Création d'une architecture de haute disponibilité et de reprise après sinistre avec des méthodes natives et hybrides pour les bases de données Microsoft SQL Server sur Amazon EC2

AWS Conseils prescriptifs

AWS Conseils prescriptifs: Création d'une architecture de haute disponibilité et de reprise après sinistre avec des méthodes natives et hybrides pour les bases de données Microsoft SQL Server sur Amazon EC2

Table of Contents

Création d'une architecture de haute disponibilité et de reprise après sinistre avec des méthodes natives et hybrides pour les bases de données Microsoft SQL Server sur Amazon EC2

Ram Yellapragada et Alysia Tran, Amazon Web Services ()AWS

Février 2022 (historique du document)

Microsoft SQL Server propose de nombreuses options natives pour prendre en charge la haute disponibilité (HA) et la reprise après sinistre (DR), afin de garantir la continuité des activités de vos charges de travail de base de données. Ce guide décrit une configuration idéale pour SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2) dans le cloud Amazon Web Services (AWS). Le réhébergement de SQL Server sur Amazon EC2 fournit un système autogéré dans lequel vous pouvez conserver un contrôle total sur les opérations et la configuration de la base de données.

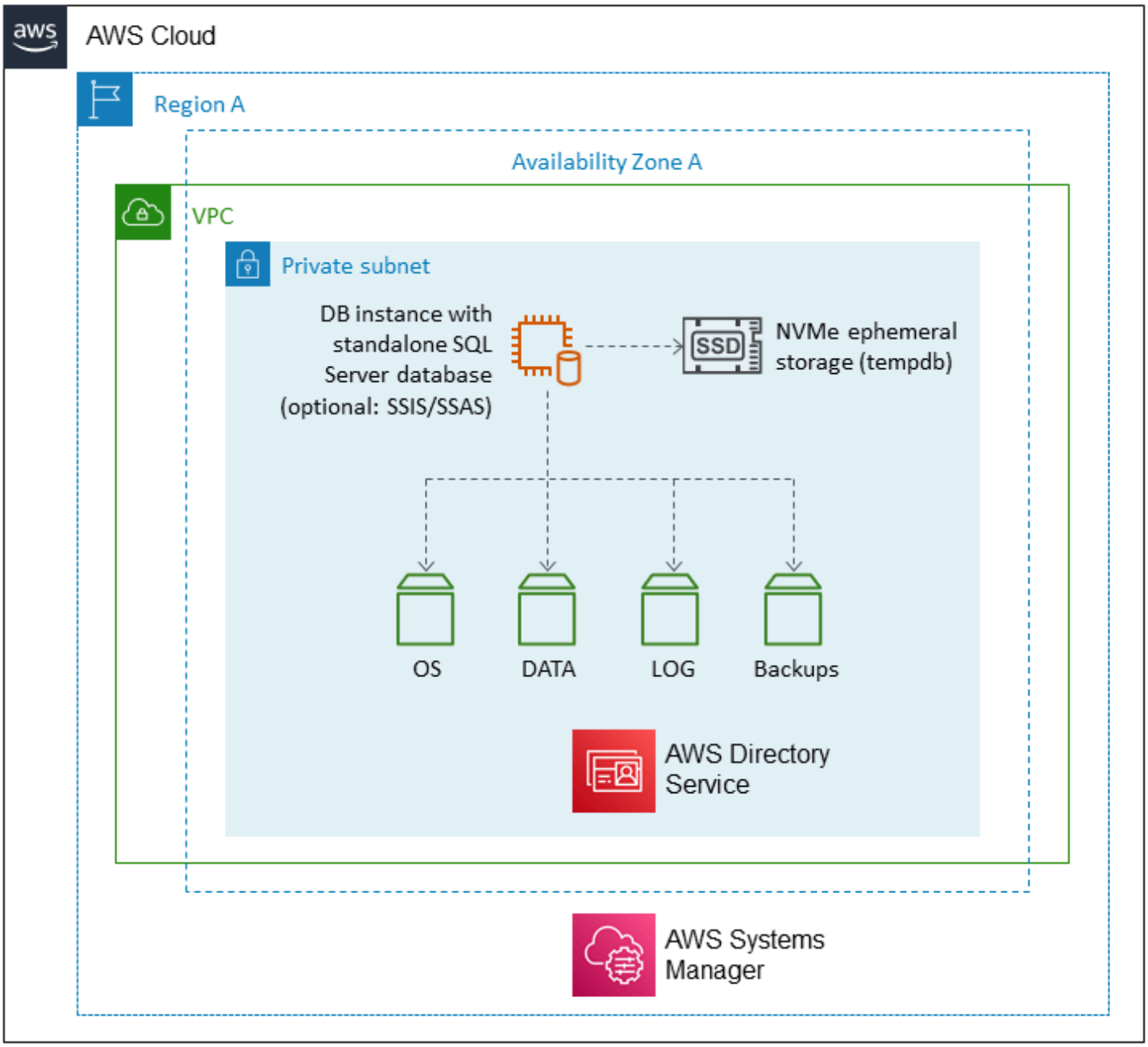
Le guide décrit les options hybrides HA/DR de SQL Server qui incluent divers AWS services et infrastructures, et fournit des conseils sur les composants et les paramètres de l'infrastructure, notamment les classes d'instances, les options de stockage, la configuration et les HA/DR setup. This document also explains how a given HA/DR strategy might fit into an example use case that has specific recovery time objective (RTO) and recovery point objective (RPO) requirements, and covers a few recovery scenarios, including relevant architecture diagrams. This guide doesn’t provide solutions designed for specific applications or requirements. It presents some HA/DR options basées sur le RTO et le RPO, afin que vous puissiez choisir une architecture qui répond à vos besoins.

En outre, à titre d'exercice de dimensionnement, le guide définit les options HA/DR pour une charge de travail de traitement des transactions en ligne (OLTP) typique de SQL Server et fournit une side-by-side comparaison de ces options. Pour une discussion sur le réhébergement de SQL Server sur AWS, consultez la section Amazon EC2 pour SQL Server du guide Migration des bases de données Microsoft SQL Server vers le AWS cloud. Pour plus d'informations sur les autres options de migration, consultez la section Stratégies de migration de base de données SQL Server de ce guide. Pour en savoir plus, consultez la Prochaines étapes et ressources section.

Architecture à EC2 nœud unique de SQL Server sur Amazon

Le schéma suivant illustre une architecture recommandée pour un serveur SQL à nœud unique sur Amazon Elastic Compute Cloud (Amazon EC2) avant d'ajouter la prise en charge de la haute disponibilité (HA) et de la reprise après sinistre (DR).

Dans cette architecture, la base de données SQL Server est déployée sur une EC2 instance, à l'aide d'une Amazon Machine Image (AMI) pour SQL Server et de volumes distincts pour le système d'exploitation, les données, le journal et les sauvegardes. Le stockage express (NVMe) en mémoire non volatile est attaché directement à l' EC2 instance et utilisé pour la base de données tempdb de SQL Server. AWS Directory Service est utilisé pour configurer l'authentification Windows pour la base de données SQL Server. Vous pouvez également l'utiliser AWS Systems Manager pour détecter et installer les correctifs et les mises à jour de SQL Server.



