Cadre d'analyse de résilience

AWS Conseils prescriptifs

AWS Conseils prescriptifs: Cadre d'analyse de résilience

Table of Contents

Cadre d'analyse de résilience

John Formento, Bruno Emer, Steven Hooper, Jason Barto et Michael Haken, Amazon Web Services (AWS)

septembre 2023(historique du document)

Des normes et des processus cohérents et reproductibles jouent un rôle important dans l'amélioration continue. Cela vaut également pour la résilience des systèmes distribués. L'objectif de ce guide est de présenter un cadre d'analyse de résilience qui fournit un moyen cohérent d'analyser les modes de défaillance et leur impact potentiel sur vos charges de travail. L'utilisation de ce cadre tout au long du cycle de vie de votre charge de travail, de la conception à l'exploitation, vous permet d'améliorer en permanence la résilience de vos charges de travail face à un plus large éventail de modes de défaillance potentiels de manière cohérente et reproductible. Cela permet de garantir que vous atteignez vos objectifs de résilience et que vous conservez les propriétés de résilience souhaitées pour vos charges de travail.

Ce framework a été développé grâce à l'expérience des équipes de terrain chargées de l'architecture des solutions AWS dans le cadre de leur travail avec des clients de tous les secteurs. Il cible les constructeurs qui peuvent avoir de nombreux titres de poste, notamment les chefs de produit, les développeurs de logiciels, les ingénieurs systèmes, les équipes opérationnelles et les architectes. Ce sont les personnes qui connaissent le mieux le système, le service ou le produit analysé. L'utilisation du cadre dans le cadre d'exercices continus peut vous aider à réaliser des progrès progressifs et à atteindre vos objectifs de résilience à long terme.

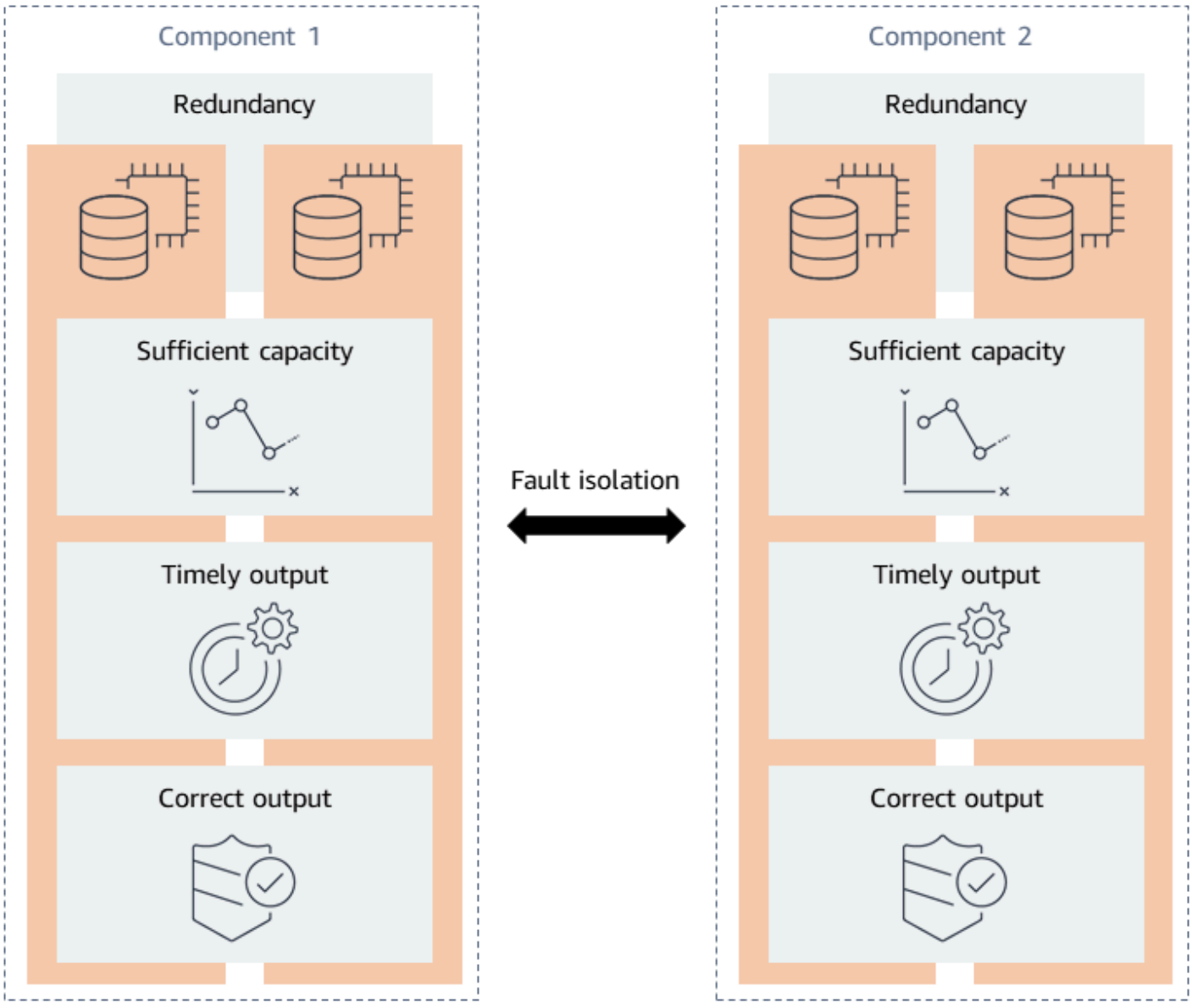
L'objectif du cadre est d'identifier les modes de défaillance potentiels et les contrôles préventifs et correctifs que vous pouvez utiliser pour atténuer leur impact. Même si les défaillances se produisent dans des composants qui ne sont pas directement sous votre contrôle, comme l'augmentation des taux d'erreur dans une dépendance, vous devez réfléchir à l'impact que ces défaillances peuvent avoir sur votre charge de travail et à la manière de concevoir cette charge de travail pour répondre à ces défaillances. En fin de compte, vous devez vous concentrer surdéfaillances auxquelles vous pouvez répondreen utilisant une mesure d'atténuation qui est sous votre contrôle.

