Une approche progressive pour l'ingénierie des performances dans le AWS Cloud

AWS Conseils prescriptifs

AWS Conseils prescriptifs: Une approche progressive pour l'ingénierie des performances dans le AWS Cloud

Table of Contents

Une approche progressive pour l'ingénierie des performances dans le AWS Cloud

Amazon Web Services (contributeurs)

Avril 2024 (historique du document)

Ce guide décrit les meilleures pratiques en matière de planification, de création et d'activation de l'ingénierie des performances pour les charges de travail des applications exécutées sur Amazon Web Services (AWS). Il définit quatre piliers pour l'ingénierie des performances et suggère différentes approches pour répondre aux exigences de performance des applications. Pour chaque pilier, ce guide répertorie les outils et les solutions permettant de configurer les tests de performance et l'environnement de test.

Qu'est-ce que l'ingénierie de performance ?

L'ingénierie des performances englobe les techniques appliquées au cours du cycle de développement d'un système pour garantir le respect des exigences de performances non fonctionnelles (telles que le débit, la latence ou l'utilisation de la mémoire).

Avant de démarrer les tests de performances, vous devez configurer l'environnement de performance. Un environnement de performance typique repose sur les piliers suivants :

- Génération de données de test
- Tester l'observabilité
- Automatisation des tests
- Rapport de test

Pourquoi utiliser l'ingénierie de performance ?

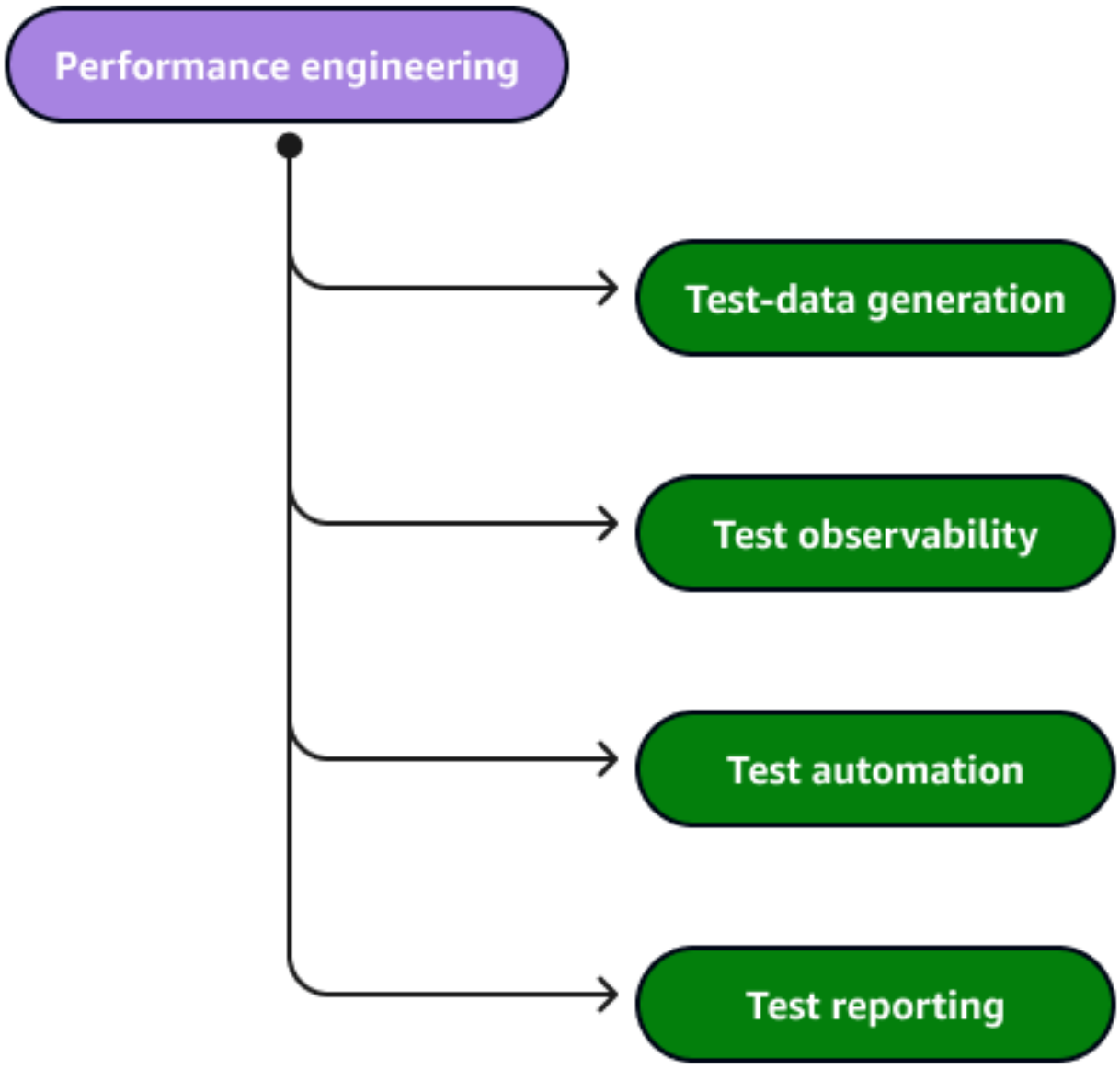
L'ingénierie des performances est le processus d'optimisation continue des performances de l'application dès le début de la phase de conception. Il apporte une grande valeur à l'entreprise en évitant le remaniement et le refactoring du code à un stade ultérieur du cycle de développement. Commencer l'ingénierie des performances dès la phase de conception permet d'obtenir une application plus performante car les performances peuvent être prises en compte dans la conception.

L'ingénierie des performances nécessite la participation active des architectes système, des développeurs et de l'assurance qualité. DevOps

Les piliers de l'ingénierie de performance

Pour mettre en place un état d'esprit d'ingénierie des performances, il est important de créer une base solide lors de la mise en place de l'ingénierie des performances pour l'application. L'ingénierie de la performance nécessite la mise en place de quatre piliers majeurs :

- Génération de données de test — Les ingénieurs de performance configurent des outils pour générer les données de test.
- Observabilité des tests : les ingénieurs de performance configurent l'environnement d'observabilité pour garantir que les performances peuvent être enregistrées et tracées, et que les ressources chargées de gérer les charges sont surveillées.
- Automatisation des tests — Les ingénieurs de performance développent des tests automatisés qui simulent le trafic utilisateur et la charge du système à l'aide d'outils tels qu'Apache JMeter ou ghz.
- Rapports de test — Des données sont collectées sur la configuration de chaque test ainsi que sur les résultats de performance. Les données permettent de corréler les modifications de configuration aux performances et fournissent des informations précieuses.



L'intégration de ces piliers encouragera l'esprit de performance dès les phases initiales de la conception. Cela permettra d'éviter toute modification de l'application ou de l'environnement lors des phases ultérieures de développement et de test.

Génération de données de test

La génération de données de test implique la génération et la maintenance d'une grande quantité de données pour exécuter le scénario de test de performance. Ces données générées servent d'entrée aux scénarios de test afin que l'application puisse être testée sur un ensemble de données diversifié.