Le tableau suivant récapitule les recommandations relatives à la configuration de cette architecture. Ces recommandations sont examinées en détail dans les sections qui suivent.

Type d'instance/AMI	• Type d'instance optimisé pour Amazon Elastic Block Store (Amazon EBS) en termes de performances • NVMe par exemple le stockage (temporaire)

	• Amazon EC2 AMIs pour SQL Server
Édition SQL Server	• Édition SQL Server Developer (hors production) • Éditions SQL Server Standard et Enterprise (production)
Type de stockage	• Amazon EBS • NVMe(tempdb) (gp2//io1) io2
Volumes	• Système d'exploitation • DATA • LOG • tempdb • Espace libre pour le stockage et le télécharg ement des sauvegardes
Options de reprise après sinistre	• Amazon EC2 • Instantanés Amazon EBS • Sauvegardes natives de SQL Server

Types d'instance

AWS propose une sélection de classes d'instances pour vos charges de travail SQL Server. Vous pouvez choisir entre des types optimisés pour le calcul, pour la mémoire, pour le stockage, pour usage général, etc., en fonction de la charge de travail attendue sur le serveur de base de données, de la version, des HA/DR options, des cœurs requis et des considérations relatives aux licences. Nous vous recommandons de choisir les types d'instances optimisés pour Amazon EBS pour SQL Server. Ils offrent le meilleur débit avec des volumes EBS attachés dans un réseau dédié, ce qui est essentiel pour les charges de travail SQL Server susceptibles de nécessiter de lourdes exigences d'accès aux données. Pour les charges de travail de base de données standard, vous pouvez exécuter des classes d'instance optimisées pour la mémoire telles que R5, R5b, R5d et R5n. Vous pouvez également inclure le stockage d'instance ou le NVMe stockage. Elles sont toutes deux idéales pour tempdb et offrent des performances équilibrées pour les charges de travail des bases de données.

Pour les charges de travail critiques, l'instance z1d à hautes performances est optimisée pour les charges de travail impliquant des coûts de licence élevés, telles que SQL Server. L'instance z1d est construite avec un processeur Intel Xeon Scalable personnalisé qui fournit une fréquence turbo intégrale soutenue allant jusqu'à 4.0 GHz, ce qui est nettement plus rapide que les autres instances. Pour les charges de travail nécessitant un traitement séquentiel plus rapide, vous pouvez exécuter moins de cœurs avec une instance z1d et obtenir des performances identiques ou supérieures à celles des autres instances dotées de plus de cœurs.

Amazon propose également un service dédié à AMIsSQL Server sur Microsoft Windows Server pour vous aider à héberger les dernières éditions de SQL Server sur Amazon EC2.

Stockage

Certains types d'instances proposent des volumes de stockage d' NVMe instance. NVMe est une option de stockage temporaire (éphémère). Ce stockage est directement attaché à l' EC2 instance. Bien que le NVMe stockage soit temporaire et que les données soient perdues au redémarrage, il offre des performances optimales. Il convient donc à la base de données tempdb de SQL Server, qui présente des modèles d'accès aux données élevés I/O et aléatoires. L'utilisation d'un magasin d' NVMe instance pour tempdb est gratuite. Pour obtenir des conseils supplémentaires, consultez la section Placer tempdb dans un magasin d'instance du guide Meilleures pratiques pour le déploiement de SQL Server sur Amazon. EC2

Amazon EBS est une solution de stockage durable qui répond aux exigences de SQL Server en matière de stockage rapide et disponible. Microsoft recommande de séparer les volumes de données et de journaux pour des performances optimales. Les raisons de cette séparation sont notamment les suivantes :

- Différentes méthodes d'accès aux données. Les volumes de données utilisent un accès aléatoire aux données par traitement des transactions en ligne (OLTP), tandis que les volumes de journaux utilisent un accès série.
- De meilleures options de restauration. La perte d'un volume n'affecte pas l'autre volume et contribue à la restauration des données.
- Différents types de charge de travail. Les volumes de données sont destinés aux charges de travail OLTP, tandis que les volumes de journaux ciblent les charges de travail de traitement analytique en ligne (OLAP).

- Différentes exigences de performance. Les volumes de données et de journaux ont des exigences différentes en termes d'IOPS et de latence, des débits minimaux et des critères de performance similaires.

Pour sélectionner le bon type de volume Amazon EBS, vous devez analyser les méthodes d'accès, les IOPS et le débit de votre base de données. Collectez des statistiques à la fois pendant les heures de travail normales et pendant les pics d'utilisation. SQL Server utilise des extensions pour stocker les données. L'unité atomique de stockage dans SQL Server est une page d'une taille de 8 Ko. Huit pages physiquement contiguës constituent une étendue d'une taille de 64 Ko. Par conséquent, sur une machine SQL Server, la taille de l'unité d'allocation NTFS pour l'hébergement des fichiers de base de données SQL (y compris tempdb) doit être de 64 Ko. Pour savoir comment vérifier la taille d'allocation NTFS de vos disques, consultez le guide Meilleures pratiques pour le déploiement de SQL Server sur Amazon EC2.

Le choix du volume EBS dépend de la charge de travail, c'est-à-dire du fait que la base de données soit intensive en lecture ou en écriture, qu'elle nécessite des IOPS élevées, un stockage d'archives, etc. Le tableau suivant présente un exemple de configuration.

Ressource Amazon EBS	Type	Description
Disque du système d'exploitation	`gp3`	Stockage à usage général.
disque de données	`io1`/`io2`	Stockage nécessitant beaucoup d'écriture.
Disque LOG	`gp3` ou `io2`	Stockage à usage général pour les charges de travail intensives.
Disque de sauvegarde	`st1`	Stockage d'archives moins coûteux. Pour de meilleures performances, les sauvegardes peuvent également être stockées sur un disque plus rapide si elles sont copiées régulièrement sur Amazon

Ressource Amazon EBS	Type	Description
		Simple Storage Service (Amazon S3).

Considérations relatives à Amazon EBS et Amazon S3

Le tableau suivant présente une comparaison entre Amazon EBS et Amazon S3 pour le stockage. Utilisez ces informations pour comprendre les différences entre les deux services et pour choisir l'approche la mieux adaptée à votre cas d'utilisation.

Service	Disponibilité	Durabilité	Remarques
Amazon EBS	• Tous les types de volumes EBS offrent des fonctionnalités de capture instantanée durables et sont conçus pour une disponibilité de 99,999 %. • Vous pouvez utiliser des instantanés pour approvisionner de nouvelles instances dans différentes AWS régions en cas de sinistre.	• Les données du volume EBS sont répliquées sur plusieurs serveurs dans une seule zone de disponibilité afin d'éviter toute perte de données due à la défaillance d'un composant. • Les volumes EBS sont conçus pour un taux de défaillance annuel (AFR) compris entre 0,1 et 0,2 %, la défaillance faisant référence à une perte totale ou partielle du volume, en fonction de la taille et des	• Une instance optimisée pour Amazon EBS utilise une pile de configuration optimisée et fournit une bande passante dédiée supplémentaire pour Amazon EBS I/O. This optimization provides the best performance for your EBS volumes by minimizing contention between Amazon EBS I/O et le reste du trafic provenant de votre instance.