Ce guide décrit le cadre, puis explique comment identifier et documenter une charge de travail, comment appliquer le cadre à cette charge de travail et comment évaluer les stratégies d'atténuation des défaillances potentielles que vous rencontrez.

Table des matières

Présentation du cadre

Le cadre d'analyse de résilience a été développé en identifiant les propriétés de résilience souhaitées d'une charge de travail. Les propriétés souhaitées sont ce que vous voulez que le système soit vrai. La résilience étant généralement mesurée par la disponibilité, cinq propriétés caractérisent un système distribué à haute disponibilité : redondance, capacité suffisante, sortie en temps voulu, sortie correcte et isolation des pannes. Ces propriétés sont illustrées dans le schéma suivant.



- Redondance— La tolérance aux pannes est atteinte grâce à la redondance qui élimine les points de défaillance uniques (SPOF). La redondance peut aller des composants de rechange de votre

charge de travail à des répliques complètes de l'ensemble de votre pile d'applications. Lorsque vous envisagez la redondance de vos applications, il est important de prendre en compte le niveau de redondance fourni par l'infrastructure, les magasins de données et les dépendances que vous utilisez. Par exemple, Amazon DynamoDB et Amazon Simple Storage Service (Amazon S3) assurent la redondance en répliquant les données sur plusieurs zones de disponibilité d'une région, etAWS Lambdaexécute vos fonctions sur plusieurs nœuds de travail dans plusieurs zones de disponibilité. Pour chaque service que vous utilisez, tenez compte de ce qu'il fournit et de ce que vous devez concevoir.

- Capacité suffisante— Votre charge de travail nécessite des ressources suffisantes pour fonctionner comme prévu. Les ressources incluent la mémoire, les cycles du processeur, les threads, le stockage, le débit, les quotas de service et bien d'autres.
- Sortie en temps opportun— Lorsque les clients utilisent votre charge de travail, ils s'attendent à ce qu'elle remplisse la fonction prévue dans un délai raisonnable. À moins que le service ne fournisse un accord de niveau de service (SLA) pour la latence, leurs attentes sont généralement basées sur des preuves empiriques, c'est-à-dire sur leur propre expérience. Ceexpérience client moyenneest généralement considérée comme la latence médiane (P50) de votre système. Si votre charge de travail prend plus de temps que prévu, cette latence peut affecter l'expérience de vos clients.
- Sortie correcte— La sortie correcte du logiciel de votre charge de travail est nécessaire pour que celui-ci fournisse les fonctionnalités prévues. Un résultat incorrect ou incomplet peut être pire que l'absence de réponse du tout.
- Isolation des défauts— L'isolation des défauts limite l'étendue de l'impact au conteneur de défauts prévu en cas de défaillance. Cela garantit que des composants spécifiques de votre charge de travail tombent en panne ensemble tout en empêchant une défaillance de se répercuter sur d'autres composants involontaires. Cela permet également de limiter l'impact de votre charge de travail sur les clients. L'isolation des défauts est quelque peu différente des quatre propriétés précédentes, car elle accepte le fait qu'une défaillance s'est déjà produite mais qu'elle doit être maîtrisée. Vous pouvez isoler les défaillances de votre infrastructure, de vos dépendances et de vos fonctions logicielles.

Lorsqu'une propriété souhaitée n'est pas respectée, une charge de travail peut être ou perçue comme étant indisponible. Sur la base de ces propriétés de résilience souhaitées et de notre expérience de travail avec de nombreuxAWSclients, nous avons identifié cinq catégories de défaillances courantes : points de défaillance uniques, charge excessive, latence excessive, erreurs de configuration et bogues, et destin partagé, que nous avons abrégée en SEEMS. Ils fournissent

