétaires des ressources des mesures correctives nécessaires. AWS Organizations, AWS ConfigAWS CloudTrail, et AWS Control Towerne sont que quelques-uns des nombreux outils Services AWS qui peuvent vous aider à gérer et à gouverner vos charges de travail dans le AWS Cloud.

Répétition

À l'aide des leçons tirées de vos premiers efforts, mettez en place un processus léger et reproductible pour rester aligné à l'avenir. Définissez les rôles dont vous avez besoin, à quelle fréquence, le degré de précision des données, la manière dont les données seront partagées et qui agira en conséquence.

Ressources

AWS documentation

- Intégration de services tiers dans le AWS Cloud (directives AWS prescriptives)
- Autorisation SaaS multi-locataires et contrôle d'accès aux API (directives AWS prescriptives)
- Gérez les locataires de plusieurs produits SaaS sur un seul plan de contrôle (directives AWS prescriptives)
- Qu'est-ce que c'est AWS Direct Connect ? (Direct Connect documentation)
- Qu'est-ce qu' AWS PrivateLink ? (documentation Amazon VPC)
- Qu'est-ce que c'est AWS Site-to-Site VPN ? (AWS Site-to-Site VPN documentation)
- Qu'est-ce que c'est AWS Transit Gateway ? (documentation Amazon VPC)
- Qu'est-ce que le peering VPC ? (documentation Amazon VPC)

Autres AWS ressources

- Options de connectivité Amazon Virtual Private Cloud (AWS livre blanc)
- AWS re:Invent 2021 - Comment choisir le bon équilibreur de charge pour vos AWS charges de travail () YouTube
- Qu'est-ce que le SaaS ? (AWS site Web)
- AWS Programme SaaS Factory (AWS Partner programme)
- Conseils pour les architectures multi-locataires sur AWS (bibliothèque de AWS solutions)

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide. Pour être averti des mises à jour à venir, abonnez-vous à un fil RSS.

Modification	Description	Date
Publication initiale	—	12 septembre 2025

AWS Glossaire des directives prescriptives

Les termes suivants sont couramment utilisés dans les stratégies, les guides et les modèles fournis par les directives AWS prescriptives. Pour suggérer des entrées, veuillez utiliser le lien Faire un commentaire à la fin du glossaire.

Nombres

7 R

Sept politiques de migration courantes pour transférer des applications vers le cloud. Ces politiques s'appuient sur les 5 R identifiés par Gartner en 2011 et sont les suivantes :

- Refactor/re-architect — Déplacez une application et modifiez son architecture en tirant pleinement parti des fonctionnalités natives du cloud pour améliorer l'agilité, les performances et l'évolutivité. Cela implique généralement le transfert du système d'exploitation et de la base de données. Exemple : migrez votre base de données Oracle sur site vers l' PostgreSQL-Compatible édition Amazon Aurora.
- Replateformer (déplacer et remodeler) : transférez une application vers le cloud et introduisez un certain niveau d'optimisation pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Amazon Relational Database Service (Amazon RDS) pour Oracle dans le. AWS Cloud
- Racheter (rachat) : optez pour un autre produit, généralement en passant d'une licence traditionnelle à un modèle SaaS. Exemple : migrez votre système de gestion de la relation client (CRM) vers Salesforce.com.
- Réhéberger (lift and shift) : transférez une application vers le cloud sans apporter de modifications pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Oracle sur une instance EC2 dans le. AWS Cloud
- Relocaliser (lift and shift au niveau de l'hyperviseur) : transférez l'infrastructure vers le cloud sans acheter de nouveau matériel, réécrire des applications ou modifier vos opérations existantes. Vous migrez des serveurs d'une plateforme sur site vers un service cloud pour la même plateforme. Exemple : migrer une Microsoft Hyper-V application vers AWS.
- Retenir : conservez les applications dans votre environnement source. Il peut s'agir d'applications nécessitant une refactorisation majeure, que vous souhaitez retarder, et d'applications existantes que vous souhaitez retenir, car rien ne justifie leur migration sur le plan commercial.

- Retirer : mettez hors service ou supprimez les applications dont vous n'avez plus besoin dans votre environnement source.

A

A2 (1) Agent-to-Agent

Protocole dynamique pour la collaboration agent-agent prenant en charge la délégation de tâches et le transfert d'état.

ABAC

Voir contrôle d'accès basé sur les attributs.

services abstraits

Consultez la section Services gérés.

ACIDE

Voir atomicité, consistance, isolation, durabilité.

migration active-active

Méthode de migration de base de données dans laquelle la synchronisation des bases de données source et cible est maintenue (à l'aide d'un outil de réplication bidirectionnelle ou d'opérations d'écriture double), tandis que les deux bases de données gèrent les transactions provenant de la connexion d'applications pendant la migration. Cette méthode prend en charge la migration par petits lots contrôlés au lieu d'exiger un basculement ponctuel. Elle est plus flexible mais demande plus de travail qu'une migration active-passive.

migration active-passive

Méthode de migration de base de données dans laquelle les bases de données source et cible sont synchronisées, mais seule la base de données source gère les transactions liées à la connexion des applications pendant que les données sont répliquées vers la base de données cible. La base de données cible n'accepte aucune transaction pendant la migration.

Agent

Un système d'IA capable de raisonner, de planifier et de prendre des mesures de manière autonome à l'aide d'outils pour atteindre des objectifs.

Agent Ops

Pratiques opérationnelles pour la création, le test, le déploiement et l'exécution d'agents d'IA en production à grande échelle.

fonction d'agrégation

Fonction SQL qui agit sur un groupe de lignes et calcule une valeur de retour unique pour le groupe. Des exemples de fonctions d'agrégation incluent `SUM` et`MAX`.

AI

Voir intelligence artificielle.

AIOps

Voir les opérations d'intelligence artificielle.

anonymisation

Processus de suppression définitive d'informations personnelles dans un ensemble de données. L'anonymisation peut contribuer à protéger la vie privée. Les données anonymisées ne sont plus considérées comme des données personnelles.

anti-motif

Solution fréquemment utilisée pour un problème récurrent lorsque la solution est contre-productive, inefficace ou moins efficace qu'une solution alternative.

contrôle des applications

Une approche de sécurité qui permet d'utiliser uniquement des applications approuvées afin de protéger un système contre les logiciels malveillants.

portefeuille d'applications

Ensemble d'informations détaillées sur chaque application utilisée par une organisation, y compris le coût de génération et de maintenance de l'application, ainsi que sa valeur métier. Ces informations sont essentielles pour le processus de découverte et d'analyse du portefeuille et permettent d'identifier et de prioriser les applications à migrer, à moderniser et à optimiser.

intelligence artificielle (IA)

Domaine de l'informatique consacré à l'utilisation des technologies de calcul pour exécuter des fonctions cognitives généralement associées aux humains, telles que l'apprentissage, la résolution de problèmes et la reconnaissance de modèles. Pour plus d'informations, veuillez consulter Qu'est-ce que l'intelligence artificielle ?

opérations d'intelligence artificielle (AIOps)

Processus consistant à utiliser des techniques de machine learning pour résoudre les problèmes opérationnels, réduire les incidents opérationnels et les interventions humaines, mais aussi améliorer la qualité du service. Pour plus d'informations sur la façon dont les AIOps sont utilisées dans la stratégie de migration AWS , veuillez consulter le guide d'intégration des opérations.

chiffrement asymétrique

Algorithme de chiffrement qui utilise une paire de clés, une clé publique pour le chiffrement et une clé privée pour le déchiffrement. Vous pouvez partager la clé publique, car elle n'est pas utilisée pour le déchiffrement, mais l'accès à la clé privée doit être très restreint.

atomicité, cohérence, isolement, durabilité (ACID)

Ensemble de propriétés logicielles garantissant la validité des données et la fiabilité opérationnelle d'une base de données, même en cas d'erreur, de panne de courant ou d'autres problèmes.

contrôle d'accès par attributs (ABAC)

Pratique qui consiste à créer des autorisations détaillées en fonction des attributs de l'utilisateur, tels que le service, le poste et le nom de l'équipe. Pour plus d'informations, consultez ABAC pour AWS dans la documentation Gestion des identités et des accès AWS (IAM).

source de données faisant autorité

Emplacement où vous stockez la version principale des données, considérée comme la source d'information la plus fiable. Vous pouvez copier les données de la source de données officielle vers d'autres emplacements à des fins de traitement ou de modification des données, par exemple en les anonymisant, en les expurgant ou en les pseudonymisant.

Zone de disponibilité

Un emplacement distinct au sein d'un Région AWS réseau isolé des défaillances dans d'autres zones de disponibilité et fournissant une connectivité réseau peu coûteuse et à faible latence aux autres zones de disponibilité de la même région.

AWS Cadre d'adoption du cloud (AWS CAF)

Un cadre de directives et de meilleures pratiques visant AWS à aider les entreprises à élaborer un plan efficace pour réussir leur migration vers le cloud. AWS La CAF organise ses conseils en six domaines prioritaires appelés perspectives : les affaires, les personnes, la gouvernance, les plateformes, la sécurité et les opérations. Les perspectives d'entreprise, de personnes et de gouvernance mettent l'accent sur les compétences et les processus métier, tandis que les

perspectives relatives à la plateforme, à la sécurité et aux opérations se concentrent sur les compétences et les processus techniques. Par exemple, la perspective liée aux personnes cible les parties prenantes qui s'occupent des ressources humaines (RH), des fonctions de dotation en personnel et de la gestion des personnes. Dans cette perspective, la AWS CAF fournit des conseils pour le développement du personnel, la formation et les communications afin de préparer l'organisation à une adoption réussie du cloud. Pour plus d'informations, veuillez consulter le site Web AWS CAF et le livre blanc AWS CAF.

AWS Cadre de qualification de la charge de travail (AWS WQF)

Outil qui évalue les charges de travail liées à la migration des bases de données, recommande des stratégies de migration et fournit des estimations de travail. AWS Le WQF est inclus avec AWS Schema Conversion Tool (AWS SCT). Il analyse les schémas de base de données et les objets de code, le code d'application, les dépendances et les caractéristiques de performance, et fournit des rapports d'évaluation.

B

mauvais bot

Un bot destiné à perturber ou à nuire à des individus ou à des organisations.