Service	Disponibilité	Durabilité	Remarques
		performances du volume.	• Les restaurations rapides d'instant anés sont prises en charge pour un maximum de 50 instantanés en même temps. Vous devez activer cette fonctionn alité de manière explicite pour chaque instantané. • Une instance optimisée pour Amazon EBS offre des performan ces provisionnées complètes lors de l'initialisation, de sorte qu'aucun temps de préchauff age n'est nécessair e.

Service	Disponibilité	Durabilité	Remarques

Amazon S3	• Hautement disponible.
	• Conçu pour une disponibilité de 99,99 % sur une année donnée.
	• Plusieurs classes de stockage sont disponibles, telles que S3 Standard et S3 Standard-Infrequent Access ((S3 Standard-IA)). Vous pouvez déplacer les fichiers de sauvegarde vers une classe de stockage en fonction d'une période de conservation.
		• Amazon S3, Amazon Glacier et S3 Glacier Deep Archive sont conçus pour une durabilité de 99,999999 999 % (11 neuf). Amazon S3 et Amazon Glacier offrent tous deux une sauvegarde fiable des données, avec la réplication d'objets dans au moins trois zones de disponibilité géographiquement dispersées.	• Vous pouvez utiliser Amazon S3 pour des sauvegardes à long terme au niveau des fichiers de SQL Server (y compris les sauvegardes complètes et les journaux de transactions).
			• Amazon S3 prend en charge :
			  • Contrôle du temps de réplication (RTC)
			  • Réplication entre régions via la gestion du cycle de vie S3 et AWS Backup
			  • Hiérarchisation intelligente
			• Amazon S3 fournit le stockage le moins coûteux. Des frais de transfert de données entre régions s'appliquent.

SQL Server sur le serveur de fichiers Amazon FSx pour Windows

Le serveur de fichiers Amazon FSx pour Windows fournit des performances rapides avec un débit de base pouvant atteindre 2 GB/second par système de fichiers, des centaines de milliers d'IOPS et des latences constantes inférieures à la milliseconde. Pour fournir les bonnes performances à vos instances SQL Server, vous pouvez choisir un niveau de débit indépendant de la taille de votre système de fichiers. Des niveaux de capacité de débit plus élevés s'accompagnent également de niveaux d'IOPS plus élevés que le serveur de fichiers peut transmettre aux instances SQL Server qui y accèdent. La capacité de stockage détermine non seulement la quantité de données que vous pouvez stocker, mais également le nombre d' I/O opérations par seconde (IOPS) que vous pouvez effectuer sur le stockage : chaque Go de stockage fournit 3 IOPS. Vous pouvez configurer chaque système de fichiers pour une taille maximale de 64 TiB (contre 16 TiB pour Amazon EBS). Vous pouvez également utiliser les FSx systèmes Amazon comme témoin de partage de fichiers pour les déploiements de Windows Server Failover Cluster (WSFC).

Options et considérations relatives à l'HA/DR

Bien qu'il soit extrêmement rare qu'une zone de AWS disponibilité ou une région soit complètement hors ligne, nous recommandons une approche à plusieurs volets pour la sauvegarde et la restauration en cas de sinistre, afin de garantir la redondance et de minimiser les pertes de données. Les processus de sauvegarde et de restauration doivent inclure le niveau de granularité approprié pour atteindre les objectifs de temps de restauration (RTO) et de point de restauration (RPO) pour la charge de travail et pour prendre en charge les processus métier, et dépendent souvent de l'application. Dans le cas des bases de données, prend AWS également en charge toutes les recommandations Microsoft relatives à l'installation et à la configuration de SQL Server pour la haute disponibilité et la reprise après sinistre (HA/DR). Different editions of SQL Server support various HA/DR options, and you should consider special cases such as very large databases (VLDBs) on a case-by-case basis. As with any DR configuration, testing is essential to ensure that each application meets its service-level agreements (SLAs) for HA/DR. For your test/developmentenvironnement). Pensez à utiliser l'édition SQL Server Developer, qui est gratuite mais comporte des limitations.

Pour un cas d'utilisation nécessitant un RPO de 15 minutes et un RTO de 4 heures, vous pouvez envisager une combinaison des options HA/DR suivantes :

- Options HA/DR natives de SQL Server avec mise en veille prolongée (au niveau de la base de données) — Pour des illustrations de certaines de ces architectures, consultez la Schémas d' EC2 architecture de SQL Server sur Amazon section plus loin dans ce guide.
  - Deux nœuds, multi-AZ dans une seule région (mode de validation synchrone) ou dans plusieurs régions (mode de validation asynchrone, groupe de disponibilité de base)
  - Trois nœuds (ou plus), multi-AZ dans plusieurs régions (modes de validation synchrone et de validation asynchrone)
  - Deux nœuds, multi-AZ et expédition de journaux dans plusieurs régions (avec des sauvegardes de journaux toutes les 5 minutes)
- Sauvegardes natives de SQL Server vers Amazon S3 (au niveau de la base de données, DR uniquement) — Sauvegardes complètes (une fois par jour)
  - Sauvegardes différentielles (toutes les 2 à 4 heures).
  - Enregistrez les sauvegardes (toutes les 5 à 10 minutes).

- Les sauvegardes doivent être effectuées et copiées sur Amazon Simple Storage Service (Amazon S3) à l'aide de scripts personnalisés ou d'une option telle qu'une passerelle de fichiers pour une sauvegarde et un transfert efficaces.
- Si vous possédez des centaines de bases de données, vous pouvez continuer à utiliser vos outils de sauvegarde existants (tels que Commvault ou Litespeed) pour gérer efficacement les sauvegardes et les stocker directement dans Amazon S3.
- Utilisez Amazon S3 Cross-Region Replication (CRR) avec S3 Replication Time Control (RTC) pour contrôler et surveiller la réplication d'objets dans le cadre d'un SLA de 15 minutes.
- Pour des raisons de conformité et de réduction des coûts, vous pouvez également utiliser la gestion du cycle de vie S3 pour déplacer et stocker les anciennes sauvegardes en vue d'un stockage à long terme.
- Si vous utilisez des sauvegardes natives de SQL Server et que vous les déplacez régulièrement vers Amazon S3, en cas de sinistre, les sauvegardes seront facilement disponibles dans la région cible. Ainsi, il n'est plus nécessaire de transférer des sauvegardes ou de restaurer des instantanés.
- Nous vous recommandons d'utiliser SQL Native Backup Compression pour réduire la taille des fichiers.