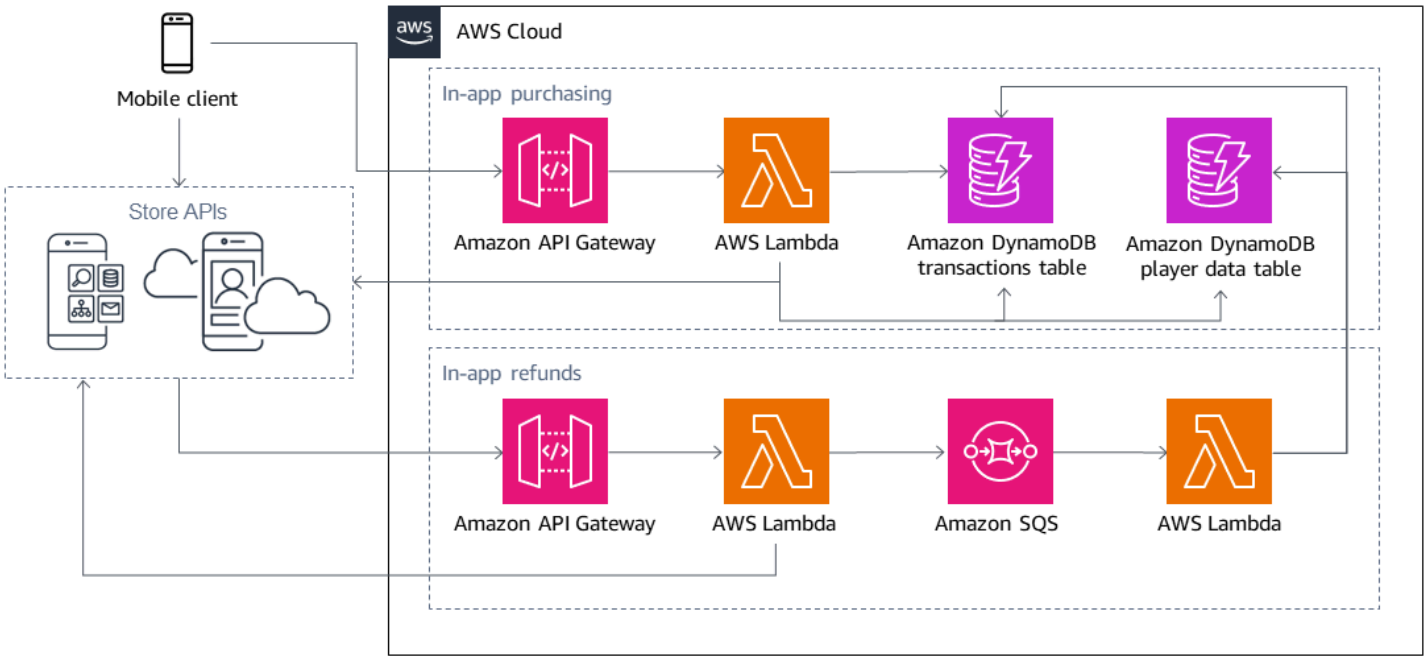
une méthode cohérente pour catégoriser les modes de défaillance potentiels et sont décrits dans le tableau suivant.

Catégorie de défaillance	Violation	Définition
Points de défaillance uniques (SPOF)	Redondance	Une défaillance d'un seul composant perturbe le système en raison de l'absence de redondance du composant.
Charge excessive	Capacité suffisante	La surconsommation d'une ressource due à une demande ou à un trafic excessif empêche la ressource de remplir la fonction attendue. Cela peut inclure l'atteinte de limites et de quotas, ce qui entraîne la limitation et le rejet des demandes.
Latence excessive	Sortie en temps opportun	La latence du traitement du système ou du trafic réseau dépasse le délai prévu, les objectifs de niveau de service (SLO) ou les accords de niveau de service (SLA).
Mauvaise configuration et bogues	Sortie correcte	Des bogues logiciels ou une mauvaise configuration du système entraînent une sortie incorrecte.
Un destin partagé	Isolation des défauts	Un défaut causé par l'une des catégories de défaillan ce précédentes dépasse les limites d'isolation des défauts prévues et se répercute sur

d'autres parties du système ou sur d'autres clients.

Comprendre la charge de travail

Pour appliquer le framework, commencez par comprendre la charge de travail que vous souhaitez analyser. Un schéma d'architecture du système fournit un point de départ pour documenter les détails les plus pertinents du système. Cependant, essayer d'analyser une charge de travail complète peut s'avérer complexe, car de nombreux systèmes comportent de nombreux composants et interactions. Nous vous recommandons plutôt de vous concentrer surtémoignages d'utilisateurs, qui sontexplications générales informelles des fonctionnalités du logiciel rédigées du point de vue de l'utilisateur final. Leur objectif est d'expliquer comment une fonctionnalité logicielle apporte de la valeur au client. Vous pouvez ensuite modéliser ces user stories à l'aide de diagrammes d'architecture et de diagrammes de flux de données afin de faciliter l'évaluation des composants techniques qui fournissent les fonctionnalités métier décrites. Par exemple, une solution d'achat de jeux mobile intégrée à une application peut comporter deux histoires utilisateur, « acheter des crédits intégrés à l'application » et « obtenir des remboursements intégrés à l'application », comme le montre le schéma suivant. (Cet exemple d'architecture montre comment décomposer un système en récits d'utilisateurs ; il n'est pas destiné à représenter une application hautement résiliente.)



Chaque user story comprend quatre composants communs : le code et la configuration, l'infrastructure, les magasins de données et les dépendances externes. Vos diagrammes doivent inclure tous ces composants et refléter les interactions entre les composants. Par exemple, si votre point de terminaison Amazon API Gateway est soumis à une charge excessive, réfléchissez à la manière dont cette charge se répercute sur d'autres composants du système, tels que votreAWS

Lambdafonctions ou tables Amazon DynamoDB. Le suivi de ces interactions vous permet de comprendre comment le mode de défaillance peut avoir un impact sur le récit de l'utilisateur. Vous pouvez capturer ce flux visuellement à l'aide d'un diagramme de flux de données ou à l'aide de simples flèches de flux dans un diagramme d'architecture, comme dans l'illustration précédente. Pour chaque composant, pensez à saisir des détails tels que le type d'informations transmises, les informations reçues, si la communication est synchrone ou asynchrone et les limites de défaillance franchies. Dans l'exemple, les tables DynamoDB sont partagées dans les deux récits utilisateur, comme le montrent les flèches indiquant que le composant Lambda de l'histoire des remboursements intégrés à l'application accède aux tables DynamoDB de l'historique des achats intégrés. Cela signifie qu'un échec causé par l'histoire utilisateur des achats intégrés à l'application peut se répercuter sur l'histoire utilisateur des remboursements intégrés à l'application en raison d'un destin partagé.

En outre, il est important de comprendre la configuration de base de chaque composant. La configuration de référence identifie les contraintes telles que le nombre moyen et maximal de transactions par seconde, la taille maximale d'une charge utile, le délai d'expiration du client et les quotas de service par défaut ou actuels pour la ressource. Si vous modélisez une nouvelle conception, nous vous recommandons de documenter les exigences fonctionnelles de la conception et de prendre en compte les limites. Cela vous permet de comprendre comment les modes de défaillance peuvent se manifester dans le composant.