BCP

Consultez la section Planification de la continuité des activités.

graphique de comportement

Vue unifiée et interactive des comportements des ressources et des interactions au fil du temps. Vous pouvez utiliser un graphique de comportement avec Amazon Detective pour examiner les tentatives de connexion infructueuses, les appels d'API suspects et les actions similaires. Pour plus d'informations, veuillez consulter Data in a behavior graph dans la documentation Detective.

système de poids fort

Système qui stocke d'abord l'octet le plus significatif. Voir aussi endianité.

classification binaire

Processus qui prédit un résultat binaire (l'une des deux classes possibles). Par exemple, votre modèle de machine learning peut avoir besoin de prévoir des problèmes tels que « Cet e-mail est-il du spam ou non ? » ou « Ce produit est-il un livre ou une voiture ? ».

filtre de Bloom

Structure de données probabiliste et efficace en termes de mémoire qui est utilisée pour tester si un élément fait partie d'un ensemble.

blue/green déploiement

Stratégie de déploiement dans laquelle vous créez deux environnements distincts mais identiques. Vous exécutez la version actuelle de l'application dans un environnement (bleu) et la nouvelle version de l'application dans l'autre environnement (vert). Cette stratégie vous permet de revenir rapidement en arrière avec un impact minimal.

bot

Application logicielle qui exécute des tâches automatisées sur Internet et simule l'activité ou l'interaction humaine. Certains robots sont utiles ou bénéfiques, comme les robots d'exploration Web qui indexent des informations sur Internet. D'autres robots, appelés « bots malveillants », sont destinés à perturber ou à nuire à des individus ou à des organisations.

botnet

Réseaux de robots infectés par des logiciels malveillants et contrôlés par une seule entité, connue sous le nom d'herder ou d'opérateur de bots. Les botnets sont le mécanisme le plus connu pour faire évoluer les bots et leur impact.

branche

Zone contenue d'un référentiel de code. La première branche créée dans un référentiel est la branche principale. Vous pouvez créer une branche à partir d'une branche existante, puis développer des fonctionnalités ou corriger des bogues dans la nouvelle branche. Une branche que vous créez pour générer une fonctionnalité est communément appelée branche de fonctionnalités. Lorsque la fonctionnalité est prête à être publiée, vous fusionnez à nouveau la branche de fonctionnalités dans la branche principale. Pour plus d'informations, consultez À propos des branches (GitHub documentation).

accès par brise-vitre

Dans des circonstances exceptionnelles et par le biais d'un processus approuvé, c'est un moyen rapide pour un utilisateur d'accéder à un accès auquel Compte AWS il n'est généralement pas autorisé. Pour plus d'informations, consultez l'indicateur Mettre en œuvre des procédures permettant de briser le verre dans le AWS Well-Architected guide.

stratégie existante (brownfield)

L'infrastructure existante de votre environnement. Lorsque vous adoptez une stratégie existante pour une architecture système, vous concevez l'architecture en fonction des contraintes des systèmes et de l'infrastructure actuels. Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et greenfield (inédites).

cache de tampon

Zone de mémoire dans laquelle sont stockées les données les plus fréquemment consultées.

capacité métier

Ce que fait une entreprise pour générer de la valeur (par exemple, les ventes, le service client ou le marketing). Les architectures de microservices et les décisions de développement peuvent être dictées par les capacités métier. Pour plus d'informations, veuillez consulter la section Organisation en fonction des capacités métier du livre blanc Exécution de microservices conteneurisés sur AWS.

planification de la continuité des activités (BCP)

Plan qui tient compte de l'impact potentiel d'un événement perturbateur, tel qu'une migration à grande échelle, sur les opérations, et qui permet à une entreprise de reprendre ses activités rapidement.

C

CAF

Voir le cadre d'adoption du AWS cloud.

déploiement de Canary

Diffusion lente et progressive d'une version pour les utilisateurs finaux. Lorsque vous êtes sûr, vous déployez la nouvelle version et remplacez la version actuelle dans son intégralité.

CCoE

Voir le Centre d'excellence du cloud.

CDC

Consultez la section Capture des données de modification.

capture des données de modification (CDC)

Processus de suivi des modifications apportées à une source de données, telle qu'une table de base de données, et d'enregistrement des métadonnées relatives à ces modifications. Vous pouvez utiliser la CDC à diverses fins, telles que l'audit ou la réplication des modifications dans un système cible afin de maintenir la synchronisation.

ingénierie du chaos

Introduire intentionnellement des défaillances ou des événements perturbateurs pour tester la résilience d'un système. Vous pouvez utiliser AWS Fault Injection Service (AWS FIS) pour effectuer des expériences qui stressent vos AWS charges de travail et évaluer leur réponse.

CI/CD

Découvrez l'intégration continue et la livraison continue.

classification

Processus de catégorisation qui permet de générer des prédictions. Les modèles de ML pour les problèmes de classification prédisent une valeur discrète. Les valeurs discrètes se distinguent toujours les unes des autres. Par exemple, un modèle peut avoir besoin d'évaluer la présence ou non d'une voiture sur une image.

Développeur citoyen

Un utilisateur professionnel qui crée des applications d'intelligence artificielle à l'aide de plateformes sans code/low code sans compétences techniques spécialisées.

chiffrement côté client

Chiffrement des données localement, avant que la cible ne les Service AWS reçoive.

Centre d'excellence cloud (CCoE)

Une équipe multidisciplinaire qui dirige les efforts d'adoption du cloud au sein d'une organisation, notamment en développant les bonnes pratiques en matière de cloud, en mobilisant des ressources, en établissant des délais de migration et en guidant l'organisation dans le cadre de transformations à grande échelle. Pour plus d'informations, consultez les articles du CCoE sur le blog de stratégie AWS Cloud d'entreprise.

cloud computing

Technologie cloud généralement utilisée pour le stockage de données à distance et la gestion des appareils IoT. Le cloud computing est généralement associé à la technologie informatique de pointe.

modèle d'exploitation du cloud

Dans une organisation informatique, modèle d'exploitation utilisé pour créer, faire évoluer et optimiser un ou plusieurs environnements cloud. Pour plus d'informations, consultez la section Création de votre modèle d'exploitation cloud.

étapes d'adoption du cloud

Les quatre phases que les entreprises traversent généralement lorsqu'elles migrent vers AWS Cloud :

- Projet : exécution de quelques projets liés au cloud à des fins de preuve de concept et d'apprentissage
- Base : réaliser des investissements fondamentaux pour mettre à l'échelle l'adoption du cloud (par exemple, en créant une zone de destination, en définissant un CCoE ou en établissant un modèle opérationnel)
- Migration : migration d'applications individuelles
- Re-invention — Optimisation des produits et services et innovation dans le cloud

Ces étapes ont été définies par Stephen Orban dans le billet de blog The Journey Toward Cloud-First & the Stages of Adoption publié sur le blog AWS Cloud Enterprise Strategy. Pour plus d'informations sur leur lien avec la stratégie de AWS migration, consultez le guide de préparation à la migration.

CMDB

Consultez la base de données de gestion des configurations.

référentiel de code

Emplacement où le code source et d'autres ressources, comme la documentation, les exemples et les scripts, sont stockés et mis à jour par le biais de processus de contrôle de version. Les référentiels cloud courants incluent GitHub ouBitbucket Cloud. Chaque version du code est appelée branche. Dans une structure de microservice, chaque référentiel est consacré à une seule fonctionnalité. Un CI/CD pipeline unique peut utiliser plusieurs référentiels.

cache passif

Cache tampon vide, mal rempli ou contenant des données obsolètes ou non pertinentes. Cela affecte les performances, car l'instance de base de données doit lire à partir de la mémoire principale ou du disque, ce qui est plus lent que la lecture à partir du cache tampon.

données gelées

Données rarement consultées et généralement historiques. Lorsque vous interrogez ce type de données, les requêtes lentes sont généralement acceptables. Le transfert de ces données vers des niveaux ou classes de stockage moins performants et moins coûteux peut réduire les coûts.

vision par ordinateur (CV)

Domaine de l'IA qui utilise l'apprentissage automatique pour analyser et extraire des informations à partir de formats visuels tels que des images numériques et des vidéos. Par exemple, Amazon SageMaker AI fournit des algorithmes de traitement d'image pour les CV.

dérive de configuration

Pour une charge de travail, une modification de configuration par rapport à l'état attendu. Cela peut entraîner une non-conformité de la charge de travail, et cela est généralement progressif et involontaire.

base de données de gestion des configurations (CMDB)

Référentiel qui stocke et gère les informations relatives à une base de données et à son environnement informatique, y compris les composants matériels et logiciels ainsi que leurs configurations. Vous utilisez généralement les données d'une CMDB lors de la phase de découverte et d'analyse du portefeuille de la migration.

pack de conformité

Ensemble de AWS Config règles et d'actions correctives que vous pouvez assembler pour personnaliser vos contrôles de conformité et de sécurité. Vous pouvez déployer un pack de conformité en tant qu'entité unique dans une région Compte AWS et, ou au sein d'une organisation, à l'aide d'un modèle YAML. Pour plus d'informations, consultez la section Packs de conformité dans la AWS Config documentation.

intégration continue et livraison continue (CI/CD)

Processus d'automatisation des étapes de source, de construction, de test, de préparation et de production du processus de publication du logiciel. CI/CD est communément décrit comme un pipeline. CI/CD peut vous aider à automatiser les processus, à améliorer la productivité, à améliorer la qualité du code et à accélérer les livraisons. Pour plus d'informations, veuillez consulter Avantages de la livraison continue. CD peut également signifier déploiement continu. Pour plus d'informations, veuillez consulter Livraison continue et déploiement continu.

CV

Voir vision par ordinateur.