- AWS instantanés (niveau d'instance et de volume, DR uniquement)
  - Amazon Elastic Compute Cloud (Amazon EC2) Sauvegardes Amazon Machine Image (AMI) pour reconstruire des bases de données à partir de zéro
  - Instantanés de volumes Amazon Elastic Block Store (Amazon EBS) pour associer des volumes EBS à Amazon EC2

Gestion des ressources HA/DR dans AWS Backup

AWS Backupest un service entièrement géré qui permet de créer des plans et des plannings de sauvegarde, et d'affecter les AWS ressources impliquées dans la configuration HA/DR, telles que les volumes Amazon EBS pour créer des instantanés et Amazon, à ces plans de sauvegarde. EC2 AMIs Vous pouvez également l'utiliser AWS Backup pour planifier des copies multirégionales de ces instantanés EBS. Pour une utilisation optimale, AWS Backup il faut un mécanisme de balisage efficace pour que les ressources soient en place. AWS Backup prend également en charge les sauvegardes cohérentes avec les applications via le service Windows Volume Shadow Copy (VSS), que vous pouvez utiliser pour SQL Server. Pour une protection au niveau du stockage, nous vous recommandons d'utiliser des instantanés EBS. Les instantanés EBS initiaux sont complets et les

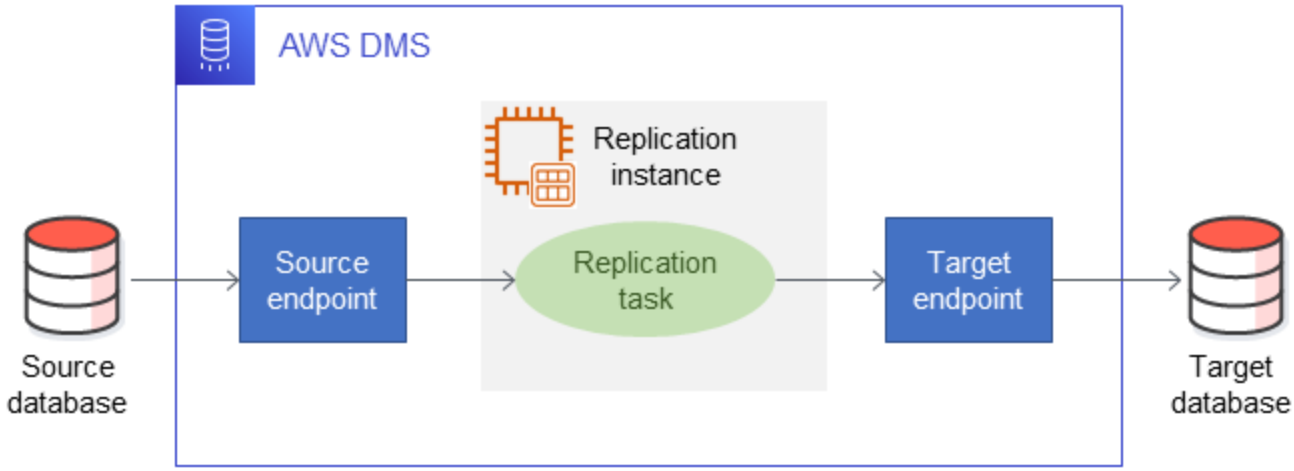
instantanés suivants sont incrémentiels. Bien que les instantanés EBS offrent une protection au niveau du stockage, ils ne remplacent pas les sauvegardes natives basées sur des fichiers SQL Server qui offrent une restauration. point-in-time

Utilisation AWS DMS pour HA/DR

Si vous recherchez une alternative aux options Always On de SQL Server pour la réplication ou si vous disposez de bases de données source et cible hétérogènes, que ce soit dans une configuration hybride ou dans AWS, vous pouvez utiliser AWS Database Migration Service (AWS DMS) de la manière suivante.

Si vous l'utilisez AWS DMS avec SQL Server dans un contexte autogéré (hébergé sur Amazon EC2 ou sur site), il prend en charge la réplication ponctuelle et continue selon deux modes : en utilisant MS-REPLICATION (pour capturer les modifications apportées aux tables dotées de clés primaires) et MS-CDC (pour capturer les modifications apportées aux tables dépourvues de clés primaires). Toutefois, si vous utilisez Amazon Relational Database Service (Amazon RDS) comme source AWS DMS pour, seul le MS-CDC est pris en charge. AWS DMS propose une gamme de points de terminaison source et cible, prend en charge des moteurs de base de données hétérogènes et offre un contrôle précis du processus de réplication. Vous pouvez également utiliser le AWS Schema Conversion Tool (AWS SCT) avec AWS DMS pour les migrations de bases de données hétérogènes. AWS SCT automatise les modifications au niveau du schéma et produit également des rapports pour la préparation et la planification de la migration.

Vous ajoutez des bases de données source et cible en tant que points de terminaison AWS DMS, comme illustré dans le schéma suivant. Ce service met en œuvre un processus de réplication logique à l'aide de MS-REPLICATION ou de MS-CDC. Si vous avez une configuration hybride, vous pouvez la configurer AWS DMS pour une réplication continue entre une configuration sur site et AWS. Pendant le passage, la tâche de AWS DMS migration peut être arrêtée et l'application pourra se connecter à la base de données déjà synchronisée avec la base de données locale sans plus tarder. L'utilisation de AWS DMS for SQL Server en tant que source comporte quelques limites, qui sont décrites dans la AWS DMS documentation.