Enfin, vous devez prioriser les user stories en fonction de la valeur commerciale qu'ils apportent. Cette hiérarchisation vous permet de vous concentrer d'abord sur les fonctionnalités les plus critiques de votre charge de travail. Vous pouvez ensuite concentrer votre analyse sur les composants de la charge de travail qui font partie du chemin critique pour cette fonctionnalité, et tirer parti de l'utilisation plus rapide de la structure. Au fur et à mesure du processus, vous pouvez examiner d'autres récits d'utilisateurs selon des priorités différentes.

Appliquer le cadre

La meilleure façon d'appliquer le cadre d'analyse de la résilience est de commencer par un ensemble de questions standard, organisées par catégorie de défaillance, que vous devez poser à propos de chaque composant de l'histoire utilisateur que vous analysez. Si certaines questions ne s'appliquent pas à tous les composants de votre charge de travail, utilisez les questions les plus pertinentes.

Vous pouvez aborder la réflexion sur les modes de défaillance sous deux angles :

- Quel est l'impact de la défaillance sur la capacité du composant à soutenir l'histoire de l'utilisateur ?
- Quel est l'impact de la défaillance sur les interactions du composant avec les autres composants ?

Par exemple, lorsque vous pensez aux magasins de données et à une charge excessive, vous pouvez penser aux modes de défaillance dans lesquels la base de données est soumise à une charge excessive et où les requêtes expirent. Vous pouvez également penser à la façon dont votre client de base de données risque de surcharger la base de données à force de tentatives ou de ne pas fermer les connexions à la base de données, épuisant ainsi le pool de connexions. Un autre exemple est un processus d'authentification, qui peut comporter plusieurs étapes. Vous devez réfléchir à la manière dont la défaillance d'une application d'authentification multifactorielle (MFA) ou d'un fournisseur d'identité tiers (IdP) pourrait avoir un impact sur un témoignage utilisateur dans ce système d'authentification.

Lorsque vous répondez aux questions suivantes, vous devez prendre en compte la source de l'échec. Par exemple, la surcharge a-t-elle été causée par une augmentation du nombre de clients ou par un opérateur humain qui a mis un trop grand nombre de nœuds hors service lors d'une activité de maintenance ? Vous pourriez être en mesure d'identifier plusieurs sources de défaillance dans chaque question, ce qui peut nécessiter différentes mesures d'atténuation. Lorsque vous posez les questions, notez les modes de défaillance potentiels que vous découvrez, les composants auxquels ils s'appliquent et la source de chaque défaillance.

Points de défaillance uniques

- Le composant est-il conçu pour la redondance ?
- Que se passe-t-il si le composant tombe en panne ?
- Votre application peut-elle tolérer la perte partielle ou totale d'une seule zone de disponibilité ?

Latence excessive

- Que se passe-t-il si ce composant subit une latence accrue, ou si un composant avec lequel il interagit présente une latence accrue (ou si le réseau est interrompu, par exemple en cas de réinitialisation du protocole TCP) ?
- Avez-vous correctement configuré les délais d'expiration avec une stratégie de nouvelle tentative ?
- Est-ce que vous échouez rapidement ou lentement ? Y a-t-il des effets en cascade, tels que l'envoi involontaire de tout le trafic vers une ressource altérée en raison d'une défaillance rapide ?
- Quelles sont les demandes les plus coûteuses adressées à ce composant ?

Charge excessive

- Qu'est-ce qui peut submerger ce composant ? Comment ce composant peut-il surpasser les autres composants ?
- Comment pouvez-vous éviter de gaspiller des ressources dans des tâches qui ne seront jamais couronnées de succès ?
- Disposez-vous d'un disjoncteur configuré pour le composant ?
- Est-ce que quelque chose peut créer un arriéré insurmontable ?
- Où ce composant peut-il présenter un comportement bimodal ?
- Quelles limites ou quels quotas de service peuvent être dépassés (y compris la capacité de stockage) ?
- Comment le composant évolue-t-il sous charge ?

Mauvaise configuration et bogues

- Comment éviter que les erreurs de configuration et les bogues ne soient déployés en production ?
- Pouvez-vous annuler automatiquement un mauvais déploiement ou détourner le trafic du conteneur d'erreurs dans lequel la mise à jour ou la modification a été déployée ?
- Quels garde-corps avez-vous mis en place pour éviter les erreurs des opérateurs ?
- Quels éléments (tels que les informations d'identification ou les certificats) peuvent expirer ?

Un destin partagé

- Quelles sont vos limites d'isolation des pannes ?

- Les modifications apportées aux unités de déploiement sont-elles au moins aussi minimes que prévu ?limites d'isolation des défautsmais idéalement plus petit, comme un environnement monobloc (une seule instance dans la limite d'isolation des pannes) ?
- Ce composant est-il partagé entre les user stories ou d'autres charges de travail ?
- Quels autres composants sont étroitement couplés à ce composant ?
- Que se passe-t-il si ce composant ou ses dépendances subissent une défaillance partielle ou grise ?

Après avoir posé ces questions, vous pouvez également utiliser SEEMS pour développer d'autres questions spécifiques à votre charge de travail et à chaque composant. Il est préférable d'utiliser SEEMS comme moyen structuré de réfléchir aux modes de défaillance et comme source d'inspiration lorsque vous effectuez une analyse de résilience. Il ne s'agit pas d'une taxonomie rigide. Ne perdez pas de temps à vous demander à quelle catégorie appartient un mode de défaillance en particulier, cela n'a pas d'importance. Quoiestl'important est que vous ayez pensé à l'échec et que vous l'ayez noté. Il n'y a pas de mauvaises réponses ; il est bénéfique de faire preuve de créativité et de sortir des sentiers battus. De plus, ne partez pas du principe qu'un mode de défaillance est déjà atténué ; incluez tous les modes de défaillance potentiels auxquels vous pouvez penser.