D

données au repos

Données stationnaires dans votre réseau, telles que les données stockées.

classification des données

Processus permettant d'identifier et de catégoriser les données de votre réseau en fonction de leur sévérité et de leur sensibilité. Il s'agit d'un élément essentiel de toute stratégie de gestion des risques de cybersécurité, car il vous aide à déterminer les contrôles de protection et de conservation appropriés pour les données. La classification des données est une composante du pilier de sécurité du AWS Well-Architected cadre. Pour plus d'informations, veuillez consulter Classification des données.

dérive des données

Une variation significative entre les données de production et les données utilisées pour entraîner un modèle ML, ou une modification significative des données d'entrée au fil du temps. La dérive des données peut réduire la qualité, la précision et l'équité globales des prédictions des modèles ML.

données en transit

Données qui circulent activement sur votre réseau, par exemple entre les ressources du réseau.

maillage de données

Un cadre architectural qui fournit une propriété des données distribuée et décentralisée avec une gestion et une gouvernance centralisées.

minimisation des données

Le principe de collecte et de traitement des seules données strictement nécessaires. La pratique de la minimisation des données AWS Cloud peut réduire les risques liés à la confidentialité, les coûts et l'empreinte carbone de vos analyses.

périmètre de données

Ensemble de garde-fous préventifs dans votre AWS environnement qui permettent de garantir que seules les identités fiables accèdent aux ressources fiables des réseaux attendus. Pour plus d'informations, voir Création d'un périmètre de données sur AWS.

prétraitement des données

Pour transformer les données brutes en un format facile à analyser par votre modèle de ML. Le prétraitement des données peut impliquer la suppression de certaines colonnes ou lignes et le traitement des valeurs manquantes, incohérentes ou en double.

provenance des données

Le processus de suivi de l'origine et de l'historique des données tout au long de leur cycle de vie, par exemple la manière dont les données ont été générées, transmises et stockées.

sujet des données

Personne dont les données sont collectées et traitées.

entrepôt des données

Un système de gestion des données qui prend en charge les informations commerciales, telles que les analyses. Les entrepôts de données contiennent généralement de grandes quantités de données historiques et sont généralement utilisés pour les requêtes et les analyses.

langage de définition de base de données (DDL)

Instructions ou commandes permettant de créer ou de modifier la structure des tables et des objets dans une base de données.

langage de manipulation de base de données (DML)

Instructions ou commandes permettant de modifier (insérer, mettre à jour et supprimer) des informations dans une base de données.

DDL

Voir langage de définition de base de données.

ensemble profond

Sert à combiner plusieurs modèles de deep learning à des fins de prédiction. Vous pouvez utiliser des ensembles profonds pour obtenir une prévision plus précise ou pour estimer l'incertitude des prédictions.

deep learning

Un sous-champ de ML qui utilise plusieurs couches de réseaux neuronaux artificiels pour identifier le mappage entre les données d'entrée et les variables cibles d'intérêt.

défense en profondeur

Approche de la sécurité de l'information dans laquelle une série de mécanismes et de contrôles de sécurité sont judicieusement répartis sur l'ensemble d'un réseau informatique afin de protéger la confidentialité, l'intégrité et la disponibilité du réseau et des données qu'il contient. Lorsque vous adoptez cette stratégie AWS, vous ajoutez plusieurs contrôles à différentes couches de la AWS Organizations structure afin de sécuriser les ressources. Par exemple, une approche de défense approfondie peut combiner l'authentification multifactorielle, la segmentation du réseau et le chiffrement.

administrateur délégué

Dans AWS Organizations, un service compatible peut enregistrer un compte AWS membre pour administrer les comptes de l'organisation et gérer les autorisations pour ce service. Ce compte est appelé administrateur délégué pour ce service. Pour plus d'informations et une liste des services compatibles, veuillez consulter la rubrique Services qui fonctionnent avec AWS Organizations dans la documentation AWS Organizations .

déploiement

Processus de mise à disposition d'une application, de nouvelles fonctionnalités ou de corrections de code dans l'environnement cible. Le déploiement implique la mise en œuvre de modifications dans une base de code, puis la génération et l'exécution de cette base de code dans les environnements de l'application.

environnement de développement

Voir environnement.

contrôle de détection

Contrôle de sécurité conçu pour détecter, journaliser et alerter après la survenue d'un événement. Ces contrôles constituent une deuxième ligne de défense et vous alertent en cas d'événements de sécurité qui ont contourné les contrôles préventifs en place. Pour plus d'informations, veuillez consulter la rubrique Contrôles de détection dans Implementing security controls on AWS.

cartographie de la chaîne de valeur du développement (DVSM)

Processus utilisé pour identifier et hiérarchiser les contraintes qui nuisent à la rapidité et à la qualité du cycle de vie du développement logiciel. DVSM étend le processus de cartographie de la chaîne de valeur initialement conçu pour les pratiques de production allégée. Il met l'accent sur les étapes et les équipes nécessaires pour créer et transférer de la valeur tout au long du processus de développement logiciel.

jumeau numérique

Représentation virtuelle d'un système réel, tel qu'un bâtiment, une usine, un équipement industriel ou une ligne de production. Les jumeaux numériques prennent en charge la maintenance prédictive, la surveillance à distance et l'optimisation de la production.

tableau des dimensions

Dans un schéma en étoile, table plus petite contenant les attributs de données relatifs aux données quantitatives d'une table de faits. Les attributs des tables de dimensions sont généralement des champs de texte ou des nombres discrets qui se comportent comme du texte. Ces attributs sont couramment utilisés pour la contrainte des requêtes, le filtrage et l'étiquetage des ensembles de résultats.

catastrophe

Un événement qui empêche une charge de travail ou un système d'atteindre ses objectifs commerciaux sur son site de déploiement principal. Ces événements peuvent être des catastrophes naturelles, des défaillances techniques ou le résultat d'actions humaines, telles qu'une mauvaise configuration involontaire ou une attaque de logiciel malveillant.

reprise après sinistre (DR)

La stratégie et le processus que vous utilisez pour minimiser les temps d'arrêt et les pertes de données causés par un sinistre. Pour plus d'informations, consultez la section Reprise après sinistre des charges de travail sur AWS : Restauration dans le cloud dans le AWS Well-Architected Framework.

DML

Voir langage de manipulation de base de données.

conception axée sur le domaine

Approche visant à développer un système logiciel complexe en connectant ses composants à des domaines évolutifs, ou objectifs métier essentiels, que sert chaque composant. Ce concept a été introduit par Eric Evans dans son livre, Domain-Driven Design : Tackling Complexity in the Heart of Software (Boston : Addison-Wesley Professional, 2003). Pour plus d'informations sur la manière dont vous pouvez utiliser la conception axée sur le domaine avec le modèle Strangler Fig, consultez la section Modernisation incrémentielle des anciens services Web ASP.NET Microsoft (ASMX) à l'aide de conteneurs et d'Amazon API Gateway.

DR

Consultez la section Reprise après sinistre.

détection de dérive

Suivi des écarts par rapport à une configuration de référence. Par exemple, vous pouvez l'utiliser AWS CloudFormation pour détecter la dérive des ressources du système ou AWS Control Tower pour détecter les modifications de votre zone d'atterrissage susceptibles d'affecter le respect des exigences de gouvernance.

DVSM

Voir la cartographie de la chaîne de valeur du développement.

E

EDA

Voir analyse exploratoire des données.

EDI

Voir échange de données informatisé.

informatique de périphérie

Technologie qui augmente la puissance de calcul des appareils intelligents en périphérie d'un réseau IoT. Comparé au cloud computing, l'informatique de pointe peut réduire la latence des communications et améliorer le temps de réponse.

échange de données informatisé (EDI)

L'échange automatique de documents commerciaux entre les organisations. Pour plus d'informations, voir Qu'est-ce que l'échange de données informatisé ?

chiffrement

Processus informatique qui transforme des données en texte clair, lisibles par l'homme, en texte chiffré.

clé de chiffrement

Chaîne cryptographique de bits aléatoires générée par un algorithme cryptographique. La longueur des clés peut varier, et chaque clé est conçue pour être imprévisible et unique.

endianisme

Ordre dans lequel les octets sont stockés dans la mémoire de l'ordinateur. Big-endian les systèmes stockent d'abord l'octet le plus significatif. Little-endian les systèmes stockent d'abord l'octet le moins significatif.

point de terminaison

Voir point de terminaison de service.

service de point de terminaison

Service que vous pouvez héberger sur un cloud privé virtuel (VPC) pour le partager avec d'autres utilisateurs. Vous pouvez créer un service de point de terminaison avec AWS PrivateLink et accorder des autorisations à d'autres principaux Comptes AWS ou à Gestion des identités et des accès AWS (IAM) principaux. Ces comptes ou principaux peuvent se connecter à votre service de point de terminaison de manière privée en créant des points de terminaison d'un VPC d'interface. Pour plus d'informations, veuillez consulter Création d'un service de point de terminaison dans la documentation Amazon Virtual Private Cloud (Amazon VPC).

planification des ressources d'entreprise (ERP)

Système qui automatise et gère les principaux processus métier (tels que la comptabilité, le MES et la gestion de projet) pour une entreprise.

chiffrement d'enveloppe

Processus de chiffrement d'une clé de chiffrement à l'aide d'une autre clé de chiffrement. Pour plus d'informations, consultez la section Chiffrement des enveloppes dans la documentation AWS Key Management Service (AWS KMS).

environment

Instance d'une application en cours d'exécution. Les types d'environnement les plus courants dans le cloud computing sont les suivants :

- Environnement de développement : instance d'une application en cours d'exécution à laquelle seule l'équipe principale chargée de la maintenance de l'application peut accéder. Les environnements de développement sont utilisés pour tester les modifications avant de les promouvoir dans les environnements supérieurs. Ce type d'environnement est parfois appelé environnement de test.
- Environnements inférieurs : tous les environnements de développement d'une application, tels que ceux utilisés pour les générations et les tests initiaux.

- Environnement de production : instance d'une application en cours d'exécution à laquelle les utilisateurs finaux peuvent accéder. Dans un CI/CD pipeline, l'environnement de production est le dernier environnement de déploiement.
- Environnements supérieurs : tous les environnements accessibles aux utilisateurs autres que l'équipe de développement principale. Ils peuvent inclure un environnement de production, des environnements de préproduction et des environnements pour les tests d'acceptation par les utilisateurs.

épopée

Dans les méthodologies agiles, catégories fonctionnelles qui aident à organiser et à prioriser votre travail. Les épopées fournissent une description détaillée des exigences et des tâches d'implémentation. Par exemple, les points forts de la AWS CAF en matière de sécurité incluent la gestion des identités et des accès, les contrôles de détection, la sécurité des infrastructures, la protection des données et la réponse aux incidents. Pour plus d'informations sur les épopées dans la stratégie de migration AWS , veuillez consulter le guide d'implémentation du programme.

ERP

Voir Planification des ressources d'entreprise.

analyse exploratoire des données (EDA)

Processus d'analyse d'un jeu de données pour comprendre ses principales caractéristiques. Vous collectez ou agrégez des données, puis vous effectuez des enquêtes initiales pour trouver des modèles, détecter des anomalies et vérifier les hypothèses. L'EDA est réalisée en calculant des statistiques récapitulatives et en créant des visualisations de données.