La génération de données de test est souvent un processus complexe. Cependant, l'utilisation d'un jeu de données mal créé peut entraîner un comportement imprévisible des applications dans l'environnement de production. La génération de données de test pour les tests de performance diffère des approches traditionnelles de génération de données de test. Cela nécessite des scénarios réels, et la plupart des clients souhaitent tester leurs charges de travail avec des données similaires à leurs données de production réelles. Les données de test générées doivent également généralement être réinitialisées ou actualisées dans leur état d'origine après chaque exécution de test, ce qui augmente le temps et les efforts.

La génération de données de test inclut les principales considérations suivantes :

- Exactitude — La précision des données est importante dans tous les aspects des tests. Des données inexactes créent des résultats inexacts. Par exemple, lorsqu'une transaction par carte de crédit est générée, elle ne devrait pas être pour une date future.
- Validité — Les données doivent être valides pour le cas d'utilisation. Par exemple, lorsque vous testez des transactions par carte de crédit, il n'est pas conseillé de générer 10 000 transactions par utilisateur et par jour, car cela s'écarte considérablement du scénario d'utilisation valide.
- Automatisation — L'automatisation de la génération des données de test peut apporter des avantages en termes de temps et d'efforts. Cela conduit également à une automatisation efficace des tests. La génération manuelle de données de test peut avoir des conséquences sur les exigences en termes de qualité et de temps.

  Il existe différents mécanismes que l'on peut adopter en fonction des cas d'utilisation, comme suit :

  - Pilotée par API : dans ce cas, le développeur fournit une API de génération de données de test que le testeur peut utiliser pour générer des données. À l'aide d'outils de test tels que JMeter, les testeurs peuvent adapter la génération de données à l'aide d'une API métier. Par exemple, si vous disposez d'une API pour ajouter un utilisateur, vous pouvez utiliser la même API pour créer des centaines d'utilisateurs aux profils différents. De même, vous pouvez supprimer les utilisateurs en appelant l'opération d'API de suppression. Pour les applications de flux de travail complexes, le développeur peut fournir une API composite capable de générer des ensembles de données à travers différents composants. Grâce à cette approche, les testeurs peuvent créer des automatismes pour générer et supprimer les ensembles de données en fonction de leurs besoins.

    Toutefois, si le système est complexe ou si le temps de réponse de l'API par appel est élevé, la configuration et le démontage des données peuvent prendre un certain temps.

- Pilotée par des instructions SQL : une autre approche consiste à utiliser des instructions SQL principales pour générer un volume important de données. Le développeur peut fournir des instructions SQL basées sur des modèles pour la génération de données de test. Les testeurs peuvent utiliser les instructions pour renseigner les données, ou ils peuvent créer des scripts wrapper au-dessus de ces instructions pour automatiser la génération des données de test. Grâce à cette approche, les testeurs peuvent renseigner et extraire des données très rapidement si celles-ci doivent être réinitialisées une fois le test terminé. Toutefois, cette approche nécessite un accès direct à la base de données de l'application, ce qui n'est pas toujours possible dans un environnement sécurisé classique. En outre, des requêtes non valides peuvent entraîner une population de données incorrecte, ce qui peut fausser les résultats. Les développeurs doivent également continuellement mettre à jour les instructions SQL dans le code de l'application pour refléter les modifications apportées à l'application au fil du temps.

Outils de génération de données de test

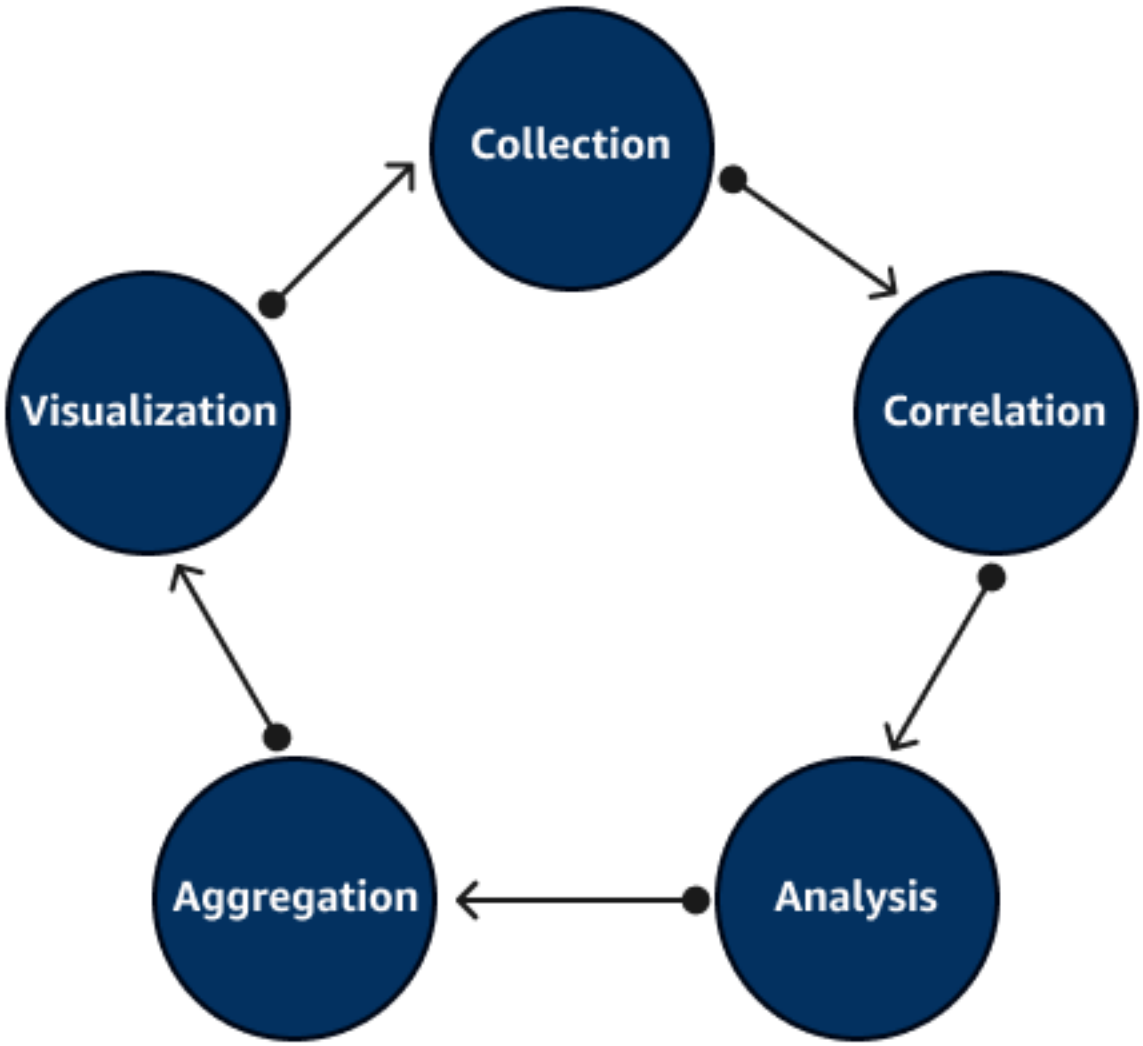
AWS fournit des outils personnalisés natifs que vous pouvez utiliser pour générer des données de test :