Il est peu probable que vous anticipiez tous les modes de défaillance potentiels lors de votre premier exercice. Plusieurs itérations du framework vous aident à générer un modèle plus complet, de sorte que vous n'avez pas à essayer de tout résoudre dès le premier passage. Vous pouvez exécuter l'analyse à une cadence régulière, hebdomadaire ou bihebdomadaire. À chaque session, concentrez-vous sur un mode de défaillance ou un composant spécifique. Cela peut vous aider à réaliser des progrès réguliers et progressifs dans l'amélioration de la résilience de votre charge de travail. Après avoir collecté une liste des modes de défaillance potentiels pour une user story, vous pouvez décider de la marche à suivre pour y remédier.

Atténuer les défaillances potentielles

Maintenant que les composants d'un user story présentent des défaillances potentielles, vous pouvez vous concentrer sur les mesures d'atténuation. Tout d'abord, examinez les compromis potentiels par rapport à l'impact potentiel et à la probabilité de chaque défaillance que vous avez découverte. Déterminez ensuite le niveau d'observabilité requis et sélectionnez une stratégie d'atténuation. Les compromis devraient inclure les efforts visant à déterminer le bon niveau d'observabilité et la stratégie d'atténuation. Enfin, déterminez la bonne cadence pour effectuer des analyses de résilience régulières.

Sections

Comprendre les compromis et les risques

Les architectures résilientes doivent utiliser une poignée de mécanismes simples, fiables et éprouvés pour répondre aux défaillances. Pour atteindre les niveaux de résilience les plus élevés, les charges de travail doivent automatiquement détecter et récupérer après autant de modes de défaillance que possible. Cela nécessite un investissement important dans la réalisation d'analyses de résilience. Cela signifie que pour atteindre des niveaux de résilience plus élevés, il faut faire des compromis. Cependant, au fur et à mesure que vous faites des compromis, vous atteignez un point où les rendements diminuent par rapport à vos objectifs de résilience. Voici les compromis les plus courants :

- Coût — Des composants redondants, une meilleure observabilité, des outils supplémentaires ou une utilisation accrue des ressources entraîneront une augmentation des coûts.
- Complexité du système — La détection et la réponse aux modes de défaillance, y compris les solutions d'atténuation, ainsi que le fait de ne pas utiliser de services gérés accroissent la complexité du système.
- Effort d'ingénierie — Des heures de développement supplémentaires sont nécessaires pour créer des solutions permettant de détecter les modes de défaillance et d'y répondre.

- Frais opérationnels — La surveillance et l'exploitation d'un système qui gère un plus grand nombre de modes de défaillance peuvent entraîner une surcharge opérationnelle, en particulier lorsque vous ne pouvez pas utiliser les services gérés pour atténuer des modes de défaillance spécifiques.
- Latence et cohérence — La création de systèmes distribués qui favorisent la disponibilité nécessite des compromis en termes de cohérence et de latence, comme décrit dans le théorème PACELC.

Lorsque vous examinez les mesures d'atténuation pour les modes de défaillance identifiés dans l'histoire de l'utilisateur, réfléchissez aux compromis que vous devez faire. Comme pour la sécurité, la résilience est un problème d'optimisation. Vous devez décider d'éviter, d'atténuer, de transférer ou d'accepter les risques posés par la défaillance identifiée. Il existe peut-être certains modes de défaillance que vous pouvez éviter, un ensemble que vous acceptez et d'autres que vous pouvez transférer. Vous pouvez choisir d'atténuer la plupart des modes de défaillance que vous avez identifiés. Pour déterminer l'approche à adopter, effectuez une évaluation en vous posant deux questions : Quelle est la probabilité que la défaillance se produise ? Quel est l'impact sur la charge de travail si elle se produit ?

La probabilité correspond à la probabilité qu'un événement se produise. Par exemple, si l'histoire utilisateur contient un composant qui fonctionne sur une seule instance Amazon Elastic Compute Cloud (Amazon EC2), le composant peut être perturbé à un moment donné pendant le

fonctionnement du système, peut-être en raison de procédures de correction ou d'erreurs du système d'exploitation. Par ailleurs, une base de données gérée par Amazon Relational Database Service (Amazon RDS) qui synchronise les données entre ses instances principales et secondaires a peu de chances de devenir totalement indisponible.

L'impact est une estimation des dommages qu'un événement peut causer. Elle doit être évaluée à la fois du point de vue financier et du point de vue de la réputation, et elle est relative à la valeur des histoires d'utilisateurs qu'elle a un impact. Par exemple, une base de données surchargée peut avoir un impact significatif sur la capacité d'un système de commerce électronique à accepter de nouvelles commandes. Cependant, la perte d'une seule instance sur un parc de 20 instances derrière un équilibreur de charge n'aurait probablement que très peu d'impact.

Vous pouvez comparer les réponses à ces questions au coût des compromis que vous devez faire pour atténuer les risques. Lorsque vous considérez ces informations au regard de votre seuil de risque et de vos objectifs de résilience, elles éclairent votre décision quant aux modes de défaillance que vous prévoyez d'atténuer activement.