F

tableau des faits

La table centrale dans un schéma en étoile. Il stocke des données quantitatives sur les opérations commerciales. Généralement, une table de faits contient deux types de colonnes : celles qui contiennent des mesures et celles qui contiennent une clé étrangère pour une table de dimensions.

échouer rapidement

Une philosophie qui utilise des tests fréquents et progressifs pour réduire le cycle de vie du développement. C'est un élément essentiel d'une approche agile.

limite d'isolation des défauts

Dans le AWS Cloud, une limite telle qu'une zone de disponibilité Région AWS, un plan de contrôle ou un plan de données qui limite l'effet d'une panne et contribue à améliorer la résilience des charges de travail. Pour plus d'informations, consultez la section Limites d'isolation des AWS pannes.

branche de fonctionnalités

Voir la succursale.

fonctionnalités

Les données d'entrée que vous utilisez pour faire une prédiction. Par exemple, dans un contexte de fabrication, les fonctionnalités peuvent être des images capturées périodiquement à partir de la ligne de fabrication.

importance des fonctionnalités

Le niveau d'importance d'une fonctionnalité pour les prédictions d'un modèle. Il s'exprime généralement sous la forme d'un score numérique qui peut être calculé à l'aide de différentes techniques, telles que la méthode Shapley Additive Explanations (SHAP) et les gradients intégrés. Pour plus d'informations, voir Interprétabilité du modèle d'apprentissage automatique avec AWS.

transformation de fonctionnalité

Optimiser les données pour le processus de ML, notamment en enrichissant les données avec des sources supplémentaires, en mettant à l'échelle les valeurs ou en extrayant plusieurs ensembles d'informations à partir d'un seul champ de données. Cela permet au modèle de ML de tirer parti des données. Par exemple, si vous décomposez la date « 2021-05-27 00:15:37 » en « 2021 », « mai », « jeudi » et « 15 », vous pouvez aider l'algorithme d'apprentissage à apprendre des modèles nuancés associés à différents composants de données.

invitation en quelques coups

Fournir à un LLM un petit nombre d'exemples illustrant la tâche et le résultat souhaité avant de lui demander d'effectuer une tâche similaire. Cette technique est une application de l'apprentissage contextuel, dans le cadre de laquelle les modèles apprennent à partir d'exemples (prises de vue) intégrés dans des instructions. Few-shot l'envoi d'instructions peut être efficace pour les tâches qui nécessitent un formatage, un raisonnement ou une connaissance du domaine spécifiques. Voir également l'invite Zero-Shot.

FGAC

Découvrez le contrôle d'accès détaillé.

contrôle d'accès détaillé (FGAC)

Utilisation de plusieurs conditions pour autoriser ou refuser une demande d'accès.

migration instantanée (flash-cut)

Méthode de migration de base de données qui utilise la réplication continue des données par le biais de la capture des données de modification afin de migrer les données dans les plus brefs délais, au lieu d'utiliser une approche progressive. L'objectif est de réduire au maximum les temps d'arrêt.

FM

Voir le modèle de fondation.

modèle de fondation (FM)

Un vaste réseau neuronal d'apprentissage profond qui s'entraîne sur des ensembles de données massifs de données généralisées et non étiquetées. Les FM sont capables d'effectuer une grande variété de tâches générales, telles que la compréhension du langage, la génération de texte et d'images et la conversation en langage naturel. Pour plus d'informations, voir Que sont les modèles de base ?

Passerelle FM

Un intermédiaire centralisé qui contrôle et normalise l'accès aux modèles de base. Également connue sous le nom de passerelle LLM.

G

IA générative

Sous-ensemble de modèles d'IA qui ont été entraînés sur de grandes quantités de données et qui peuvent utiliser une simple invite textuelle pour créer de nouveaux contenus et artefacts, tels que des images, des vidéos, du texte et du son. Pour plus d'informations, consultez Qu'est-ce que l'IA générative.

blocage géographique

Voir les restrictions géographiques.

restrictions géographiques (blocage géographique)

Sur Amazon CloudFront, option permettant d'empêcher les utilisateurs de certains pays d'accéder aux distributions de contenu. Vous pouvez utiliser une liste d'autorisation ou une liste de blocage

pour spécifier les pays approuvés et interdits. Pour plus d'informations, consultez la section Restreindre la distribution géographique de votre contenu dans la CloudFront documentation.

Flux de travail Gitflow

Approche dans laquelle les environnements inférieurs et supérieurs utilisent différentes branches dans un référentiel de code source. Le flux de travail Gitflow est considéré comme existant, et le flux de travail basé sur les troncs est l'approche moderne préférée.

image dorée

Un instantané d'un système ou d'un logiciel utilisé comme modèle pour déployer de nouvelles instances de ce système ou logiciel. Par exemple, dans le secteur de la fabrication, une image dorée peut être utilisée pour fournir des logiciels sur plusieurs appareils et contribue à améliorer la vitesse, l'évolutivité et la productivité des opérations de fabrication des appareils.

stratégie inédite

L'absence d'infrastructures existantes dans un nouvel environnement. Lorsque vous adoptez une stratégie inédite pour une architecture système, vous pouvez sélectionner toutes les nouvelles technologies sans restriction de compatibilité avec l'infrastructure existante, également appelée brownfield. Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et greenfield (inédites).

barrière de protection

Règle de haut niveau qui permet de régir les ressources, les politiques et la conformité au sein des unités d'organisation (UO). Les barrières de protection préventives appliquent des politiques pour garantir l'alignement sur les normes de conformité. Elles sont mises en œuvre à l'aide de politiques de contrôle des services et de limites des autorisations IAM. Les barrières de protection de détection détectent les violations des politiques et les problèmes de conformité, et génèrent des alertes pour y remédier. Ils sont implémentés à l'aide d'Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, d'Amazon Inspector et de AWS Lambda contrôles personnalisés.

rambardes (AI)

Des mécanismes de sécurité qui filtrent, valident et limitent les entrées et sorties des agents afin de garantir un comportement responsable et sûr de l'IA.

H

HA

Découvrez la haute disponibilité.

migration de base de données hétérogène

Migration de votre base de données source vers une base de données cible qui utilise un moteur de base de données différent (par exemple, Oracle vers Amazon Aurora). La migration hétérogène fait généralement partie d'un effort de réarchitecture, et la conversion du schéma peut s'avérer une tâche complexe. AWS propose AWS SCT qui facilite les conversions de schémas.

haute disponibilité (HA)

Capacité d'une charge de travail à fonctionner en continu, sans intervention, en cas de difficultés ou de catastrophes. Les systèmes HA sont conçus pour basculer automatiquement, fournir constamment des performances de haute qualité et gérer différentes charges et défaillances avec un impact minimal sur les performances.

modernisation des historiens

Approche utilisée pour moderniser et mettre à niveau les systèmes de technologie opérationnelle (OT) afin de mieux répondre aux besoins de l'industrie manufacturière. Un historien est un type de base de données utilisé pour collecter et stocker des données provenant de diverses sources dans une usine.

données de rétention

Partie de données historiques étiquetées qui n'est pas divulguée dans un ensemble de données utilisé pour entraîner un modèle d'apprentissage automatique. Vous pouvez utiliser les données de blocage pour évaluer les performances du modèle en comparant les prévisions du modèle aux données de blocage.

humain dans la boucle (HiTL)

Un modèle de flux de travail dans lequel l'exécution des agents s'arrête pour examen et approbation par l'homme aux points de décision critiques.

migration de base de données homogène

Migration de votre base de données source vers une base de données cible qui partage le même moteur de base de données (par exemple, Microsoft SQL Server vers Amazon RDS for SQL Server). La migration homogène s'inscrit généralement dans le cadre d'un effort de

réhébergement ou de replateforme. Vous pouvez utiliser les utilitaires de base de données natifs pour migrer le schéma.

données chaudes

Données fréquemment consultées, telles que les données en temps réel ou les données translationnelles récentes. Ces données nécessitent généralement un niveau ou une classe de stockage à hautes performances pour fournir des réponses rapides aux requêtes.

correctif

Solution d'urgence à un problème critique dans un environnement de production. En raison de son urgence, un correctif est généralement créé en dehors du flux de travail de DevOps publication habituel.

période de soins intensifs

Immédiatement après le basculement, période pendant laquelle une équipe de migration gère et surveille les applications migrées dans le cloud afin de résoudre les problèmes éventuels. En règle générale, cette période dure de 1 à 4 jours. À la fin de la période de soins intensifs, l'équipe de migration transfère généralement la responsabilité des applications à l'équipe des opérations cloud.

I

IaC

Considérez l'infrastructure comme un code.

politique basée sur l’identité

Politique attachée à un ou plusieurs principaux IAM qui définit leurs autorisations au sein de l' AWS Cloud environnement.

application inactive

Application dont l'utilisation moyenne du processeur et de la mémoire se situe entre 5 et 20 % sur une période de 90 jours. Dans un projet de migration, il est courant de retirer ces applications ou de les retenir sur site.

IIoT

Voir Internet industriel des objets.

infrastructure immuable

Modèle qui déploie une nouvelle infrastructure pour les charges de travail de production au lieu de mettre à jour, d'appliquer des correctifs ou de modifier l'infrastructure existante. Les infrastructures immuables sont intrinsèquement plus cohérentes, fiables et prévisibles que les infrastructures mutables. Pour plus d'informations, consultez les meilleures pratiques de déploiement à l'aide d'une infrastructure immuable dans le AWS Well-Architected Framework.

VPC entrant (d'entrée)

Dans une architecture AWS multi-comptes, un VPC qui accepte, inspecte et achemine les connexions réseau depuis l'extérieur d'une application. L'architecture de référence de sécuritéAWS recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

migration incrémentielle

Stratégie de basculement dans le cadre de laquelle vous migrez votre application par petites parties au lieu d'effectuer un basculement complet unique. Par exemple, il se peut que vous ne transfériez que quelques microservices ou utilisateurs vers le nouveau système dans un premier temps. Après avoir vérifié que tout fonctionne correctement, vous pouvez transférer progressivement des microservices ou des utilisateurs supplémentaires jusqu'à ce que vous puissiez mettre hors service votre système hérité. Cette stratégie réduit les risques associés aux migrations de grande ampleur.