- Amazon Kinesis Data Generator : le générateur de données Amazon Kinesis (KDG) simplifie la génération de données et leur envoi à Amazon Kinesis. L'outil fournit une interface utilisateur conviviale qui s'exécute directement dans votre navigateur. Pour plus d'informations et une implémentation de référence, consultez le billet de blog Testez votre solution de données de streaming avec le nouvel Amazon Kinesis Data Generator.
- AWS Glue Générateur de données de test — Le générateur de données de AWS Glue test fournit un cadre configurable pour la génération de données de test à l'aide de tâches sans AWS Glue PySpark serveur. La description des données de test requise est entièrement configurable via un fichier de configuration YAML. Pour plus d'informations et une implémentation de référence, consultez le GitHub référentiel AWS Glue Test Data Generator.

Tester l'observabilité

L'observabilité des tests prend en charge la collecte, la corrélation, l'agrégation et l'analyse de la télémétrie au sein de votre réseau, de votre infrastructure et de vos applications pendant les tests de performance. Vous obtenez des informations complètes sur le comportement, les performances et l'état de santé de votre système. Ces informations vous aident à détecter, étudier et résoudre les

problèmes plus rapidement. En ajoutant l'intelligence artificielle et l'apprentissage automatique, vous pouvez réagir de manière proactive aux problèmes, les prévoir et les prévenir.



L'observabilité repose sur la journalisation, la surveillance et le suivi. La responsabilité de la mise en œuvre réussie de ces activités incombe aux équipes chargées des applications et de l'infrastructure.

Au début de la phase de conception, les équipes d'application doivent comprendre l'état actuel de leur environnement d'observabilité, notamment en matière de journalisation, de surveillance et de suivi. Ils peuvent ensuite choisir des outils qui s'intègrent plus facilement dans la pile d'observabilité.

De même, l'équipe chargée de l'infrastructure est chargée de gérer et de faire évoluer l'infrastructure d'observabilité.

Tenez compte des aspects suivants en ce qui concerne l'observabilité des tests :

- Disponibilité des journaux et des traces des applications
- Corrélation entre les journaux et les traces
- Disponibilité des nœuds, des conteneurs et des métriques des applications
- Automatisation pour configurer et mettre à jour l'infrastructure d'observabilité à la demande
- Possibilité de visualiser la télémétrie
- Mise à l'échelle de l'infrastructure d'observabilité

Logging

La journalisation est le processus qui consiste à conserver des données sur les événements qui se produisent dans un système. Le journal peut inclure des problèmes, des erreurs ou des informations sur l'opération en cours. Les journaux peuvent être classés en différents types, tels que les suivants :

- Journal des événements
- Journal du serveur
- Journal du système
- Journaux d'autorisation et d'accès
- Journaux d’audit

Un développeur peut rechercher dans les journaux des codes ou modèles d'erreur spécifiques, les filtrer en fonction de champs spécifiques ou les archiver de manière sécurisée pour une analyse future. Les journaux aident le développeur à analyser les causes profondes des problèmes de performance et à établir une corrélation entre les composants du système.

La création d'une solution de journalisation efficace implique une étroite coordination entre les équipes chargées de l'application et de l'infrastructure. Les journaux d'applications ne sont utiles que s'il existe une infrastructure de journalisation évolutive prenant en charge des cas d'utilisation tels que l'analyse, le filtrage, la mise en mémoire tampon et la corrélation des journaux. Les cas d'utilisation courants, tels que la génération d'un ID de corrélation, la journalisation du temps d'exécution pour les méthodes critiques et la définition de modèles de journalisation, peuvent être simplifiés.

L'équipe de candidature

Le développeur d'applications doit s'assurer que les journaux générés respectent les meilleures pratiques en matière de journalisation. Les meilleures pratiques sont les suivantes :

- Génération de corrélations IDs pour suivre les demandes uniques
- Enregistrement du temps nécessaire aux méthodes critiques pour l'entreprise
- Journalisation à un niveau de journalisation approprié
- Partage d'une bibliothèque de journalisation commune

Lorsque vous concevez des applications qui interagissent avec différents microservices, utilisez ces principes de conception de journalisation pour simplifier le filtrage et l'extraction des journaux sur le backend.

Génération de corrélations IDs pour suivre les demandes uniques

Lorsque l'application reçoit la demande, elle peut vérifier si un identifiant de corrélation est déjà présent dans l'en-tête. Si aucun identifiant n'est présent, l'application doit en générer un. Par exemple, un Application Load Balancer ajoute un en-tête appelé. `X-Amzn-Trace-Id` L'application peut utiliser l'en-tête pour corréler la demande provenant de l'équilibreur de charge avec l'application. De même, l'application doit effectuer une injection `traceId` si elle appelle des microservices dépendants afin que les journaux générés par les différents composants d'un flux de demandes soient corrélés.

Enregistrement du temps nécessaire aux méthodes critiques pour l'entreprise

Lorsque l'application reçoit une demande, elle interagit avec un autre composant. L'application doit enregistrer le temps nécessaire aux méthodes critiques pour l'entreprise selon un schéma défini. Cela peut faciliter l'analyse des journaux dans le backend. Cela peut également vous aider à générer des informations utiles à partir des journaux. Vous pouvez utiliser des approches telles que la programmation orientée aspect (AOP) pour générer de tels journaux afin de séparer les préoccupations liées à la journalisation de votre logique métier.

Journalisation à un niveau de journalisation approprié

L'application doit écrire des journaux contenant une quantité d'informations utiles. Utilisez les niveaux de journalisation pour classer les événements en fonction de leur gravité. Par exemple, utilisation `WARNING` et `ERROR` niveaux pour les événements importants nécessitant une enquête. Utilisez `INFO` et `DEBUG` pour le suivi détaillé et les événements à volume élevé. Configurez les gestionnaires de journaux pour qu'ils capturent uniquement les niveaux nécessaires à la production. Générer trop de journalisation au `INFO` niveau n'est pas utile et cela ajoute de la pression sur l'infrastructure principale. `DEBUG`la journalisation peut être utile, mais elle doit être utilisée avec

prudence. L'utilisation de DEBUG journaux peut générer un volume important de données, elle n'est donc pas recommandée dans un environnement de test de performances.

Partage d'une bibliothèque de journalisation commune

Les équipes d'application doivent utiliser une bibliothèque de journalisation commune, par exemple avec un modèle de journalisation commun prédéfini que les développeurs peuvent utiliser comme dépendances dans leur projet. AWS SDK pour Java

Équipe chargée de l'infrastructure

DevOps les ingénieurs peuvent réduire leurs efforts en utilisant les principes de conception de journalisation suivants lors du filtrage et de l'extraction des journaux sur le backend. L'équipe chargée de l'infrastructure doit configurer et prendre en charge les ressources suivantes.