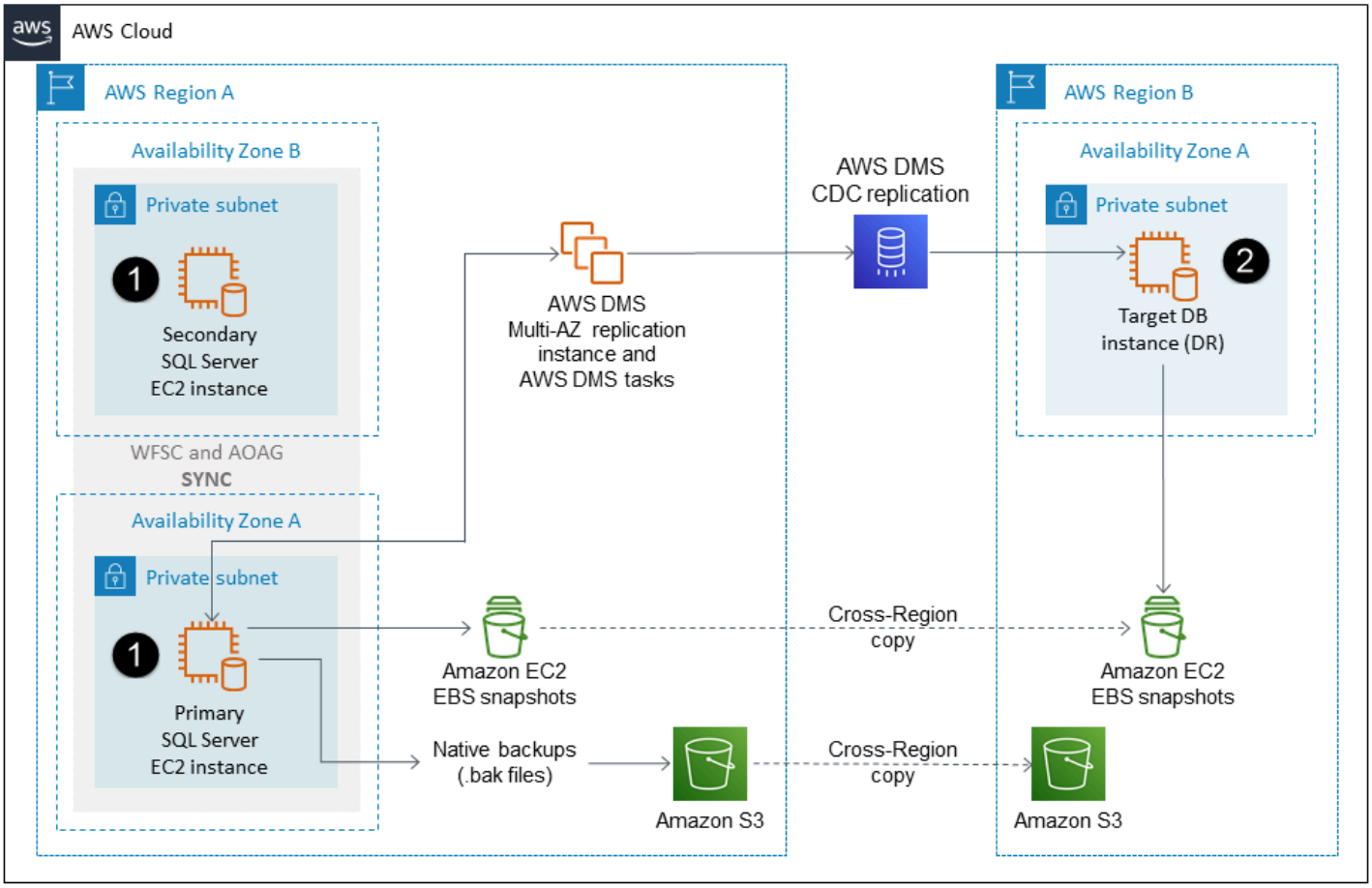


Envisagez d'utiliser des méthodes HA/DR natives à la AWS DMS place des méthodes HA/DR natives dans les scénarios suivants :

- Lorsque vous souhaitez économiser sur les coûts de licence. Par exemple, si vous utilisez une version avancée telle que SQL Server Enterprise uniquement pour ses options Always On, vous pouvez envisager de la configurer à la AWS DMS place, car elle peut fournir une option de réplication logique sans le coût d'une licence d'édition Enterprise.
- Lorsque vous avez des sources et des cibles hétérogènes. Les versions de SQL Server sur les nœuds principaux et de reprise après sinistre n'ont pas besoin de correspondre (dans AWS DMS certaines limites), ce qui offre une flexibilité significative.
- Pour éviter la surcharge liée à Windows, au clustering SQL Server et à la configuration et à la gestion des groupes de disponibilité distribués. AWS DMS permet une configuration simple et une gestion aisée des tâches de réplication.
- Pour les cas d'utilisation professionnels tels que le transfert en temps quasi réel (en fonction de l'instance de réplication, de la configuration réseau et du volume de données), le masquage des données, le filtrage sélectif, le mappage des schémas/tables (homogènes et hétérogènes), les évaluations préalables à la migration et le support JSON.
- Pour dupliquer, arrêter et démarrer facilement des tâches selon les besoins en fonction des numéros de séquence du journal (LSNs), des horodatages et d'options similaires.

Le schéma suivant montre une autre approche permettant de AWS DMS fournir un support de réplication. Dans cette configuration, la source est un cluster de groupes de disponibilité SQL Server Always On qui AWS DMS utilise l'option de capture des données de modification (CDC)

pour répliquer en continu les données vers une cible située dans une autre AWS région. Pour des performances optimales, il est essentiel de s'assurer que l'instance de réplication est correctement dimensionnée et qu'elle reste dans la région source.



Il n'est pas nécessaire que les moteurs source et cible correspondent. Dans le diagramme, les nœuds principal et secondaire marqués comme (1) peuvent être un cluster SQL Server dans une configuration mono-AZ ou multi-AZ. La source peut également être un nœud SQL Server unique qui prend en charge MS-CDC ou MS-REPLICATION.

L'instance de base de données cible, marquée comme (2) dans le diagramme, peut être n'importe quelle version de SQL Server sur Amazon RDS EC2, Amazon ou toute autre cible hétérogène. Il n'est pas nécessaire qu'elle corresponde aux instances principale et secondaire ou qu'elle prenne en charge les groupes de disponibilité Always On. Par exemple, la source peut être un cluster de groupes de disponibilité SQL Server Always On, et la cible peut être l'édition compatible avec Amazon Aurora PostgreSQL.

Utilisation AWS Application Migration Service pour la DR

Nous vous recommandons d'utiliser le AWS Application Migration Service pour lift-and-shift les migrations vers AWS. Le service de migration d'applications réplique en permanence vos machines (y compris le système d'exploitation, la configuration de l'état du système, les bases de données, les applications et les fichiers) dans une zone de transit peu coûteuse de votre AWS compte cible et de votre région préférée. En cas de sinistre, vous pouvez utiliser le service de migration d'applications pour lancer automatiquement des milliers de machines entièrement provisionnées en quelques minutes.

Considérations supplémentaires

La liste suivante identifie les points d'étranglement possibles que vous devez prendre en compte lors de la conception d'une stratégie HA/DR.

- Bande passante, latence, complexité du réseau et connectivité dans une configuration de nœud multirégional.
- Taille des EC2 instantanés Amazon EBS ou Amazon, et temps nécessaire pour les copier à l'aide de. AWS Backup
  - Les EC2 instantanés Amazon EBS et Amazon sont stockés dans Amazon S3 à l'aide de. AWS Backup
  - Un instantané EBS n'est pas répliqué dans la région cible dans Amazon S3 tant que le cliché actuel n'est pas terminé. La durée de réplication dépend également de la taille du volume.
  - Lorsque l'instantané est terminé, la durée de copie des instantanés peut être de 15 minutes seulement pour 99,99 % des objets. Cependant, des tests approfondis sont nécessaires pour des cas d'utilisation spécifiques et des volumes importants critiques.
- Temps nécessaire pour restaurer les volumes EBS dans la zone de disponibilité et la région cibles.
- Temps nécessaire pour restaurer les EC2 images Amazon dans la zone de disponibilité et la région cibles.
- Si vous créez à partir de zéro, il faut du temps pour provisionner l'infrastructure pour l' EC2image Amazon ou pour restaurer les instantanés EBS dans la zone de disponibilité et la région cibles.
- En cas de restauration à partir de zéro, temps nécessaire pour restaurer les sauvegardes complètes, différentielles et journaux natives de SQL Server dans la zone de disponibilité et la région cibles.

- Dépendances applicatives et externes qui doivent être disponibles dans toutes les régions.
- Limitations relatives à la taille des fichiers pour les volumes et pour le téléchargement vers Amazon S3.