Industry 4.0

Terme introduit par Klaus Schwab en 2016 pour désigner la modernisation des processus de fabrication grâce aux avancées en matière de connectivité, de données en temps réel, d'automatisation, d'analyse et. AI/ML

infrastructure

Ensemble des ressources et des actifs contenus dans l'environnement d'une application.

infrastructure en tant que code (IaC)

Processus de mise en service et de gestion de l'infrastructure d'une application via un ensemble de fichiers de configuration. IaC est conçue pour vous aider à centraliser la gestion de l'infrastructure, à normaliser les ressources et à mettre à l'échelle rapidement afin que les nouveaux environnements soient reproductibles, fiables et cohérents.

internet industriel des objets (IIoT)

L'utilisation de capteurs et d'appareils connectés à Internet dans les secteurs industriels tels que la fabrication, l'énergie, l'automobile, les soins de santé, les sciences de la vie et l'agriculture. Pour plus d'informations, veuillez consulter Building an industrial Internet of Things (IIoT) digital transformation strategy.

VPC d'inspection

Dans une architecture AWS multi-comptes, un VPC centralisé qui gère les inspections du trafic réseau entre les VPC (identiques ou Régions AWS différents), Internet et les réseaux sur site. L'architecture de référence de sécuritéAWS recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

Internet des objets (IoT)

Réseau d'objets physiques connectés dotés de capteurs ou de processeurs intégrés qui communiquent avec d'autres appareils et systèmes via Internet ou via un réseau de communication local. Pour plus d'informations, veuillez consulter la section Qu'est-ce que l'IoT ?.

interprétabilité

Caractéristique d'un modèle de machine learning qui décrit dans quelle mesure un être humain peut comprendre comment les prédictions du modèle dépendent de ses entrées. Pour plus d'informations, voir Interprétabilité du modèle d'apprentissage automatique avec AWS.

IoT

Voir Internet des objets.

Bibliothèque d'informations informatiques (ITIL)

Ensemble de bonnes pratiques pour proposer des services informatiques et les aligner sur les exigences métier. L'ITIL constitue la base de l'ITSM.

gestion des services informatiques (ITSM)

Activités associées à la conception, à la mise en œuvre, à la gestion et à la prise en charge de services informatiques d'une organisation. Pour plus d'informations sur l'intégration des opérations cloud aux outils ITSM, veuillez consulter le guide d'intégration des opérations.

ITIL

Consultez la bibliothèque d'informations informatiques.

ITSM

Voir Gestion des services informatiques.

L

contrôle d'accès basé sur des étiquettes (LBAC)

Une implémentation du contrôle d'accès obligatoire (MAC) dans laquelle une valeur d'étiquette de sécurité est explicitement attribuée aux utilisateurs et aux données elles-mêmes. L'intersection entre l'étiquette de sécurité utilisateur et l'étiquette de sécurité des données détermine les lignes et les colonnes visibles par l'utilisateur.

zone de destination

Une zone d'atterrissage est un AWS environnement multi-comptes bien conçu, évolutif et sécurisé. Il s'agit d'un point de départ à partir duquel vos entreprises peuvent rapidement lancer et déployer des charges de travail et des applications en toute confiance dans leur environnement de sécurité et d'infrastructure. Pour plus d'informations sur les zones de destination, veuillez consulter Setting up a secure and scalable multi-account AWS environment.

grand modèle de langage (LLM)

Un modèle d'intelligence artificielle basé sur le deep learning qui est préentraîné sur une grande quantité de données. Un LLM peut effectuer plusieurs tâches, telles que répondre à des questions, résumer des documents, traduire du texte dans d'autres langues et compléter des phrases. Pour plus d'informations, voir Que sont les LLM.

migration de grande envergure

Migration de 300 serveurs ou plus.

LBAC

Voir contrôle d'accès basé sur des étiquettes.

principe de moindre privilège

Bonne pratique de sécurité qui consiste à accorder les autorisations minimales nécessaires à l'exécution d'une tâche. Pour plus d'informations, veuillez consulter la rubrique Accorder les autorisations de moindre privilège dans la documentation IAM.

lift and shift

Voir 7 Rs.

système de poids faible

Système qui stocke d'abord l'octet le moins significatif. Voir aussi endianité.

LLM

Voir le grand modèle de langage.

environnements inférieurs

Voir environnement.

M

machine learning (ML)

Type d'intelligence artificielle qui utilise des algorithmes et des techniques pour la reconnaissance et l'apprentissage de modèles. Le ML analyse et apprend à partir de données enregistrées, telles que les données de l'Internet des objets (IoT), pour générer un modèle statistique basé sur des modèles. Pour plus d'informations, veuillez consulter Machine Learning.

branche principale

Voir la succursale.

malware

Logiciel conçu pour compromettre la sécurité ou la confidentialité de l'ordinateur. Les logiciels malveillants peuvent perturber les systèmes informatiques, divulguer des informations sensibles ou obtenir un accès non autorisé. Parmi les malwares, on peut citer les virus, les vers, les rançongiciels, les chevaux de Troie, les logiciels espions et les enregistreurs de frappe.

services gérés

Services AWS pour lequel AWS fonctionnent la couche d'infrastructure, le système d'exploitation et les plateformes, et vous accédez aux points de terminaison pour stocker et récupérer des données. Amazon Simple Storage Service (Amazon S3) et Amazon DynamoDB sont des exemples de services gérés. Ils sont également connus sous le nom de services abstraits.

système d'exécution de la fabrication (MES)

Un système logiciel pour le suivi, la surveillance, la documentation et le contrôle des processus de production qui convertissent les matières premières en produits finis dans l'atelier.

MAP

Voir Migration Acceleration Program.

MCP

Voir Model Context Protocol.

Protocole de contexte du modèle (MCP)

Protocole sans état pour la communication entre un agent et un outil.

serveur MCP

Service qui expose un ou plusieurs outils via le protocole Model Context.

mécanisme

Processus complet au cours duquel vous créez un outil, favorisez son adoption, puis inspectez les résultats afin de procéder aux ajustements nécessaires. Un mécanisme est un cycle qui se renforce et s'améliore au fur et à mesure de son fonctionnement. Pour plus d'informations, voir Création de mécanismes dans le AWS Well-Architected cadre.

compte membre

Tous, à l' Comptes AWS exception du compte de gestion, qui font partie d'une organisation dans AWS Organizations. Un compte ne peut être membre que d'une seule organisation à la fois.

MAILLES

Voir le système d'exécution de la fabrication.

Transport télémétrique en file d'attente de messages (MQTT)

Un protocole de communication léger de machine à machine (M2M), basé sur le publish/subscribemodèle, pour les appareils IoT aux ressources limitées.

microservice

Petit service indépendant qui communique via des API bien définies et qui est généralement détenu par de petites équipes autonomes. Par exemple, un système d'assurance peut inclure des microservices qui mappent à des capacités métier, telles que les ventes ou le marketing, ou à des sous-domaines, tels que les achats, les réclamations ou l'analytique. Les avantages des microservices incluent l'agilité, la flexibilité de la mise à l'échelle, la facilité de déploiement, la réutilisation du code et la résilience. Pour plus d'informations, consultez la section Intégration de microservices à l'aide de services AWS sans serveur.

architecture de microservices

Approche de création d'une application avec des composants indépendants qui exécutent chaque processus d'application en tant que microservice. Ces microservices communiquent via une interface bien définie à l'aide d'API légères. Chaque microservice de cette architecture peut être mis à jour, déployé et mis à l'échelle pour répondre à la demande de fonctions spécifiques d'une application. Pour plus d'informations, consultez la section Implémentation de microservices sur AWS.

Programme d'accélération des migrations (MAP)

Un AWS programme qui fournit un support de conseil, des formations et des services pour aider les entreprises à établir une base opérationnelle solide pour passer au cloud, et pour aider à compenser le coût initial des migrations. MAP inclut une méthodologie de migration pour exécuter les migrations héritées de manière méthodique, ainsi qu'un ensemble d'outils pour automatiser et accélérer les scénarios de migration courants.

migration à grande échelle

Processus consistant à transférer la majeure partie du portefeuille d'applications vers le cloud par vagues, un plus grand nombre d'applications étant déplacées plus rapidement à chaque vague. Cette phase utilise les bonnes pratiques et les enseignements tirés des phases précédentes pour implémenter une usine de migration d'équipes, d'outils et de processus en vue de rationaliser la migration des charges de travail grâce à l'automatisation et à la livraison agile. Il s'agit de la troisième phase de la stratégie de migration AWS.

usine de migration

Cross-functional des équipes qui rationalisent la migration des charges de travail grâce à des approches automatisées et agiles. Les équipes de Migration Factory comprennent généralement des responsables des opérations, des analystes commerciaux et des propriétaires, des ingénieurs de migration, des développeurs et DevOps des professionnels travaillant dans le cadre de sprints. Entre 20 et 50 % du portefeuille d'applications d'entreprise est constitué de modèles répétés qui peuvent être optimisés par une approche d'usine. Pour plus d'informations, veuillez consulter la rubrique discussion of migration factories et le guide Cloud Migration Factory dans cet ensemble de contenus.

métadonnées de migration

Informations relatives à l'application et au serveur nécessaires pour finaliser la migration. Chaque modèle de migration nécessite un ensemble de métadonnées de migration différent. Les

exemples de métadonnées de migration incluent le sous-réseau cible, le groupe de sécurité et le AWS compte.

modèle de migration

Tâche de migration reproductible qui détaille la stratégie de migration, la destination de la migration et l'application ou le service de migration utilisé. Exemple : réorganisez la migration vers Amazon EC2 AWS avec le service de migration d'applications.

Évaluation du portefeuille de migration (MPA)

Outil en ligne qui fournit des informations pour valider l'analyse de rentabilisation en faveur de la migration vers le. AWS Cloud La MPA propose une évaluation détaillée du portefeuille (dimensionnement approprié des serveurs, tarification, comparaison du coût total de possession, analyse des coûts de migration), ainsi que la planification de la migration (analyse et collecte des données d'applications, regroupement des applications, priorisation des migrations et planification des vagues). L'outil MPA (connexion requise) est disponible gratuitement pour tous les AWS consultants et consultants APN Partner.

Évaluation de la préparation à la migration (MRA)

Processus qui consiste à obtenir des informations sur l'état de préparation d'une organisation au cloud, à identifier les forces et les faiblesses et à élaborer un plan d'action pour combler les lacunes identifiées, à l'aide du AWS CAF. Pour plus d'informations, veuillez consulter le guide de préparation à la migration. La MRA est la première phase de la stratégie de migration AWS.

stratégie de migration

L'approche utilisée pour migrer une charge de travail vers le AWS Cloud. Pour plus d'informations, reportez-vous aux 7 R de ce glossaire et à Mobiliser votre organisation pour accélérer les migrations à grande échelle.