Agent de journalisation

Un agent de journalisation (expéditeur de journaux) est un programme qui lit les journaux depuis un emplacement et les envoie vers un autre emplacement. Les agents de journalisation sont utilisés pour lire les fichiers journaux stockés sur un ordinateur et télécharger les événements du journal vers le backend à des fins de centralisation.

Les journaux sont des données non structurées qui doivent être structurées pour que vous puissiez en tirer des informations pertinentes. Les agents de journalisation utilisent des analyseurs pour lire les instructions du journal et extraire les champs pertinents tels que l'horodatage, le niveau du journal et le nom du service, et ils structurent ces données au format JSON. Il est utile de disposer d'un agent de journalisation léger à la périphérie, car cela réduit l'utilisation des ressources. L'agent de journalisation peut envoyer les données directement au serveur principal ou utiliser un redirecteur de journal intermédiaire qui envoie les données vers le serveur principal. L'utilisation d'un redirecteur de journal décharge le travail des agents de journal à la source.

Analyseur de journaux

Un analyseur de journaux convertit les journaux non structurés en journaux structurés. Les analyseurs des agents de journalisation enrichissent également les journaux en ajoutant des métadonnées. L'analyse des données peut être effectuée à la source (côté application) ou centralisée. Le schéma de stockage des journaux doit être extensible afin que vous puissiez ajouter de nouveaux champs. Nous vous recommandons d'utiliser des formats de journal standard tels que JSON. Cependant, dans certains cas, les journaux doivent être transformés au format JSON pour

une meilleure recherche. L'écriture de la bonne expression d'analyseur permet une transformation efficace.

Backend de journaux

Un service de gestion des journaux collecte, ingère et visualise les données des journaux provenant de diverses sources. L'agent de journalisation peut écrire directement dans le backend ou utiliser un redirecteur de journal intermédiaire. Lors des tests de performance, veillez à stocker les journaux afin qu'ils puissent être consultés ultérieurement. Stockez les journaux séparément dans le backend pour chaque application. Par exemple, utilisez un index dédié pour une application et utilisez un modèle d'index pour rechercher des journaux répartis entre différentes applications connexes. Nous vous recommandons de sauvegarder au moins 7 jours de données pour la recherche dans les journaux. Cependant, le stockage des données pendant une plus longue durée peut entraîner des coûts de stockage inutiles. Étant donné qu'un grand volume de journaux est généré pendant le test de performance, il est important que l'infrastructure de journalisation adapte et dimensionne correctement le backend de journalisation.

Visualisation du journal

Pour obtenir des informations pertinentes et exploitables à partir des journaux d'applications, utilisez des outils de visualisation dédiés pour traiter et transformer les données brutes des journaux en représentations graphiques. Les visualisations telles que les tableaux, les graphiques et les tableaux de bord peuvent aider à découvrir les tendances, les modèles et les anomalies qui peuvent ne pas être facilement apparents lorsque l'on examine les journaux bruts.

Les principaux avantages de l'utilisation d'outils de visualisation incluent la possibilité de corréler les données entre plusieurs systèmes et applications afin d'identifier les dépendances et les goulots d'étranglement. Les tableaux de bord interactifs permettent d'explorer les données à différents niveaux de granularité pour résoudre les problèmes ou identifier les tendances d'utilisation. Les plateformes de visualisation de données spécialisées fournissent des fonctionnalités telles que l'analyse, les alertes et le partage de données qui peuvent améliorer la surveillance et l'analyse.

En utilisant la puissance de la visualisation des données sur les journaux des applications, les équipes de développement et d'exploitation peuvent gagner en visibilité sur les performances du système et des applications. Les informations obtenues peuvent être utilisées à diverses fins, notamment pour optimiser l'efficacité, améliorer l'expérience utilisateur, renforcer la sécurité et planifier les capacités. Le résultat final est des tableaux de bord adaptés aux différentes parties prenantes, fournissant des at-a-glance vues qui résument les données des journaux en informations exploitables et pertinentes.

Automatisation de l'infrastructure de journalisation

Étant donné que les différentes applications ont des exigences différentes, il est important d'automatiser l'installation et le fonctionnement de l'infrastructure de journalisation. Utilisez des outils d'infrastructure en tant que code (IaC) pour approvisionner le backend de l'infrastructure de journalisation. Vous pouvez ensuite fournir l'infrastructure de journalisation sous la forme d'un service partagé ou d'un déploiement indépendant sur mesure pour une application particulière.

Nous recommandons aux développeurs d'utiliser des pipelines de livraison continue (CD) pour automatiser les opérations suivantes :

- Déployez l'infrastructure de journalisation à la demande et démolissez-la lorsqu'elle n'est pas nécessaire.
- Déployez des agents de journalisation sur différentes cibles.
- Déployez les configurations de l'analyseur de journaux et du redirecteur.
- Déployez des tableaux de bord d'applications.

Outils de journalisation

AWS fournit des services natifs de journalisation, d'alarme et de tableau de bord. Les ressources suivantes sont populaires Services AWS et concernent la journalisation :

- Amazon OpenSearch Service aide les entreprises à collecter, à ingérer et à visualiser les données de journal provenant de diverses sources. Pour plus d'informations, consultez la section Journalisation centralisée avec OpenSearch.
- Amazon CloudWatch agent et AWS for Fluent Bit sont les agents de journalisation les plus populaires sur AWS. Pour plus d'informations sur l'utilisation de l' CloudWatch agent avec Amazon CloudWatch Logs Insights, consultez le billet de blog Simplifying Apache server logs with Amazon CloudWatch Logs Insights. AWS Pour l'implémentation de référence de Fluent Bit, consultez le billet de blog Centralized Container Logging with Fluent Bit.

Contrôle

La surveillance consiste à collecter différentes métriques, telles que le processeur et la mémoire, et à les stocker dans une base de données chronologique telle qu'Amazon Managed Service for Prometheus. Le système de surveillance peut être basé sur le push ou le pull. Dans les systèmes

basés sur le push, la source envoie régulièrement des métriques à la base de données de séries chronologiques. Dans les systèmes basés sur le pull, le scraper extrait les métriques de diverses sources et les stocke dans la base de données de séries chronologiques. Les développeurs peuvent analyser les indicateurs, les filtrer et les tracer au fil du temps pour visualiser les performances. La mise en œuvre réussie de la surveillance peut être divisée en deux grands domaines : l'application et l'infrastructure.

Pour les développeurs d'applications, les indicateurs suivants sont essentiels :

- Latence : temps nécessaire pour recevoir une réponse
- Débit de demandes : nombre total de demandes traitées par seconde
- Taux d'erreur des demandes : nombre total d'erreurs

Capturez l'utilisation des ressources, la saturation et le nombre d'erreurs pour chaque ressource (telle que le conteneur d'applications, la base de données) impliquée dans la transaction commerciale. Par exemple, lorsque vous surveillez l'utilisation du processeur, vous pouvez suivre l'utilisation moyenne du processeur, la charge moyenne et la charge maximale pendant l'exécution des tests de performances. Lorsqu'une ressource atteint la saturation lors d'un test de stress, mais qu'elle risque de ne pas atteindre la saturation pendant une période plus courte pendant une période plus courte.