Scénarios de reprise après sinistre

Cette section fournit des exemples de défaillance d'une seule zone de disponibilité ou d'une seule AWS région, et décrit les options de reprise après sinistre (DR). Les exemples supposent un objectif de point de récupération (RPO) de 15 minutes et un objectif de temps de restauration (RTO) de 4 heures.

Défaillance de zone de disponibilité

Vous pouvez utiliser l'une des options suivantes pour effectuer une restauration après une défaillance d'une seule zone de disponibilité dans les paramètres définis (RPO de 15 minutes, RTO de 4 heures).

- Provisionnez la restauration de l'application à l'aide de la dernière sauvegarde d'image Amazon Elastic Compute Cloud (Amazon EC2) et connectez-vous à l'instance de base de données Warm Standby existante via le déploiement d'un groupe de disponibilité Always On ou l'expédition de journaux.
  - Une configuration de groupe de disponibilité SQL Server Always On pour la reprise après sinistre avec deux nœuds ou plus permet un basculement automatique vers le nœud secondaire via le mode de validation synchrone ou asynchrone, de sorte que la base de données est disponible immédiatement. Pour une configuration HA, les deux nœuds sont disponibles pour les opérations de lecture. Cette option répond facilement aux exigences RTO et RPO. Dans l'édition standard de SQL Server, l'utilisation de groupes de disponibilité de base est également une option, mais elle est limitée à deux nœuds, car un groupe de disponibilité ne peut inclure qu'une seule base de données. Toutefois, vous pouvez configurer plusieurs groupes de disponibilité au sein d'une même région ou entre plusieurs régions. Cette configuration permet de réaliser des économies, car il n'y a aucun coût supplémentaire pour le nœud secondaire, qui n'est pas accessible pour les opérations de lecture. L'édition SQL Server Enterprise fournit des fonctionnalités complètes et un basculement sur incident pour toutes les bases de données au sein d'un même groupe de disponibilité. Pour des exemples de cette option, consultez les diagrammes d'architecture suivants :
    - Architecture HA/DR à deux nœuds avec cluster de groupes de disponibilité Always On (région unique, multi-AZ)
    - Architecture HA/DR à trois nœuds (région unique, multi-AZ)

    - Architecture HA/DR à quatre nœuds avec cluster de groupes de disponibilité distribués Always On (multirégion, multi-AZ)
    - Architecture HA/DR à trois nœuds avec groupe de disponibilité unique (multirégion)
  - L'expédition des journaux SQL Server en tant que solution de reprise après sinistre nécessite un basculement manuel vers un serveur de secours et dépend de la fréquence des sauvegardes des journaux. Il s'agit de l'une des options de reprise après sinistre les moins coûteuses. Il n'est pas nécessaire que les éditions de SQL Server du site de reprise après sinistre principal et du site DR expédié par journal correspondent. Cette option répond au RPO (utilisation de sauvegardes du journal des transactions toutes les 5 minutes et au RTO), mais nécessite une maintenance par le biais de scripts manuels personnalisés. Pour un exemple de cette option, consultez le schéma d'architecture suivant :
    - Architecture HA/DR à trois nœuds avec expédition de journaux (multirégion)
- Si vous disposez d'une application telle qu'une application SQL Server Reporting Services (SSRS) dotée d'un déploiement évolutif, l'équilibreur de charge peut rediriger tout le trafic vers le nœud secondaire.
- Vous pouvez utiliser Amazon EC2 base AMIs pour le serveur d'applications et de base de données afin de provisionner l'infrastructure. Les bases de données peuvent être restaurées dans une nouvelle zone de disponibilité, en fonction de leur taille et de leur fréquence de sauvegarde, à partir des sauvegardes natives les plus récentes (sauvegarde complète, sauvegarde différentielle ou sauvegarde du journal des transactions toutes les 5 minutes) ou à l'aide de snapshots EBS. Cette option répond aux exigences RPO et RTO mais nécessite un script personnalisé. Vous devez également tenir compte du temps nécessaire pour provisionner l'infrastructure, et répondre aux exigences de RPO et de RTO peut s'avérer difficile.
- EC2 Les images Amazon (y compris les volumes EBS) pour les applications et le serveur de base de données peuvent être restaurées dans une nouvelle zone de disponibilité. Le RPO peut s'avérer difficile, en fonction de la sauvegarde la plus récente, mais cette option peut être associée aux journaux de transactions les plus récents pour répondre aux exigences. Cette option prend en charge les instantanés Windows Volume Shadow Copy Service (VSS).

Défaillance régionale

Vous pouvez utiliser l'une des options suivantes pour effectuer une restauration après une défaillance d'une seule AWS région dans les limites des paramètres définis (RPO de 15 minutes, RTO de 4 heures).

- Vous pouvez utiliser Amazon Machine Images (AMIs) de EC2 base Amazon pour le serveur d'applications et de base de données afin de provisionner l'infrastructure. Les bases de données peuvent être restaurées dans une nouvelle région, en fonction de leur taille et de leur fréquence de sauvegarde, à partir des sauvegardes natives les plus récentes (sauvegarde complète, sauvegarde différentielle ou sauvegarde du journal des transactions toutes les 5 minutes). Cette option répond aux exigences RPO et RTO mais nécessite un script personnalisé.
  - L'expédition des journaux SQL Server en tant que solution de reprise après sinistre nécessite un basculement manuel vers un serveur de secours et dépend de la fréquence des sauvegardes des journaux. Il s'agit de l'une des options de reprise après sinistre les moins coûteuses. Il n'est pas nécessaire que les éditions de SQL Server du site de reprise après sinistre principal et du site DR expédié par journal correspondent. Cette option répond au RPO (en utilisant des sauvegardes du journal des transactions toutes les 5 minutes) et au RTO, mais nécessite une maintenance par le biais de scripts manuels personnalisés. Les bases de données volumineuses nécessitent de longs délais de restauration.
- Vous pouvez utiliser une EC2 AMI Amazon pour l'application et le serveur de base de données et le restaurer sur une cible dans une nouvelle région. Le RPO dépend de la taille et de la fréquence des sauvegardes.
  - Les images d'application les plus récentes peuvent être restaurées à l'aide d'une AMI. Vous pouvez utiliser des sauvegardes différentielles ou du journal des transactions natives récentes toutes les 5 minutes pour mettre la base de données à jour afin de respecter le RPO.
  - Le RTO dépend de la taille et du temps nécessaires pour transférer et restaurer les instantanés vers la nouvelle région, si la source n'est pas déjà synchronisée avec la cible.
- La solution présentant le moins de temps d'arrêt consiste à restaurer l'image de sauvegarde de l'application et à disposer d'un nœud SQL Server de secours dans une région distante en utilisant une configuration de groupe de disponibilité à deux, trois ou quatre nœuds (de base, classique ou distribué) et à se connecter au serveur de base de données de secours après un basculement. La réplique en mode de validation synchrone répond aux exigences du RPO, tandis que la réplique en mode de validation asynchrone peut être retardée en fonction du volume de transactions. Vous pouvez utiliser une configuration de groupe de disponibilité distribué pour étendre les nœuds de base de données dans une nouvelle région, si nécessaire. Cette configuration réduit également la complexité car elle utilise deux groupes de disponibilité indépendants au lieu d'un seul groupe de disponibilité réparti entre les régions en mode de validation synchrone ou asynchrone, et répond facilement aux exigences RTO et RPO. Il est également possible d'utiliser les groupes de disponibilité de base de SQL Server dans l'édition Standard. Cependant, il présente des limites car il ne prend en charge que deux nœuds et qu'une seule base de données peut figurer