ML

Voir apprentissage automatique.

modernisation

Transformation d'une application obsolète (héritée ou monolithique) et de son infrastructure en un système agile, élastique et hautement disponible dans le cloud afin de réduire les coûts, de gagner en efficacité et de tirer parti des innovations. Pour plus d'informations, consultez la section Stratégie de modernisation des applications dans le AWS Cloud.

évaluation de la préparation à la modernisation

Évaluation qui permet de déterminer si les applications d'une organisation sont prêtes à être modernisées, d'identifier les avantages, les risques et les dépendances, et qui détermine dans quelle mesure l'organisation peut prendre en charge l'état futur de ces applications. Le résultat de l'évaluation est un plan de l'architecture cible, une feuille de route détaillant les phases de développement et les étapes du processus de modernisation, ainsi qu'un plan d'action pour combler les lacunes identifiées. Pour plus d'informations, consultez la section Évaluation de l'état de préparation à la modernisation des applications dans le AWS Cloud.

applications monolithiques (monolithes)

Applications qui s'exécutent en tant que service unique avec des processus étroitement couplés. Les applications monolithiques ont plusieurs inconvénients. Si une fonctionnalité de l'application connaît un pic de demande, l'architecture entière doit être mise à l'échelle. L'ajout ou l'amélioration des fonctionnalités d'une application monolithique devient également plus complexe lorsque la base de code s'élargit. Pour résoudre ces problèmes, vous pouvez utiliser une architecture de microservices. Pour plus d'informations, veuillez consulter Decomposing monoliths into microservices.

MPA

Voir Évaluation du portefeuille de migration.

MQTT

Voir Message Queuing Telemetry Transport.

classification multi-classes

Processus qui permet de générer des prédictions pour plusieurs classes (prédiction d'un résultat parmi plus de deux). Par exemple, un modèle de ML peut demander « Ce produit est-il un livre, une voiture ou un téléphone ? » ou « Quelle catégorie de produits intéresse le plus ce client ? ».

infrastructure mutable

Modèle qui met à jour et modifie l'infrastructure existante pour les charges de travail de production. Pour améliorer la cohérence, la fiabilité et la prévisibilité, le AWS Well-Architected Framework recommande l'utilisation d'une infrastructure immuable comme meilleure pratique.

O

OAC

Voir Contrôle d'accès à l'origine.

OAI

Voir l'identité d'accès à l'origine.

OCM

Voir gestion du changement organisationnel.

migration hors ligne

Méthode de migration dans laquelle la charge de travail source est supprimée au cours du processus de migration. Cette méthode implique un temps d'arrêt prolongé et est généralement utilisée pour de petites charges de travail non critiques.

OI

Consultez la section Intégration des opérations.

OLA

Voir l'accord au niveau opérationnel.

migration en ligne

Méthode de migration dans laquelle la charge de travail source est copiée sur le système cible sans être mise hors ligne. Les applications connectées à la charge de travail peuvent continuer à fonctionner pendant la migration. Cette méthode implique un temps d'arrêt nul ou minimal et est généralement utilisée pour les charges de travail de production critiques.

OPC-UA

Voir Open Process Communications - Architecture unifiée.

Communications par processus ouvert - Architecture unifiée (OPC-UA)

Protocole de communication machine à machine (M2M) pour l'automatisation industrielle. OPC-UA fournit une norme d'interopérabilité avec des schémas de chiffrement, d'authentification et d'autorisation des données.

accord au niveau opérationnel (OLA)

Accord qui précise ce que les groupes informatiques fonctionnels s'engagent à fournir les uns aux autres, afin de prendre en charge un contrat de niveau de service (SLA).

examen de l'état de préparation opérationnelle (ORR)

Une liste de questions et de bonnes pratiques associées qui vous aident à comprendre, à évaluer, à prévenir ou à réduire l'ampleur des incidents et des défaillances possibles. Pour plus d'informations, voir Examens de l'état de préparation opérationnelle (ORR) dans le AWS Well-Architected cadre.

technologie opérationnelle (OT)

Systèmes matériels et logiciels qui fonctionnent avec l'environnement physique pour contrôler les opérations, les équipements et les infrastructures industriels. Dans le secteur manufacturier, l'intégration des systèmes OT et des technologies de l'information (IT) est au cœur des transformations de l'industrie 4.0.

intégration des opérations (OI)

Processus de modernisation des opérations dans le cloud, qui implique la planification de la préparation, l'automatisation et l'intégration. Pour en savoir plus, veuillez consulter le guide d'intégration des opérations.

journal de suivi d'organisation

Un parcours créé par AWS CloudTrail qui enregistre tous les événements pour tous les membres Comptes AWS d'une organisation dans AWS Organizations. Ce journal de suivi est créé dans chaque Compte AWS qui fait partie de l'organisation et suit l'activité de chaque compte. Pour plus d'informations, consultez la section Création d'un suivi pour une organisation dans la CloudTrail documentation.

gestion du changement organisationnel (OCM)

Cadre pour gérer les transformations métier majeures et perturbatrices du point de vue des personnes, de la culture et du leadership. L'OCM aide les organisations à se préparer et à effectuer la transition vers de nouveaux systèmes et de nouvelles politiques en accélérant l'adoption des changements, en abordant les problèmes de transition et en favorisant des changements culturels et organisationnels. Dans la stratégie de AWS migration, ce cadre est appelé accélération du personnel, en raison de la rapidité du changement requise dans les projets d'adoption du cloud. Pour plus d'informations, veuillez consulter le guide OCM.

contrôle d'accès d'origine (OAC)

Dans CloudFront, une option améliorée pour restreindre l'accès afin de sécuriser votre contenu Amazon Simple Storage Service (Amazon S3). OAC prend en charge tous les compartiments S3 dans leur ensemble Régions AWS, le chiffrement côté serveur avec AWS KMS (SSE-KMS) et les `DELETE` requêtes dynamiques `PUT` adressées au compartiment S3.

identité d'accès d'origine (OAI)

Dans CloudFront, une option permettant de restreindre l'accès afin de sécuriser votre contenu Amazon S3. Lorsque vous utilisez OAI, il CloudFront crée un principal auprès duquel Amazon S3 peut s'authentifier. Les principaux authentifiés ne peuvent accéder au contenu d'un compartiment S3 que par le biais d'une distribution spécifique CloudFront . Voir également OAC, qui fournit un contrôle d'accès plus précis et amélioré.

ORR

Voir l'examen de l'état de préparation opérationnelle.

DE

Voir technologie opérationnelle.

VPC sortant (de sortie)

Dans une architecture AWS multi-comptes, un VPC qui gère les connexions réseau initiées depuis une application. L'architecture de référence de sécuritéAWS recommande de configurer votre compte réseau avec des VPC entrants, sortants et d'inspection afin de protéger l'interface bidirectionnelle entre votre application et Internet en général.

P

limite des autorisations

Politique de gestion IAM attachée aux principaux IAM pour définir les autorisations maximales que peut avoir l'utilisateur ou le rôle. Pour plus d'informations, veuillez consulter la rubrique Limites des autorisations dans la documentation IAM.

informations personnelles identifiables (PII)

Informations qui, lorsqu'elles sont consultées directement ou associées à d'autres données connexes, peuvent être utilisées pour déduire raisonnablement l'identité d'une personne. Les

exemples d'informations personnelles incluent les noms, les adresses et les informations de contact.

PII

Voir les informations personnelles identifiables.

manuel stratégique

Ensemble d'étapes prédéfinies qui capturent le travail associé aux migrations, comme la fourniture de fonctions d'opérations de base dans le cloud. Un manuel stratégique peut revêtir la forme de scripts, de runbooks automatisés ou d'un résumé des processus ou des étapes nécessaires au fonctionnement de votre environnement modernisé.

PLC

Voir contrôleur logique programmable.

PLM

Consultez la section Gestion du cycle de vie des produits.

policy

Objet capable de définir les autorisations (voir la politique basée sur l'identité), de spécifier les conditions d'accès (voir la politique basée sur les ressources) ou de définir les autorisations maximales pour tous les comptes d'une organisation dans AWS Organizations (voir la politique de contrôle des services).

persistance polyglotte

Choix indépendant de la technologie de stockage de données d'un microservice en fonction des modèles d'accès aux données et d'autres exigences. Si vos microservices utilisent la même technologie de stockage de données, ils peuvent rencontrer des difficultés d'implémentation ou présenter des performances médiocres. Les microservices sont plus faciles à mettre en œuvre, atteignent de meilleures performances, ainsi qu'une meilleure capacité de mise à l'échelle s'ils utilisent l'entrepôt de données le mieux adapté à leurs besoins.

évaluation du portefeuille

Processus de découverte, d'analyse et de priorisation du portefeuille d'applications afin de planifier la migration. Pour plus d'informations, veuillez consulter Evaluating migration readiness.

predicate

Une condition de requête qui renvoie `true` ou`false`, généralement située dans une `WHERE` clause.

prédicat pushdown

Technique d'optimisation des requêtes de base de données qui filtre les données de la requête avant le transfert. Cela réduit la quantité de données qui doivent être extraites et traitées à partir de la base de données relationnelle et améliore les performances des requêtes.

contrôle préventif

Contrôle de sécurité conçu pour empêcher qu'un événement ne se produise. Ces contrôles constituent une première ligne de défense pour empêcher tout accès non autorisé ou toute modification indésirable de votre réseau. Pour plus d'informations, veuillez consulter Preventative controls dans Implementing security controls on AWS.

principal

Entité capable d'effectuer AWS des actions et d'accéder à des ressources. Cette entité est généralement un utilisateur root pour un Compte AWS rôle IAM ou un utilisateur. Pour plus d'informations, veuillez consulter la rubrique Principal dans Termes et concepts relatifs aux rôles, dans la documentation IAM.

confidentialité dès la conception

Une approche d'ingénierie système qui prend en compte la confidentialité tout au long du processus de développement.

zones hébergées privées

Conteneur qui contient des informations concernant la façon dont vous souhaitez qu'Amazon Route 53 réponde aux requêtes DNS pour un domaine et ses sous-domaines dans un ou plusieurs VPC. Pour plus d'informations, veuillez consulter Working with private hosted zones dans la documentation Route 53.