Observabilité du mode de défaillance

Pour atténuer un mode de défaillance, vous devez d'abord détecter qu'il impacte actuellement ou est sur le point d'avoir un impact sur votre charge de travail. Une atténuation n'est efficace que si un signal indique qu'une action doit être prise. Cela signifie qu'une partie de la création de toute mesure d'atténuation implique, à tout le moins, de vérifier que vous avez ou que vous êtes en train de développer l'observabilité nécessaire pour détecter l'impact de la panne.

Vous devez considérer les symptômes observables du mode de défaillance en deux dimensions :

- Quels sont les principaux indicateurs qui indiquent que le système est sur le point d'atteindre un point tel qu'un impact pourrait bientôt être observé ?
- Quels sont les indicateurs de retard qui peuvent montrer l'impact du mode de défaillance le plus rapidement possible après son apparition ?

Par exemple, une défaillance de charge excessive appliquée à un élément de base de données peut avoir comme indicateur principal le nombre de connexions. L'augmentation constante du nombre de connexions est un indicateur indiquant que la base de données pourrait bientôt dépasser la limite de connexions. Vous pouvez donc prendre des mesures, telles que mettre fin aux connexions les moins récemment utilisées, pour réduire le nombre de connexions. L'indicateur de retard indique lorsque la limite de connexion à la base de données a été dépassée et que les erreurs de connexion à la base

de données augmentent. Outre la collecte de mesures relatives aux applications et à l'infrastructure, pensez à recueillir des indicateurs de performance clés (KPI) pour détecter les défaillances qui ont un impact sur votre expérience client.

Dans la mesure du possible, nous vous recommandons d'inclure les deux types d'indicateurs dans votre stratégie d'observabilité. Dans certains cas, il se peut que vous ne puissiez pas créer d'indicateurs avancés, mais vous devez toujours prévoir un indicateur de retard pour chaque défaillance que vous souhaitez atténuer. Pour choisir la bonne solution d'atténuation, vous devez également déterminer si un indicateur avancé ou différé a détecté la défaillance. Prenons l'exemple d'un pic soudain de trafic vers votre site Web. Vous ne verrez probablement qu'un indicateur de retard. Dans ce cas, la mise à l'échelle automatique à elle seule n'est peut-être pas la meilleure solution, car le déploiement de nouvelles ressources prend du temps, tandis que la régulation peut empêcher la surcharge presque immédiatement et donner à votre application le temps de s'adapter ou de réduire la charge. À l'inverse, pour une augmentation progressive du trafic, vous verrez un indicateur avancé. Dans ce cas, la régulation ne serait pas appropriée, car vous avez le temps de réagir en dimensionnant automatiquement votre système.

Stratégies d'atténuation communes

Pour commencer, pensez à utiliser des mesures d'atténuation préventives pour éviter que le mode de défaillance n'ait un impact sur l'histoire de l'utilisateur. Vous devriez alors réfléchir à des mesures correctives d'atténuation. Les mesures d'atténuation correctives aident le système à s'auto-guérir ou à s'adapter aux conditions changeantes. Voici une liste des mesures d'atténuation courantes pour chaque catégorie de défaillance qui correspondent aux propriétés de résilience.

Catégorie de défaillance	Propriétés de résilience souhaitées	Atténuations
Points de défaillance uniques (SPOF)	Redondance et tolérance aux pannes	• Mettez en œuvre la redondance, par exemple en utilisant plusieurs instances EC2 derrière Elastic Load Balancing (ELB). • Supprimez les dépendances sur le plan de contrôle de service AWS global et prenez uniquement les

		dépendances sur les plans de données de service globaux. • Utilisez la dégradation progressive lorsqu'une ressource n'est pas disponible, afin que votre système soit statiquement stable jusqu'à un point de défaillance unique.
Charge excessive	Capacité suffisante	• Les principales stratégies d'atténuation sont la limitation du débit, le délestage et la priorisation des tâches, le travail constant, les retards exponentiels et les nouvelles tentatives avec instabilité ou absence de nouvelle tentative, la maîtrise du petit service, la gestion de la profondeur des files d'attente, la mise à l'échelle automatique, la prévention des caches froides et les disjoncteurs. • Vous devez également tenir compte de votre plan de capacité et réfléchir aux futures limites de capacité et de dimensionnement, liées à la fois aux ressources AWS et aux limites de votre système, que vous pourriez atteindre.

Latence excessive	Sortie en temps opportun	• Mettez en œuvre des délais d'attente configurés de manière appropriée ou des délais d'expiration adaptatifs (en modifiant les valeurs de délai d'attente en fonction des conditions de latence actuelles et prévues afin de permettre à une dépendanc e lente de progresser au lieu de renoncer à des demandes lentes). • Mettez en œuvre un ralentissement exponenti el et réessayez en cas de latence sur des itinérair es spécifiques, en utilisant des technologies telles que le TCP multipath lorsque vous vous connectez à des services cloud depuis des environnements sur site et que vous subissez une latence sur des itinérair es spécifiques, en utilisant des interactions asynchron es avec des systèmes faiblement couplés, en mettant en cache et en évitant de gaspiller du travail.

Mauvaise configuration et bogues	Sortie correcte	• Le principal moyen de détecter les erreurs fonctionnelles répétables dans les logiciels consiste à effectuer des tests rigoureux au moyen de mécanismes tels que l'analyse statique, les tests unitaires, les tests d'intégration, les tests de régression, les tests de charge et les tests de résilience. • Mettez en œuvre des stratégies telles que l'infrastructure sous forme de code (IaC) et l'automatisation de l'intégration continue et de la livraison continue (CI/CD) pour atténuer les menaces liées aux erreurs de configuration. • Utilisez des techniques de déploiement telles que les déploiements à boîtier unique, les déploiements Canary, les déploiements fractionnés alignés sur les limites d'isolation des pannes ou les déploiements bleu/vert pour réduire les erreurs de configuration et les bogues.