Métriques

Les applications peuvent utiliser différents actionneurs, tels que des actionneurs à ressort, pour surveiller leurs applications. Ces bibliothèques de production exposent généralement un point de terminaison REST pour surveiller les informations relatives aux applications en cours d'exécution. Les bibliothèques peuvent surveiller l'infrastructure sous-jacente, les plateformes d'applications et les autres ressources. Si l'une des métriques par défaut ne répond pas aux exigences, le développeur doit implémenter des métriques personnalisées. Les métriques personnalisées peuvent aider à suivre les indicateurs de performance clés de l'entreprise (KPIs) qui ne peuvent pas être suivis à l'aide des données issues des implémentations par défaut. Par exemple, vous souhaiterez peut-être suivre une opération commerciale telle que la latence d'intégration d'API tierces ou le nombre total de transactions effectuées.

Cardinalité

La cardinalité fait référence au nombre de séries chronologiques uniques d'une métrique. Les métriques sont étiquetées pour fournir des informations supplémentaires. Par exemple, une application basée sur REST qui suit le nombre de demandes pour une API particulière indique une

cardinalité de 1. Si vous ajoutez une étiquette utilisateur pour identifier le nombre de demandes par utilisateur, la cardinalité augmente proportionnellement au nombre d'utilisateurs. En ajoutant des étiquettes qui créent une cardinalité, vous pouvez découper et découper les métriques en différents groupes. Il est important d'utiliser les bonnes étiquettes pour le bon cas d'utilisation, car la cardinalité augmente le nombre de séries de mesures dans la base de données de séries chronologiques de surveillance du backend.

Résolution

Dans une configuration de surveillance classique, l'application de surveillance est configurée pour extraire périodiquement les métriques de l'application. La périodicité du scraping définit la granularité des données de surveillance. Les mesures collectées à des intervalles plus courts ont tendance à fournir une vision plus précise des performances, car davantage de points de données sont disponibles. Toutefois, la charge sur la base de données de séries chronologiques augmente à mesure que de nouvelles entrées sont stockées. Généralement, une granularité de 60 secondes correspond à une résolution standard et de 1 seconde à une résolution élevée.

DevOps équipe

Les développeurs d'applications demandent souvent aux DevOps ingénieurs de mettre en place un environnement de surveillance pour visualiser les métriques de l'infrastructure et des applications. L' DevOps ingénieur doit mettre en place un environnement évolutif prenant en charge les outils de visualisation des données utilisés par le développeur de l'application. Cela implique de récupérer les données de surveillance provenant de différentes sources et de les envoyer à une base de données chronologique centrale telle qu'Amazon Managed Service for Prometheus.

Backend de surveillance

Un service principal de surveillance prend en charge la collecte, le stockage, l'interrogation et la visualisation des données métriques. Il s'agit généralement d'une base de données chronologique telle qu'Amazon Managed Service for InfluxData Prometheus ou InfluxDB. À l'aide d'un mécanisme de découverte des services, le collecteur de surveillance peut collecter des métriques provenant de différentes sources et les stocker. Lors des tests de performance, il est important de stocker les données des métriques afin de pouvoir les rechercher ultérieurement. Nous vous recommandons de sauvegarder au moins 15 jours de données pour les métriques. Cependant, le stockage des métriques sur une plus longue durée n'apporte pas d'avantages significatifs et entraîne des coûts de stockage inutiles. Étant donné que le test de performance peut générer un grand nombre de métriques, il est important que l'infrastructure de métriques soit évolutive tout en fournissant des

performances de requête rapides. Le service principal de surveillance fournit un langage de requête qui peut être utilisé pour afficher les données des métriques.

Visualisation

Fournissez des outils de visualisation capables d'afficher les données de l'application afin de fournir des informations pertinentes. L' DevOps ingénieur et le développeur de l'application doivent apprendre le langage de requête pour le backend de surveillance et travailler en étroite collaboration pour générer un modèle de tableau de bord réutilisable. Dans les tableaux de bord, incluez la latence et les erreurs, tout en affichant l'utilisation et la saturation des ressources de l'infrastructure et des applications.

Automatisation de l'infrastructure de surveillance

À l'instar de la journalisation, il est important d'automatiser l'installation et le fonctionnement de l'infrastructure de surveillance afin de répondre aux différentes exigences des différentes applications. Utilisez les outils IaC pour approvisionner le backend de l'infrastructure de surveillance. Vous pouvez ensuite fournir l'infrastructure de surveillance sous la forme d'un service partagé ou d'un déploiement indépendant sur mesure pour une application particulière.

Utilisez des pipelines de CD pour automatiser les opérations suivantes :

- Déployez l'infrastructure de surveillance à la demande et démolissez-la lorsqu'elle n'est pas nécessaire.
- Mettez à jour la configuration de surveillance pour filtrer ou agréger les métriques.
- Déployez des tableaux de bord d'applications.

Outils de surveillance

Amazon Managed Service for Prometheus est un service de surveillance compatible avec Prometheus pour l'infrastructure de conteneurs et les métriques d'application pour les conteneurs, que vous pouvez utiliser pour surveiller en toute sécurité les environnements de conteneurs à grande échelle. Pour plus d'informations, consultez le billet de blog Getting Started with Amazon Managed Service for Prometheus.

Amazon CloudWatch fournit une surveillance complète du stack sur. AWS CloudWatch prend en charge les solutions AWS natives et open source afin que vous puissiez comprendre à tout moment ce qui se passe dans votre infrastructure technologique.

Les AWS outils natifs sont les suivants :

- Tableaux de CloudWatch bord Amazon
- Container Insights CloudWatch
- CloudWatch métriques
- CloudWatch alarmes

Amazon CloudWatch propose des fonctionnalités spécialement conçues pour répondre à des cas d'utilisation spécifiques tels que la surveillance des conteneurs via CloudWatch Container Insights. Ces fonctionnalités sont intégrées CloudWatch afin que vous puissiez configurer les journaux, la collecte de métriques et la surveillance.

Pour vos applications conteneurisées et vos microservices, utilisez Container Insights pour collecter, agréger et résumer les métriques et les journaux. Container Insights est disponible pour les plateformes Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS) et Kubernetes sur Amazon Elastic Compute Cloud (Amazon EC2). Container Insights collecte des données sous forme d'événements du journal des performances au format métrique intégré. Ces entrées d'événements du journal des performances utilisent un schéma JSON structuré qui prend en charge l'ingestion et le stockage de données à haute cardinalité à grande échelle.

Pour plus d'informations sur la mise en œuvre de Container Insights avec Amazon EKS, consultez le billet de blog Présentation d'Amazon CloudWatch Container Insights for Amazon EKS Fargate using Distro AWS for. OpenTelemetry

Tracing

Le traçage implique l'utilisation spécialisée des informations de journalisation relatives aux processus d'un programme. Les informations issues des journaux peuvent aider les ingénieurs à déboguer des transactions individuelles et à identifier les goulots d'étranglement. Le traçage peut être activé automatiquement ou à l'aide d'instruments manuels.