contrôle proactif

Contrôle de sécurité conçu pour empêcher le déploiement de ressources non conformes. Ces contrôles analysent les ressources avant qu'elles ne soient provisionnées. Si la ressource n'est pas conforme au contrôle, elle n'est pas provisionnée. Pour plus d'informations, consultez le guide de référence sur les contrôles dans la AWS Control Tower documentation et consultez la section Contrôles proactifs dans Implémentation des contrôles de sécurité sur AWS.

gestion du cycle de vie des produits (PLM)

Gestion des données et des processus d'un produit tout au long de son cycle de vie, depuis la conception, le développement et le lancement, en passant par la croissance et la maturité, jusqu'au déclin et au retrait.

environnement de production

Voir environnement.

contrôleur logique programmable (PLC)

Dans le secteur manufacturier, un ordinateur hautement fiable et adaptable qui surveille les machines et automatise les processus de fabrication.

chaînage rapide

Utiliser le résultat d'une invite LLM comme entrée pour l'invite suivante afin de générer de meilleures réponses. Cette technique est utilisée pour décomposer une tâche complexe en sous-tâches ou pour affiner ou développer de manière itérative une réponse préliminaire. Cela permet d'améliorer la précision et la pertinence des réponses d'un modèle et permet d'obtenir des résultats plus précis et personnalisés.

pseudonymisation

Processus de remplacement des identifiants personnels dans un ensemble de données par des valeurs fictives. La pseudonymisation peut contribuer à protéger la vie privée. Les données pseudonymisées sont toujours considérées comme des données personnelles.

publish/subscribe (pub/sub)

Modèle qui permet des communications asynchrones entre les microservices afin d'améliorer l'évolutivité et la réactivité. Par exemple, dans un MES basé sur des microservices, un microservice peut publier des messages d'événements sur un canal auquel d'autres microservices peuvent s'abonner. Le système peut ajouter de nouveaux microservices sans modifier le service de publication.

Q

plan de requête

Série d'étapes, telles que des instructions, utilisées pour accéder aux données d'un système de base de données relationnelle SQL.

régression du plan de requêtes

Le cas où un optimiseur de service de base de données choisit un plan moins optimal qu'avant une modification donnée de l'environnement de base de données. Cela peut être dû à des

changements en termes de statistiques, de contraintes, de paramètres d'environnement, de liaisons de paramètres de requêtes et de mises à jour du moteur de base de données.

R

Matrice RACI

Voir responsable, responsable, consulté, informé (RACI).

RAG

Voir Retrieval Augmented Generation.

rançongiciel

Logiciel malveillant conçu pour bloquer l'accès à un système informatique ou à des données jusqu'à ce qu'un paiement soit effectué.

Matrice RASCI

Voir responsable, responsable, consulté, informé (RACI).

RCAC

Voir contrôle d'accès aux lignes et aux colonnes.

réplica en lecture

Copie d'une base de données utilisée en lecture seule. Vous pouvez acheminer les requêtes vers le réplica de lecture pour réduire la charge sur votre base de données principale.

réarchitecte

Voir 7 Rs.

objectif de point de récupération (RPO)

Durée maximale acceptable depuis le dernier point de récupération des données. Il détermine ce qui est considéré comme étant une perte de données acceptable entre le dernier point de reprise et l'interruption du service.

objectif de temps de récupération (RTO)

Le délai maximum acceptable entre l'interruption du service et le rétablissement du service.

refactoriser

Voir 7 Rs.

Région

Un ensemble de AWS ressources dans une zone géographique. Chacune Région AWS est isolée et indépendante des autres pour garantir tolérance aux pannes, stabilité et résilience. Pour plus d'informations, voir Spécifier ce que Régions AWS votre compte peut utiliser.

régression

Technique de ML qui prédit une valeur numérique. Par exemple, pour résoudre le problème « Quel sera le prix de vente de cette maison ? », un modèle de ML pourrait utiliser un modèle de régression linéaire pour prédire le prix de vente d'une maison sur la base de faits connus à son sujet (par exemple, la superficie en mètres carrés).

réhéberger

Voir 7 Rs.

version

Dans un processus de déploiement, action visant à promouvoir les modifications apportées à un environnement de production.

déplacer

Voir 7 Rs.

replateforme

Voir 7 Rs.

rachat

Voir 7 Rs.

résilience

La capacité d'une application à résister aux perturbations ou à s'en remettre. La haute disponibilité et la reprise après sinistre sont des considérations courantes lors de la planification de la résilience dans le AWS Cloud. Pour plus d'informations, consultez la section AWS Cloud Résilience.

politique basée sur les ressources

Politique attachée à une ressource, comme un compartiment Amazon S3, un point de terminaison ou une clé de chiffrement. Ce type de politique précise les principaux auxquels l'accès est autorisé, les actions prises en charge et toutes les autres conditions qui doivent être remplies.

matrice responsable, redevable, consulté et informé (RACI)

Une matrice qui définit les rôles et les responsabilités de toutes les parties impliquées dans les activités de migration et les opérations cloud. Le nom de la matrice est dérivé des types de responsabilité définis dans la matrice : responsable (R), responsable (A), consulté (C) et informé (I). Le type de support (S) est facultatif. Si vous incluez le support, la matrice est appelée matrice RASCI, et si vous l'excluez, elle est appelée matrice RACI.

contrôle réactif

Contrôle de sécurité conçu pour permettre de remédier aux événements indésirables ou aux écarts par rapport à votre référence de sécurité. Pour plus d'informations, veuillez consulter la rubrique Responsive controls dans Implementing security controls on AWS.

retain

Voir 7 Rs.

se retirer

Voir 7 Rs.

Génération augmentée de récupération (RAG)

Technologie d'IA générative dans laquelle un LLM fait référence à une source de données faisant autorité qui se trouve en dehors de ses sources de données de formation avant de générer une réponse. Par exemple, un modèle RAG peut effectuer une recherche sémantique dans la base de connaissances ou dans les données personnalisées d'une organisation. Pour plus d'informations, voir Qu'est-ce que RAG ?

rotation

Processus de mise à jour périodique d'un secret pour empêcher un attaquant d'accéder aux informations d'identification.

contrôle d'accès aux lignes et aux colonnes (RCAC)

Utilisation d'expressions SQL simples et flexibles dotées de règles d'accès définies. Le RCAC comprend des autorisations de ligne et des masques de colonnes.

RPO

Voir l'objectif du point de récupération.

RTO

Voir l'objectif en matière de temps de rétablissement.

runbook

Ensemble de procédures manuelles ou automatisées nécessaires à l'exécution d'une tâche spécifique. Elles visent généralement à rationaliser les opérations ou les procédures répétitives présentant des taux d'erreur élevés.

S

SAML 2.0

Un standard ouvert utilisé par de nombreux fournisseurs d'identité (IdPs). Cette fonctionnalité permet l'authentification unique fédérée (SSO), afin que les utilisateurs puissent se connecter AWS Management Console ou appeler les opérations de l' AWS API sans que vous ayez à créer un utilisateur dans IAM pour tous les membres de votre organisation. Pour plus d'informations sur la fédération SAML 2.0, veuillez consulter À propos de la fédération SAML 2.0 dans la documentation IAM.

SCADA

Voir Contrôle de supervision et acquisition de données.

SCP

Voir la politique de contrôle des services.

secret

Dans AWS Secrets Manager des informations confidentielles ou restreintes, telles qu'un mot de passe ou des informations d'identification utilisateur, que vous stockez sous forme cryptée. Il comprend la valeur secrète et ses métadonnées. La valeur secrète peut être binaire, une chaîne unique ou plusieurs chaînes. Pour plus d'informations, voir Que contient le secret d'un Secrets Manager ? dans la documentation de Secrets Manager.

sécurité dès la conception

Une approche d'ingénierie système qui prend en compte la sécurité tout au long du processus de développement.

contrôle de sécurité

Barrière de protection technique ou administrative qui empêche, détecte ou réduit la capacité d'un assaillant d'exploiter une vulnérabilité de sécurité. Il existe quatre principaux types de contrôles de sécurité : préventifs, détectifs, réactifs et proactifs.

renforcement de la sécurité

Processus qui consiste à réduire la surface d'attaque pour la rendre plus résistante aux attaques. Cela peut inclure des actions telles que la suppression de ressources qui ne sont plus requises, la mise en œuvre des bonnes pratiques de sécurité consistant à accorder le moindre privilège ou la désactivation de fonctionnalités inutiles dans les fichiers de configuration.

système de gestion des informations et des événements de sécurité (SIEM)

Outils et services qui associent les systèmes de gestion des informations de sécurité (SIM) et de gestion des événements de sécurité (SEM). Un système SIEM collecte, surveille et analyse les données provenant de serveurs, de réseaux, d'appareils et d'autres sources afin de détecter les menaces et les failles de sécurité, mais aussi de générer des alertes.

automatisation des réponses de sécurité

Action prédéfinie et programmée conçue pour répondre automatiquement à un événement de sécurité ou y remédier. Ces automatisations servent de contrôles de sécurité détectifs ou réactifs qui vous aident à mettre en œuvre les meilleures pratiques en matière AWS de sécurité. Parmi les actions de réponse automatique, citons la modification d'un groupe de sécurité VPC, l'application de correctifs à une instance Amazon EC2 ou la rotation des informations d'identification.

chiffrement côté serveur

Chiffrement des données à destination, par celui Service AWS qui les reçoit.

Politique de contrôle des services (SCP)

Politique qui propose un contrôle centralisé des autorisations pour tous les comptes d'une organisation dans AWS Organizations. Les SCP définissent des barrières de protection ou des limites aux actions qu'un administrateur peut déléguer à des utilisateurs ou à des rôles. Vous pouvez utiliser les SCP comme listes d'autorisation ou de refus, pour indiquer les services ou les actions autorisés ou interdits. Pour plus d'informations, consultez la section Politiques de contrôle des services dans la AWS Organizations documentation.

point de terminaison du service

URL du point d'entrée pour un Service AWS. Pour vous connecter par programmation au service cible, vous pouvez utiliser un point de terminaison. Pour plus d'informations, veuillez consulter la rubrique Service AWS endpoints dans Références générales AWS.

contrat de niveau de service (SLA)

Accord qui précise ce qu'une équipe informatique promet de fournir à ses clients, comme le temps de disponibilité et les performances des services.

indicateur de niveau de service (SLI)

Mesure d'un aspect des performances d'un service, tel que son taux d'erreur, sa disponibilité ou son débit.

objectif de niveau de service (SLO)

Mesure cible qui représente l'état d'un service, tel que mesuré par un indicateur de niveau de service.

modèle de responsabilité partagée

Un modèle décrivant la responsabilité que vous partagez en matière AWS de sécurité et de conformité dans le cloud. AWS est responsable de la sécurité du cloud, alors que vous êtes responsable de la sécurité dans le cloud. Pour de plus amples informations, veuillez consulter Modèle de responsabilité partagée.