Un destin partagé	Isolation des défauts	• Implémentez la tolérance aux pannes dans votre système et utilisez des limites logiques et physiques d'isolation des pannes, telles que plusieurs clusters de calcul ou de conteneur s, plusieurs comptes AWS, plusieurs principaux AWS Identity and Access Management (IAM), plusieurs zones de disponibi lité, voire plusieurs. Régions AWS • Des techniques telles que les architectures basées sur les cellules et le shuffle sharding peuvent également améliorer l'isolation des pannes. • Envisagez des modèles tels que le couplage lâche et la dégradation progressi ve pour éviter les défaillan ces en cascade. Lorsque vous hiérarchisez les user stories, vous pouvez également utiliser cette hiérarchisation pour faire la distinction entre les user stories qui sont essentiel les à la fonction commercia le principale et les user stories qui peuvent être dégradées avec élégance. Par exemple, sur un site

	de commerce électroni que, vous ne voudriez pas qu'une altération du widget de promotions du site Web ait un impact sur la capacité de traiter les nouvelles commandes.

Bien que certaines de ces mesures d'atténuation nécessitent un minimum d'efforts pour être mises en œuvre, d'autres (telles que l'adoption d'une architecture basée sur les cellules pour une isolation prévisible des pannes et un minimum de défaillances à destin partagé) peuvent nécessiter une refonte de l'ensemble de la charge de travail et pas seulement des composants d'une histoire utilisateur en particulier. Comme indiqué précédemment, il est important d'évaluer la probabilité et l'impact du mode de défaillance par rapport aux compromis que vous devez faire pour l'atténuer.

Outre les techniques d'atténuation applicables à chaque catégorie de mode de défaillance, vous devez réfléchir aux mesures d'atténuation nécessaires à la restauration de l'histoire utilisateur ou de l'ensemble du système. Par exemple, une panne peut interrompre un flux de travail et empêcher l'écriture des données vers les destinations prévues. Dans ce cas, vous aurez peut-être besoin d'outils opérationnels pour relancer le flux de travail ou corriger manuellement les données. Vous devrez peut-être également intégrer un mécanisme de point de contrôle à votre charge de travail pour éviter les pertes de données en cas de défaillance. Il se peut également que vous deviez créer un cordon andon pour interrompre le flux de travail et arrêter d'accepter de nouvelles tâches afin d'éviter de nouveaux dommages. Dans ces cas, vous devez réfléchir aux outils opérationnels et aux glissières de sécurité dont vous avez besoin.

Enfin, vous devez toujours partir du principe que les humains commettront des erreurs lors de l'élaboration de votre stratégie d'atténuation. Bien que DevOps les pratiques modernes visent à automatiser les opérations, les humains doivent toujours interagir avec vos charges de travail pour diverses raisons. Une action humaine incorrecte peut entraîner une défaillance dans l'une des catégories SEEMS, par exemple en supprimant un trop grand nombre de nœuds pendant la maintenance et en provoquant une surcharge, ou en définissant un indicateur de fonctionnalité de manière incorrecte. Ces scénarios constituent un véritable échec en matière de garde-corps préventifs. Une analyse des causes profondes ne doit jamais aboutir à la conclusion qu' « un humain a commis une erreur ». Il devrait plutôt aborder les raisons pour lesquelles des erreurs étaient possibles au départ. Par conséquent, votre stratégie d'atténuation doit tenir compte de la manière dont les opérateurs humains peuvent interagir avec les composants de la charge de travail et de la

manière de prévenir ou de minimiser l'impact des erreurs des opérateurs grâce à des glissières de sécurité.

Amélioration continue

La résilience est un processus continu. Au cours du cycle de vie de votre système, l'environnement dans lequel il fonctionne évoluera. Pour garantir la résilience de votre système, vous devez intégrer le cadre dans vos révisions opérationnelles et architecturales périodiques. Il se peut que vous trouviez de nouveaux modes de défaillance que vous n'aviez pas identifiés la première fois, ou que vous puissiez mettre en place de nouvelles mesures d'atténuation ou des mesures d'atténuation inédites. L'analyse de résilience doit être un processus itératif et non un exercice ponctuel.

Vous devez tester empiriquement vos stratégies d'atténuation à l'aide de processus tels que l'ingénierie du chaos ou les jours de jeu pour vérifier qu'elles fonctionnent comme prévu. Si vous ne disposez pas d'un mécanisme de test rigoureux, vous ne serez pas sûr que les mesures d'atténuation fonctionneront comme prévu lorsque vous en aurez besoin. Au cours de l'analyse de résilience, vous pouvez déterminer qu'un mode de défaillance est déjà géré par une atténuation spécifique, mais il est également important de tester ces hypothèses. Vous devez tester à la fois les mesures d'atténuation existantes et les nouvelles mesures d'atténuation créées à l'aide du cadre d'analyse de résilience.

Vous devez également évaluer dans quelle mesure vous avez effectué l'analyse par le biais de rétrospectives d'équipe. Est-ce que tout le monde savait sur quoi ils travaillaient pendant l'analyse ? Le nombre de modes de défaillance que vous avez découverts grâce à l'analyse de résilience correspondait-il aux attentes de l'équipe ? Pourriez-vous identifier des mesures d'atténuation pour tous les modes de défaillance que vous avez découverts ? L'équipe a-t-elle trouvé le processus utile ? Pensez-vous que cela améliorera la résilience de votre charge de travail ?