Comme une application s'intègre à différents services, il est important d'identifier les performances de l'application et de ses services sous-jacents. Le traçage fonctionne avec les traces et les étendues. Une trace est le processus de demande complet, et chaque trace est composée de plages. Un intervalle est un intervalle de temps étiqueté et représente l'activité au sein des composants ou services individuels d'un système. Les traces fournissent une vue d'ensemble de ce qui se passe lorsqu'une demande est envoyée à une application.

L'équipe de candidature

Les développeurs d'applications instrumentent leurs applications en envoyant des données de suivi pour les demandes entrantes et sortantes et d'autres événements au sein de l'application, ainsi que des métadonnées relatives à chaque demande. Pour générer des traces, une application doit être instrumentée pour générer des traces. L'instrumentation peut être automatique ou manuelle.

Instrumentation automatique

Vous pouvez collecter des données télémétriques à partir d'une application à l'aide d'instruments automatiques sans avoir à modifier le code source. Les agents d'instrumentation automatique peuvent générer des traces d'application d'une application ou d'un service. Généralement, vous utilisez les modifications de configuration pour ajouter l'agent ou un autre mécanisme.

L'instrumentation de bibliothèque implique d'apporter des modifications minimales au code d'application pour ajouter une instrumentation prédéfinie. L'instrumentation cible des bibliothèques ou des frameworks spécifiques, tels que le AWS SDK, les clients HTTP Apache ou les clients SQL.

Instrumentation manuelle

Dans cette approche, les développeurs d'applications ajoutent du code d'instrumentation à l'application à chaque emplacement où ils souhaitent collecter des informations de suivi. Par exemple, utilisez la programmation orientée aspect (AOP) pour collecter des données AWS X-Ray de suivi. Les développeurs peuvent l'utiliser SDKs pour instrumenter leurs applications.

Echantillonnage

Les données de suivi sont souvent générées en gros volumes. Il est important de disposer d'un mécanisme permettant de déterminer si les données de suivi doivent être exportées ou non. L'échantillonnage est le processus qui permet de déterminer quelles données doivent être exportées. Cela est généralement fait pour réduire les coûts. En personnalisant les règles d'échantillonnage, vous pouvez contrôler la quantité de données que vous enregistrez. Vous pouvez également modifier le comportement d'échantillonnage sans modifier ni redéployer votre code. Il est important de contrôler le taux d'échantillonnage pour générer la bonne quantité de traces.

Les développeurs d'applications peuvent annoter les traces en ajoutant des métadonnées sous forme de paires clé-valeur. Les annotations enrichissent les traces et aident à affiner le filtrage dans le backend.

DevOps équipe

DevOps les ingénieurs sont souvent invités à configurer un environnement de traçage permettant au développeur d'applications de visualiser les traces de l'infrastructure et des applications. La configuration de l'environnement de suivi implique de collecter des données de suivi provenant de différentes sources et de les envoyer à un magasin central à des fins de visualisation.

Backend de suivi

Un backend de suivi est un service tel AWS X-Ray que celui qui collecte des données sur les demandes traitées par votre application. Il fournit des outils que vous pouvez utiliser pour visualiser, filtrer et obtenir des informations sur ces données afin d'identifier les problèmes et les opportunités d'optimisation. Pour toute demande retracée envoyée à votre application, vous pouvez consulter des informations détaillées sur la demande et la réponse, ainsi que sur les autres appels que votre application fait aux AWS ressources en aval, aux microservices, aux bases de données et au Web APIs.

Automatiser le traçage

Étant donné que les différentes applications ont des exigences de suivi différentes, il est important d'automatiser la configuration et le fonctionnement de l'infrastructure de suivi. Utilisez les outils IaC pour approvisionner le backend de l'infrastructure de suivi.

Utilisez des pipelines de CD pour automatiser les opérations suivantes :

- Déployez l'infrastructure de suivi à la demande et démontez-la lorsqu'elle n'est pas nécessaire.
- Déployez la configuration de suivi dans toutes les applications.

Outils de traçage

AWS fournit les services suivants pour le traçage et la visualisation associée :

- AWS X-Ray reçoit des traces de votre application, en plus des traces des AWS services utilisés par votre application qui sont déjà intégrés à X-Ray. Il existe plusieurs SDKs agents et outils qui peuvent être utilisés pour instrumenter votre application pour le traçage par rayons X. Pour plus d’informations, consultez la documentation AWS X-Ray.

Les développeurs peuvent également l'utiliser AWS X-Ray SDKs pour envoyer des traces à X-Ray. AWS X-Ray fournit SDKs GoJava, Node.jsPython, .NET etRuby. Chaque kit de développement X-Ray fournit les éléments suivants :

- Des intercepteurs à ajouter à votre code pour suivre les demandes HTTP entrantes
- Des gestionnaires de clients pour instrumenter les clients du AWS SDK que votre application utilise pour appeler d'autres services AWS
- Un client HTTP pour instrumenter les appels vers d'autres services Web HTTP internes et externes

X-Ray prend SDKs également en charge l'instrumentation des appels aux bases de données SQL, l'instrumentation client automatique du AWS SDK et d'autres fonctionnalités. Au lieu d'envoyer les données de suivi directement à X-Ray, le SDK envoie des documents de segments JSON à un processus démon qui écoute le trafic UDP. Le daemon X-Ray met en mémoire tampon les segments d'une file d'attente et les télécharge dans X-Ray par lots. Pour plus d'informations sur l'instrumentation de votre application à l'aide d'un SDK X-Ray, consultez la documentation de X-Ray.

- Amazon OpenSearch Service est un service AWS géré permettant d'exécuter et de dimensionner OpenSearch des clusters, qui peut être utilisé pour stocker de manière centralisée les journaux, les métriques et les traces. Le plugin Observability offre une expérience unifiée pour la collecte et la surveillance des métriques, des journaux et des traces provenant de sources de données communes. La collecte et la surveillance des données en un seul endroit offrent une end-to-end observabilité complète de l'ensemble de votre infrastructure. Pour plus d'informations sur la mise en œuvre, consultez la documentation du OpenSearch service.
- AWS Distro for OpenTelemetry (ADOT) est une AWS distribution basée sur le projet Cloud Native Computing Foundation (CNCF). OpenTelemetry ADOT inclut actuellement le support de l'instrumentation automatique pour Java et Python. En outre, ADOT prend en charge l'instrumentation automatique des AWS Lambda fonctions et de leurs requêtes en aval à l'aide Java de Node.js et d'environnements d'Pythonexécution, via des couches Lambda gérées par ADOT. Les développeurs peuvent utiliser le collecteur ADOT pour envoyer des traces vers différents backends, notamment AWS X-Ray Amazon OpenSearch Service.