IA de l'ombre

Applications d'IA non autorisées créées ou utilisées en dehors des canaux régis au sein d'une organisation.

SIEM

Consultez les informations de sécurité et le système de gestion des événements.

point de défaillance unique (SPOF)

Défaillance d'un seul composant critique d'une application susceptible de perturber le système.

SLA

Voir le contrat de niveau de service.

SLI

Voir l'indicateur de niveau de service.

SLO

Voir l'objectif de niveau de service.

modèle split-and-seed

Modèle permettant de mettre à l'échelle et d'accélérer les projets de modernisation. Au fur et à mesure que les nouvelles fonctionnalités et les nouvelles versions de produits sont définies, l'équipe principale se divise pour créer des équipes de produit. Cela permet de mettre à l'échelle

les capacités et les services de votre organisation, d'améliorer la productivité des développeurs et de favoriser une innovation rapide. Pour plus d'informations, consultez la section Approche progressive de la modernisation des applications dans le. AWS Cloud

SPOF

Voir point de défaillance unique.

schéma en étoile

Structure organisationnelle de base de données qui utilise une grande table de faits pour stocker les données transactionnelles ou mesurées et utilise une ou plusieurs tables dimensionnelles plus petites pour stocker les attributs des données. Cette structure est conçue pour être utilisée dans un entrepôt de données ou à des fins de business intelligence.

modèle de figuier étrangleur

Approche de modernisation des systèmes monolithiques en réécrivant et en remplaçant progressivement les fonctionnalités du système jusqu'à ce que le système hérité puisse être mis hors service. Ce modèle utilise l'analogie d'un figuier de vigne qui se développe dans un arbre existant et qui finit par supplanter son hôte. Le schéma a été présenté par Martin Fowler comme un moyen de gérer les risques lors de la réécriture de systèmes monolithiques. Pour un exemple d'application de ce modèle, consultez la section Modernisation progressive des anciens services Web Microsoft ASP.NET (ASMX) à l'aide de conteneurs et d'Amazon API Gateway.

sous-réseau

Plage d'adresses IP dans votre VPC. Un sous-réseau doit se trouver dans une seule zone de disponibilité.

contrôle de supervision et acquisition de données (SCADA)

Dans le secteur manufacturier, un système qui utilise du matériel et des logiciels pour surveiller les actifs physiques et les opérations de production.

chiffrement symétrique

Algorithme de chiffrement qui utilise la même clé pour chiffrer et déchiffrer les données.

tests synthétiques

Tester un système de manière à simuler les interactions des utilisateurs afin de détecter les problèmes potentiels ou de surveiller les performances. Vous pouvez utiliser Amazon CloudWatch Synthetics pour créer ces tests.

invite du système

Technique permettant de fournir un contexte, des instructions ou des directives à un LLM afin d'orienter son comportement. Les instructions du système aident à définir le contexte et à établir des règles pour les interactions avec les utilisateurs.

T

tags

Key-value des paires qui agissent comme des métadonnées pour organiser vos AWS ressources. Les balises peuvent vous aider à gérer, identifier, organiser, rechercher et filtrer des ressources. Pour plus d'informations, veuillez consulter la rubrique Balisage de vos AWS ressources.

variable cible

La valeur que vous essayez de prédire dans le cadre du ML supervisé. Elle est également qualifiée de variable de résultat. Par exemple, dans un environnement de fabrication, la variable cible peut être un défaut du produit.

liste de tâches

Outil utilisé pour suivre les progrès dans un runbook. Liste de tâches qui contient une vue d'ensemble du runbook et une liste des tâches générales à effectuer. Pour chaque tâche générale, elle inclut le temps estimé nécessaire, le propriétaire et l'avancement.

environnement de test

Voir environnement.

entraînement

Pour fournir des données à partir desquelles votre modèle de ML peut apprendre. Les données d'entraînement doivent contenir la bonne réponse. L'algorithme d'apprentissage identifie des modèles dans les données d'entraînement, qui mettent en correspondance les attributs des données d'entrée avec la cible (la réponse que vous souhaitez prédire). Il fournit un modèle de ML qui capture ces modèles. Vous pouvez alors utiliser le modèle de ML pour obtenir des prédictions sur de nouvelles données pour lesquelles vous ne connaissez pas la cible.

outil

Fonction ou API qu'un agent peut invoquer pour effectuer des opérations dans des systèmes externes.

passerelle de transit

Hub de transit de réseau que vous pouvez utiliser pour relier vos VPC et vos réseaux sur site. Pour plus d'informations, voir Qu'est-ce qu'une passerelle de transit dans la AWS Transit Gateway documentation.

flux de travail basé sur jonction

Approche selon laquelle les développeurs génèrent et testent des fonctionnalités localement dans une branche de fonctionnalités, puis fusionnent ces modifications dans la branche principale. La branche principale est ensuite intégrée aux environnements de développement, de préproduction et de production, de manière séquentielle.

accès sécurisé

Accorder des autorisations à un service que vous spécifiez pour effectuer des tâches au sein de votre organisation AWS Organizations et dans ses comptes en votre nom. Le service de confiance crée un rôle lié au service dans chaque compte, lorsque ce rôle est nécessaire, pour effectuer des tâches de gestion à votre place. Pour plus d'informations, consultez la section Utilisation AWS Organizations avec d'autres AWS services dans la AWS Organizations documentation.

réglage

Pour modifier certains aspects de votre processus d'entraînement afin d'améliorer la précision du modèle de ML. Par exemple, vous pouvez entraîner le modèle de ML en générant un ensemble d'étiquetage, en ajoutant des étiquettes, puis en répétant ces étapes plusieurs fois avec différents paramètres pour optimiser le modèle.

équipe de deux pizzas

Une petite DevOps équipe que vous pouvez nourrir avec deux pizzas. Une équipe de deux pizzas garantit les meilleures opportunités de collaboration possible dans le développement de logiciels.

U

incertitude

Un concept qui fait référence à des informations imprécises, incomplètes ou inconnues susceptibles de compromettre la fiabilité des modèles de ML prédictifs. Il existe deux types d'incertitude : l'incertitude épistémique est causée par des données limitées et incomplètes, alors que l'incertitude aléatoire est causée par le bruit et le caractère aléatoire inhérents aux données.

tâches indifférenciées

Également connu sous le nom de « levage de charges lourdes », ce travail est nécessaire pour créer et exploiter une application, mais qui n'apporte pas de valeur directe à l'utilisateur final ni d'avantage concurrentiel. Les exemples de tâches indifférenciées incluent l'approvisionnement, la maintenance et la planification des capacités.

environnements supérieurs

Voir environnement.

V

mise à vide

Opération de maintenance de base de données qui implique un nettoyage après des mises à jour incrémentielles afin de récupérer de l'espace de stockage et d'améliorer les performances.

contrôle de version

Processus et outils permettant de suivre les modifications, telles que les modifications apportées au code source dans un référentiel.

Appairage de VPC

Connexion entre deux VPC qui vous permet d'acheminer le trafic à l'aide d'adresses IP privées. Pour plus d'informations, veuillez consulter la rubrique Qu'est-ce que l'appairage de VPC ? dans la documentation Amazon VPC.

vulnérabilités

Défaut logiciel ou matériel qui compromet la sécurité du système.

W

cache actif

Cache tampon qui contient les données actuelles et pertinentes fréquemment consultées. L'instance de base de données peut lire à partir du cache tampon, ce qui est plus rapide que la lecture à partir de la mémoire principale ou du disque.

données chaudes

Données rarement consultées. Lorsque vous interrogez ce type de données, des requêtes modérément lentes sont généralement acceptables.

fonction de fenêtre

Fonction SQL qui effectue un calcul sur un groupe de lignes liées d'une manière ou d'une autre à l'enregistrement en cours. Les fonctions de fenêtre sont utiles pour traiter des tâches, telles que le calcul d'une moyenne mobile ou l'accès à la valeur des lignes en fonction de la position relative de la ligne en cours.

charge de travail

Ensemble de ressources et de code qui fournit une valeur métier, par exemple une application destinée au client ou un processus de backend.

flux de travail

Groupes fonctionnels d'un projet de migration chargés d'un ensemble de tâches spécifique. Chaque flux de travail est indépendant, mais prend en charge les autres flux de travail du projet. Par exemple, le flux de travail du portefeuille est chargé de prioriser les applications, de planifier les vagues et de collecter les métadonnées de migration. Le flux de travail du portefeuille fournit ces actifs au flux de travail de migration, qui migre ensuite les serveurs et les applications.

VER

Voir écrire une fois, lire plusieurs.

WQF

Voir le cadre AWS de qualification de la charge de travail.

écrire une fois, lire plusieurs (WORM)

Modèle de stockage qui écrit les données une seule fois et empêche leur suppression ou leur modification. Les utilisateurs autorisés peuvent lire les données autant de fois que nécessaire, mais ils ne peuvent pas les modifier. Cette infrastructure de stockage de données est considérée comme immuable.

Z

exploit Zero-Day

Une attaque, généralement un logiciel malveillant, qui tire parti d'une vulnérabilité de type « jour zéro ».

vulnérabilité de type « jour zéro »

Une faille ou une vulnérabilité non atténuée dans un système de production. Les acteurs malveillants peuvent utiliser ce type de vulnérabilité pour attaquer le système. Les développeurs prennent souvent conscience de la vulnérabilité à la suite de l'attaque.

invite Zero-Shot

Fournir à un LLM des instructions pour effectuer une tâche, mais aucun exemple (plans) pouvant aider à la guider. Le LLM doit utiliser ses connaissances pré-entraînées pour gérer la tâche. L'efficacité de l'invite zéro dépend de la complexité de la tâche et de la qualité de l'invite. Voir également les instructions en quelques clics.

application zombie

Application dont l'utilisation moyenne du processeur et de la mémoire est inférieure à 5 %. Dans un projet de migration, il est courant de retirer ces applications.

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