Lorsque de véritables défaillances se produisent et ont un impact sur la disponibilité de votre charge de travail, enregistrez le mode de défaillance spécifique, les composants impliqués dans la défaillance et le schéma d'atténuation utilisé. Rendez ces métadonnées consultables dans votre outil d'analyse post-incident afin de déterminer les modes de défaillance et les composants sur lesquels vous devez vous concentrer à l'avenir. Tout au long de ce processus, vous pouvez impliquer votre équipe chargée des AWS comptes et les architectes de solutions.

Conclusion et ressources

Ce guide présente un cadre permettant d'effectuer une analyse de résilience de manière continue et cohérente. Ce cadre vous aide à identifier comment les points de défaillance uniques, la charge excessive, la latence excessive, les erreurs de configuration et les bogues, ainsi que le destin partagé peuvent affecter les composants de votre charge de travail. L'identification de ces modes de défaillance vous aide à déterminer une stratégie d'atténuation appropriée dans le cadre de la création d'une architecture axée sur la restauration.

Pour en savoir plus sur l'analyse de résilience, consultez les liens suivants :

- Cadre du cycle de vie de résilience (AWSdirectives prescriptives)
- Solutions pour la résilience (bibliothèque de AWS solutions)
- Vers une résilience continue (Adrian Hornsby, The Cloud Architect, 24 mars 2021)

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide. Si vous souhaitez être informé des prochaines mises à jour, vous pouvez vous abonner à unfil RSS.

Modification	Description	Date
Publication initiale	—	5 septembre 2023

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 :

- Refactorisation/réarchitecture : transférez une application et modifiez son architecture en tirant pleinement parti des fonctionnalités natives cloud pour améliorer l'agilité, les performances et la capacité de mise à l'échelle. 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 Amazon Aurora Édition compatible avec PostgreSQL.
- 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 in the Cloud. AWS
- 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. Ce scénario de migration est spécifique à VMware Cloud on AWS, qui prend en charge la compatibilité des machines virtuelles (VM) et la portabilité de la charge de travail entre votre environnement sur site et AWS. Vous pouvez utiliser les technologies VMware Cloud Foundation à partir de vos centres de données sur site lorsque vous migrez votre infrastructure vers VMware Cloud on AWS. Exemple : déplacez l'hyperviseur hébergeant votre base de données Oracle vers VMware Cloud on. 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

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 la synchronisation des bases de données source et cible est maintenue, mais seule la base de données source gère les transactions provenant de la connexion d'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.

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 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 AWS Identity and Access Management (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.

déploiement bleu/vert

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 Implementation break-glass procedures dans le guide Well-Architected AWS .

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.

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 AWS Cloud Enterprise Strategy.

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 le 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
- Réinvention : 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 on the AWS Cloud Enterprise Strategy blog. 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 ou AWS CodeCommit. Chaque version du code est appelée branche. Dans une structure de microservice, chaque référentiel est consacré à une seule fonctionnalité. Un seul pipeline CI/CD 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, AWS

Panorama propose des appareils qui ajoutent des CV aux réseaux de caméras locaux, et Amazon SageMaker 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 source, de génération, de test, intermédiaire 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 Framework. 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.

defense-in-depth

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 defense-in-depth approche 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 Disaster Recovery of Workloads on AWS : Recovery in the Cloud in the 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 ouvrage Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston : Addison-Wesley Professional, 2003). Pour plus d'informations sur l'utilisation du design piloté par domaine avec le modèle de figuier étrangleur, veuillez consulter Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and 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.

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.

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 selon lequel les octets sont stockés dans la mémoire de l'ordinateur. Les systèmes de poids fort stockent d'abord l'octet le plus significatif. Les systèmes de poids faible 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 Comptes AWS ou à AWS Identity and Access Management (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 pipeline CI/CD, 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.

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 via 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.

G

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.

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 GuardDuty AWS Trusted Advisor, d'Amazon Inspector et de AWS Lambda contrôles personnalisés.

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 de l'historien

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.

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 d'IA/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, veuillez consulter Machine learning model interpretability with 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.

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.

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 qui AWS gère 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.

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 cadre AWS Well-Architected.

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)

Protocole de communication léger machine-to-machine (M2M), basé sur le modèle de publication/abonnement, 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

Équipes interfonctionnelles qui rationalisent la migration des charges de travail grâce à des approches automatisées et agiles. Les équipes de Migration Factory comprennent généralement les opérations, les analystes commerciaux et les propriétaires, les ingénieurs de migration, les développeurs et les DevOps 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)

Un 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

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, veuillez consulter Evaluating modernization readiness for applications in the 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)

Un protocole de communication machine-to-machine (M2M) pour l'automatisation industrielle. L'OPC-UA fournit une norme d'interopérabilité avec des schémas de cryptage, 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 Operational Readiness Reviews (ORR) dans le AWS Well-Architected Framework.

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). L'OAC prend en charge tous les compartiments S3 dans leur ensemble Régions AWS, le chiffrement côté serveur avec AWS KMS (SSE-KMS) et les requêtes dynamiques `PUT` adressées au compartiment S3. `DELETE`

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.

OU

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.

politique

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. Pour plus d'informations, veuillez consulter Enabling data persistence in microservices.

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 de l'ingénierie des systèmes qui prend en compte la confidentialité tout au long du processus d'ingénierie.

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.

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.

publier/souscrire (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).

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. Cela permet de déterminer ce qui est considéré comme une perte de données acceptable entre le dernier point de restauration 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.

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 d' 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, consultez la section Secret dans la documentation de Secrets Manager.

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 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.

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.

split-and-seed modèle

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 obtenir un exemple d'application de ce modèle, veuillez consulter Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and 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.

T

balises

Des paires clé-valeur 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.

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.

Pour plus d'informations, veuillez consulter le guide Quantifying uncertainty in deep learning systems.

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

Consultez le cadre de qualification des charges de travail AWS.

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.

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.