  Pour un exemple de référence expliquant comment instrumenter votre application à l'aide du SDK ADOT, consultez la documentation. Pour un exemple de référence expliquant comment utiliser le SDK ADOT pour envoyer des données à Amazon OpenSearch Service, consultez la documentation du OpenSearch service.

Pour un exemple de référence expliquant comment instrumenter votre application exécutée sur Amazon EKS, consultez le billet de blog Metrics and traces collection using Amazon EKS add-ons for AWS Distro for OpenTelemetry.

Automatisation des tests

Les tests automatisés avec un cadre et des outils spécialisés peuvent réduire l'intervention humaine et optimiser la qualité. Les tests de performance automatisés ne sont pas différents des tests d'automatisation tels que les tests unitaires et les tests d'intégration.

Utilisez DevOps des pipelines aux différentes étapes pour les tests de performance.

Les cinq étapes du pipeline d'automatisation des tests sont les suivantes :

1. Configuration : utilisez les approches de données de test décrites dans la section Génération de données de test pour cette étape. La génération de données de test réalistes est essentielle pour obtenir des résultats de test valides. Vous devez créer avec soin diverses données de test couvrant un large éventail de cas d'utilisation et correspondant étroitement aux données de production réelles. Avant d'exécuter des tests de performances complets, vous devrez peut-être exécuter des tests d'essai initiaux pour valider les scripts de test, les environnements et les outils de surveillance.
2. Outil de test — Pour effectuer les tests de performance, sélectionnez un outil de test de charge approprié, tel que JMeter ou ghz. Déterminez la solution la mieux adaptée aux besoins de votre entreprise en termes de simulation de charges d'utilisateurs réelles.
3. Exécution des tests : une fois les outils et les environnements de test établis, exécutez des tests de end-to-end performances sur une plage de charges utilisateur et de durées attendues. Tout au long du test, surveillez de près l'état du système testé. Il s'agit généralement d'une étape de longue haleine. Surveillez les taux d'erreur pour l'invalidation automatique des tests et arrêtez le test s'il y a trop d'erreurs.

   L'outil de test de charge fournit des informations sur l'utilisation des ressources, les temps de réponse et les goulots d'étranglement potentiels.

4. Rapports de test — Collectez les résultats des tests ainsi que la configuration des applications et des tests. Automatisez la collecte de la configuration des applications, de la configuration des tests et des résultats, ce qui permet d'enregistrer les données relatives aux tests de performance et de les stocker de manière centralisée. La gestion centralisée des données de performance permet de fournir de bonnes informations et de définir les critères de réussite de manière programmatique pour votre entreprise.
5. Nettoyage : une fois que vous avez terminé un test de performance, réinitialisez l'environnement de test et les données pour préparer les essais suivants. Tout d'abord, vous annulez les modifications apportées aux données de test pendant l'exécution. Vous devez restaurer les bases de données et les autres banques de données dans leur état d'origine, en rétablissant tous les enregistrements nouveaux, mis à jour ou supprimés générés pendant le test.

Vous pouvez réutiliser le pipeline pour répéter le test plusieurs fois jusqu'à ce que les résultats reflètent les performances souhaitées. Vous pouvez également utiliser le pipeline pour vérifier que les modifications de code n'altèrent pas les performances. Vous pouvez exécuter des tests de validation de code en dehors des heures de bureau et utiliser les données de test et d'observabilité disponibles pour le dépannage.

Les meilleures pratiques sont les suivantes :

- Enregistrez les heures de début et de fin, et générez-les automatiquement URLs pour la journalisation. Cela vous aide à filtrer les données d'observabilité dans la fenêtre temporelle appropriée. Systèmes de surveillance et de suivi.
- Injectez des identifiants de test dans l'en-tête lors de l'appel des tests. Les développeurs d'applications peuvent enrichir leurs données de journalisation, de surveillance et de suivi en utilisant l'identifiant comme filtre dans le backend.
- Limitez le pipeline à une seule exécution à la fois. L'exécution de tests simultanés génère du bruit susceptible de prêter à confusion lors du dépannage. Il est également important d'exécuter le test dans un environnement de performance dédié.

Outils d'automatisation des tests

Les outils de test jouent un rôle important dans toute automatisation des tests. Les choix les plus courants pour les outils de test open source sont les suivants :

- Apache JMeter est un cheval puissant aguerri. Au fil des ans, Apache est JMeter devenu plus fiable et a ajouté des fonctionnalités. Grâce à l'interface graphique, vous pouvez créer des tests complexes sans connaître de langage de programmation. Des entreprises telles que BlazeMeter Supportent Apache JMeter.
- K6 est un outil gratuit qui offre une assistance, un hébergement de la source de charge et une interface Web intégrée pour organiser, exécuter et analyser les tests de charge.
- Le test de charge Vegeta suit un concept différent. Au lieu de définir la simultanéité ou d'imposer une charge à votre système, vous définissez un certain taux. L'outil crée ensuite cette charge indépendamment des temps de réponse de votre système.
- Hey et ab, l'outil d'analyse comparative du serveur HTTP Apache, sont des outils de base que vous pouvez utiliser depuis la ligne de commande pour exécuter la charge spécifiée sur un seul point de terminaison. C'est le moyen le plus rapide de générer une charge si vous disposez d'un serveur sur lequel vous pouvez exécuter les outils. Même un ordinateur portable local sera performant, même s'il n'est peut-être pas assez puissant pour produire une charge élevée.
- ghz est un utilitaire de ligne de commande et un package Go pour les tests de charge et l'analyse comparative des services gRPC.

AWS fournit le test de charge distribué sur la AWS solution. La solution crée et simule des milliers d'utilisateurs connectés qui génèrent des enregistrements transactionnels à un rythme constant, sans qu'il soit nécessaire de configurer des serveurs. Pour plus d'informations, consultez la bibliothèque de AWS solutions.

Vous pouvez l'utiliser AWS CodePipeline pour automatiser le pipeline de tests de performances. Pour plus d'informations sur l'automatisation de vos tests d'API en utilisant CodePipeline, consultez le AWS DevOps blog et la AWS documentation.

Rapport de test

Les rapports de test font référence à la collecte, à l'analyse et à la présentation de données relatives aux performances des systèmes, des applications, des services ou des processus. Il s'agit de mesurer divers paramètres et indicateurs pour évaluer l'efficacité, la réactivité, la fiabilité et l'efficacité globale d'un système ou d'un composant en particulier.

Le reporting des tests de performance implique le choix de mesures pertinentes en fonction du contexte et des objectifs de l'analyse. Les indicateurs de performance courants incluent les temps de

réponse, le débit, les taux d'erreur, l'utilisation des ressources (processeur, mémoire, disque) et la latence du réseau.

Une fois que les données relatives aux performances ont été collectées, elles doivent être stockées dans un référentiel central. Les résultats de ces tests peuvent provenir de différents environnements, applications et outils de test. Lorsque plusieurs charges de travail s'exécutent dans différents environnements, il est difficile de recueillir des données relatives aux performances et d'établir des corrélations entre ces points de données afin de tirer des conclusions éclairées. Nous recommandons de définir une méthode standard pour collecter des données de mesures de performance à l'aide d'un référentiel central pour le stockage et la visualisation des données.