dans un seul groupe de disponibilité, bien que plusieurs groupes de disponibilité soient pris en charge. Vous pouvez configurer l'édition standard de SQL Server dans une ou plusieurs régions. Cette édition permet de réaliser des économies car elle ne facture pas le nœud secondaire, qui n'est pas accessible pour les opérations de lecture. L'édition SQL Server Enterprise fournit des fonctionnalités complètes et prend en charge le basculement de toutes les bases de données en tant que basculement d'un seul groupe de disponibilité.

Cas d’utilisation courants

À titre d'exercice de dimensionnement, 80 % des applications SQL Server exécutées sur Amazon et EC2 dont la charge de travail est normale pour le traitement des transactions en ligne (OLTP) peuvent être regroupées dans l'une des trois catégories suivantes, en fonction de leur importance critique :

- SQL Server HA/DR avec sauvegardes SQL Server, à l'aide de deux répliques en mode de validation synchrone et d'une réplique en mode de validation asynchrone
- AWS Backup HA/DR avec sauvegardes SQL Server, à l'aide d'une EC2 AMI Amazon pour l'application et la base de données, et du stockage Amazon EBS
- AWS Backup HA/DR avec sauvegardes SQL Server, à l'aide d'une EC2 AMI de base Amazon pour le serveur de base de données, d' EC2 une image Amazon pour l'application et de snapshots Amazon EBS

Le tableau suivant fournit des informations détaillées sur chaque catégorie.

	SQL Server HA/DR avec sauvegardes SQL Server	AWS Backup HA/DR avec AMIs stockage EBS et sauvegardes SQL Server	AWS Backup HA/DR avec AMIs, instantanés EBS et sauvegardes SQL Server
Processus de restauration en cas de sinistre	• Restaurez l'AMI EC2 de base Amazon pour l'application depuis AWS Backup	• Restaurez EC2 les images Amazon à partir des sauvegardes de l'application et de la base de données	• Restaurer EC2 l'image Amazon à partir de la sauvegarde de l'application

	SQL Server HA/DR avec sauvegardes SQL Server	AWS Backup HA/DR avec AMIs stockage EBS et sauvegardes SQL Server	AWS Backup HA/DR avec AMIs, instantanés EBS et sauvegardes SQL Server
	• Basculer vers l'instance de secours de la région (en cas de défaillance de la zone de disponibilité) ou vers l'instance interrégionale (en cas de défaillance de la région) • Répond aux exigences RPO et RTO	• Fournit un support à la fois régional et interrégional • Appliquez les sauvegardes différentielles et du journal des transactions les plus récentes de SQL Server (toutes les 15 minutes) pour répondre aux exigences RPO et RTO de la base de données	• Restaurer l'AMI EC2 de base Amazon pour le serveur de base de données • Restaurez les instantanés EBS (le cas échéant) • Le cluster doit être reconstruit • Fournit un support à la fois régional et interrégional • Appliquez les sauvegardes différentielles et du journal des transactions les plus récentes à la base de données pour répondre aux exigences du RPO, mais le RTO risque de ne pas être atteint

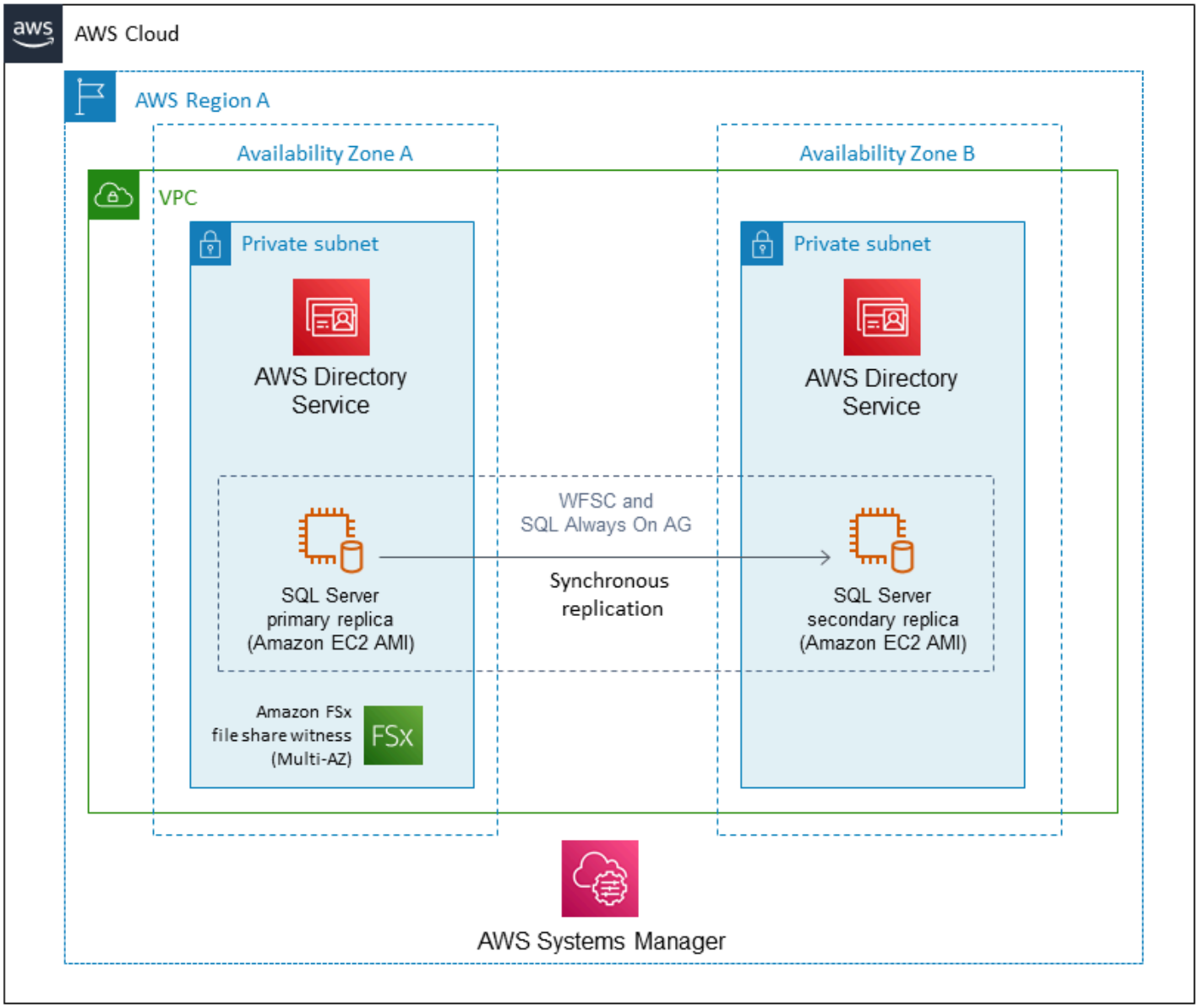
	SQL Server HA/DR avec sauvegardes SQL Server	AWS Backup HA/DR avec AMIs stockage EBS et sauvegardes SQL Server	AWS Backup HA/DR avec AMIs, instantanés EBS et sauvegardes SQL Server
Ressources principales	• Trois licences SQL Server Enterprise Edition (la licence passive pour les nœuds HA et DR est gratuite si vous avez un contrat de licence Software Assurance existant avec Microsoft ; voir l'annonce) • Espace EC2 de sauvegarde Amazon sur Amazon Simple Storage Service (Amazon S3) • Transfert de données entre régions	• Une licence SQL Server (n'importe quelle édition). • Espace EC2 de sauvegarde Amazon sur Amazon S3 • Sauvegardes SQL Server (fichiers différentiels et journaux) sur Amazon S3 • Transfert de données entre régions	• Une licence SQL Server (n'importe quelle édition). • Espace EC2 de sauvegarde Amazon sur Amazon S3 • Sauvegardes SQL Server (fichiers différentiels et journaux) sur Amazon S3 • Transfert de données entre régions
HA/DR	Offres HA et DR	Offres DR uniquement	Offres DR uniquement
RPO	Le basculement est géré par le groupe de disponibilité de SQL Server (DR est manuel)	Script manuel ou personnalisé	Script manuel ou personnalisé
RTO	Quelques secondes à quelques minutes	De quelques minutes à quelques heures	Plusieurs heures