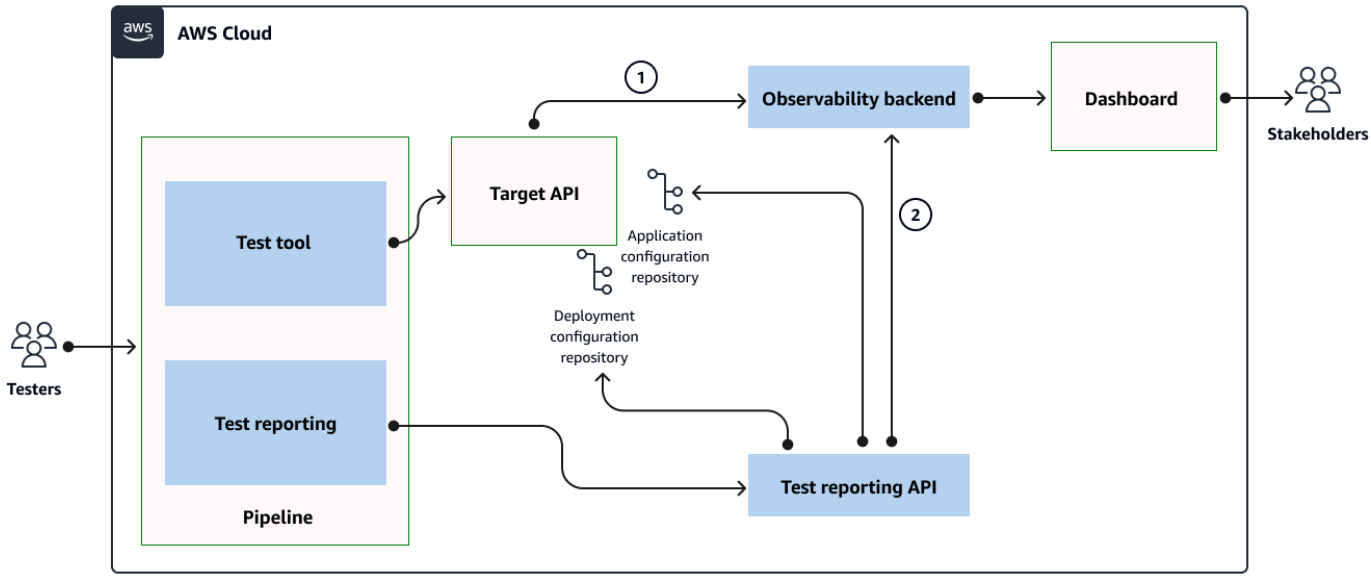
Enregistrement standardisé

Nous recommandons de standardiser la manière dont les différentes parties prenantes effectuent les tests de performance et écrivent les données obtenues dans un référentiel central. Par exemple, cela peut prendre la forme d'une API acceptant les résultats et les stockant dans une solution de stockage persistant. Dans les situations où les données doivent être extraites de sources telles qu' GitOps Amazon Managed Service pour Prometheus, l'API peut directement extraire ces informations des sources spécifiées sur la base de fichiers de schéma qui décrivent comment extraire les champs des spécifications de déploiement et des spécifications Kubernetes. Les fichiers de schéma peuvent utiliser des `JSONPath` expressions ou le langage de requête Prometheus (ProMQL). Comme mentionné précédemment, les mesures collectées doivent être pertinentes par rapport au contexte et aux objectifs de l'analyse des performances.

Les données transmises à l'API peuvent inclure des détails et des balises relatifs à l'application et à l'environnement pour lesquels le test a été effectué. Cela permet d'effectuer des analyses sur les données des tests de performance.

Les piliers de l'ingénierie de performance en action

L'architecture de référence suivante illustre les piliers de l'ingénierie des performances pour tester une API spécifique.



1. Les données de journalisation, de surveillance et de suivi sont envoyées depuis l'API cible vers le backend.
2. Lorsqu'elle est invoquée, l'API de rapports de test envoie les résultats et les informations de configuration au backend.

Le composant principal est l'API ou l'application cible en cours de test. L'API cible se synchronise avec le référentiel de configuration des applications et le référentiel de configuration de déploiement de GitOps manière à obtenir les dernières configurations d'applications et d'infrastructures. Cette synchronisation permet aux tests automatisés de s'exécuter par rapport à l'état actuel souhaité de l'application et de son infrastructure de support, tel que défini dans les référentiels Git.

Le pipeline d'automatisation des tests automatise la génération des données de test, l'exécution du test et la communication des résultats des tests pour l'API cible.

L'API cible génère des informations sur les performances (métriques, journaux et traces), en utilisant les meilleures pratiques d'observabilité, et elle diffuse les données de métriques vers le backend d'observabilité.

L'API de rapports de test collecte toutes les données de reporting liées aux tests (configuration et résultats des tests) et les stocke dans le backend d'observabilité.

L'agrégation des informations sur les performances et des données de reporting (configuration, résultats des tests) vous permet d'interroger les données relatives aux performances pour l'API cible. Par exemple, vous pouvez demander ce qui suit :

- Quelles sont les dix transactions les plus lentes ?
- Quel est le nombre moyen de chaque test (P99, P90) ?
- Comment se comparent les configurations des deux essais ?

La corrélation des cas de test avec les résultats, les configurations et les métriques sur une période donnée permet d'identifier la meilleure configuration et les meilleurs résultats de performance.

À l'aide de ces résultats de test, vous pouvez prendre des décisions plus précises, basées sur les données, pour l'API et être en mesure de mettre l'API en production en toute confiance.

Ressources

Services AWS

- Amazon CloudWatch
- AWS CodePipeline
- AWS Distro pour OpenTelemetry
- Amazon OpenSearch Service
- AWS X-Ray

Implémentations

- amazon-kinesis-data-generator
- AWS Glue Générateur de données de test
- Test de charge distribué sur AWS

Billets de blogs

- Enregistrement centralisé des conteneurs avec Fluent Bit
- Testez votre solution de streaming avec le nouveau générateur de données Amazon Kinesis
- Présentation d'Amazon CloudWatch Container Insights pour Amazon EKS AWS Fargate à l'aide de Distro pour OpenTelemetry
- Suivi des applications sur Kubernetes avec AWS X-Ray
- Collecte de métriques et de traces à l'aide des modules complémentaires Amazon EKS pour AWS Distro pour OpenTelemetry
- Commencer à utiliser Amazon Managed Service pour Prometheus

Atelier

- Introduction à l' AWS observabilité

AWS Conseils prescriptifs

- Applications de test de charge (guide)

Applications tierces

- Apache JMeter
- K6
- Vegeta
- Salut et Ab
- ghz

Collaborateurs

Les personnes qui ont contribué à ce document incluent :

- Varun Sharma, consultant principal principal, AWS
- Akash Kumar, consultant principal principal, AWS
- Archana Bhatnagar, responsable du cabinet, AWS
- Pratik Sharma, Services professionnels II, AWS

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	—	24 avril 2024

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.