	SQL Server HA/DR avec sauvegardes SQL Server	AWS Backup HA/DR avec AMIs stockage EBS et sauvegardes SQL Server	AWS Backup HA/DR avec AMIs, instantanés EBS et sauvegardes SQL Server
Risque de disparition SLAs	Faible	Medium	Élevé
Facilité de gestion	Simplicité	Moyen	Moyen
Dimensionnement	Simplicité	Moyen	Moyen
Limites de taille de fichier pour les téléchargements vers Amazon S3 ou les transferts entre régions	N/A — Géré en mode de validation synchrone ou en mode de validation asynchrone pour une mise en veille prolongée	Oui	Oui
Perte de données	Près de zéro (dépend de la charge de travail et de l'infrastructure mises en service)	Dépend de la fréquence des images de EC2 sauvegarde Amazon et des sauvegardes SQL Server	Cela dépend de la fréquence des images de EC2 sauvegarde Amazon ou des instantanés EBS et des sauvegardes SQL Server
Coût	Moyen	Faible à moyen	Faible à moyen

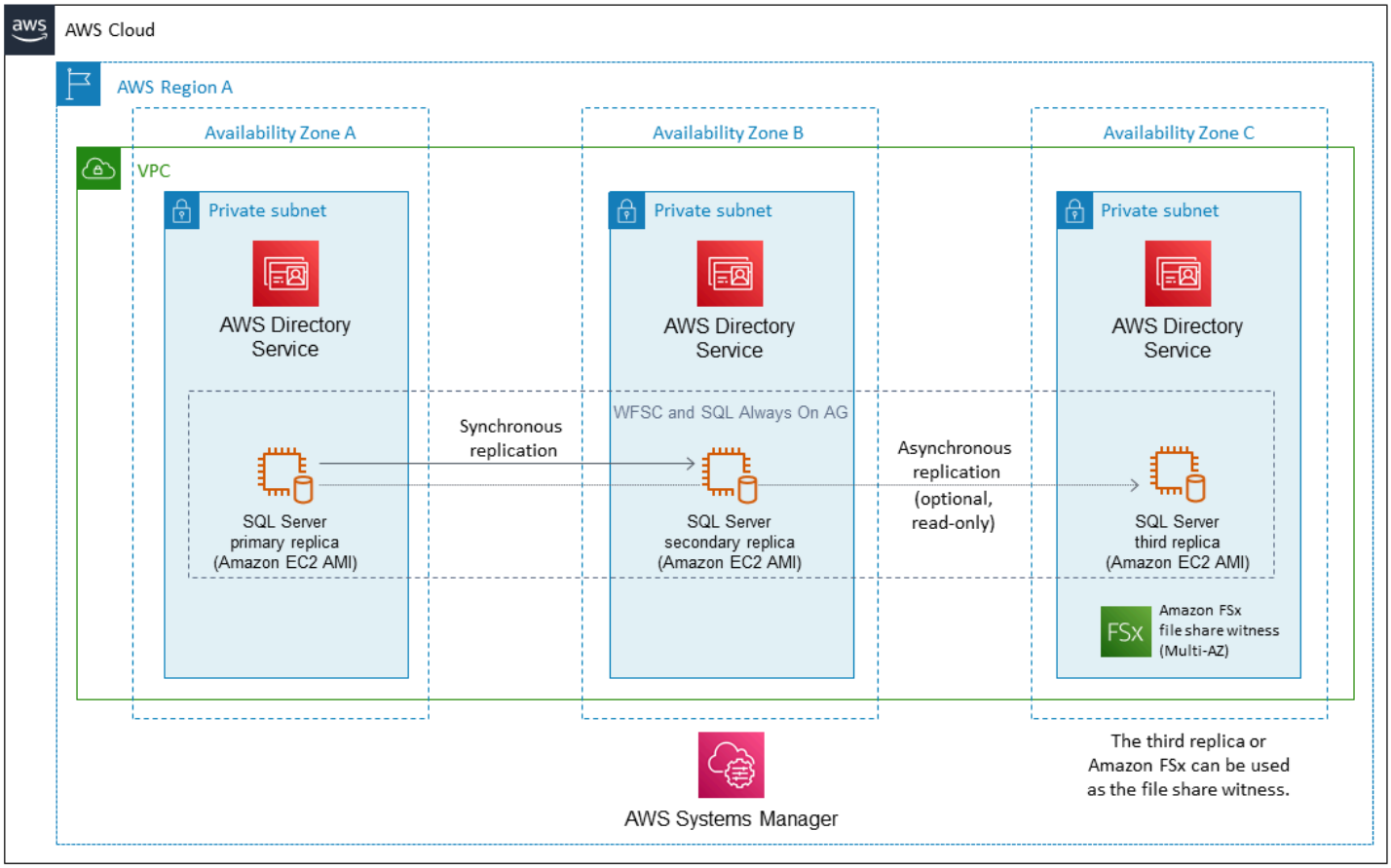
Schémas d' EC2 architecture de SQL Server sur Amazon

Cette section fournit des diagrammes d'architecture qui illustrent les stratégies HA/DR décrites dans les sections précédentes.

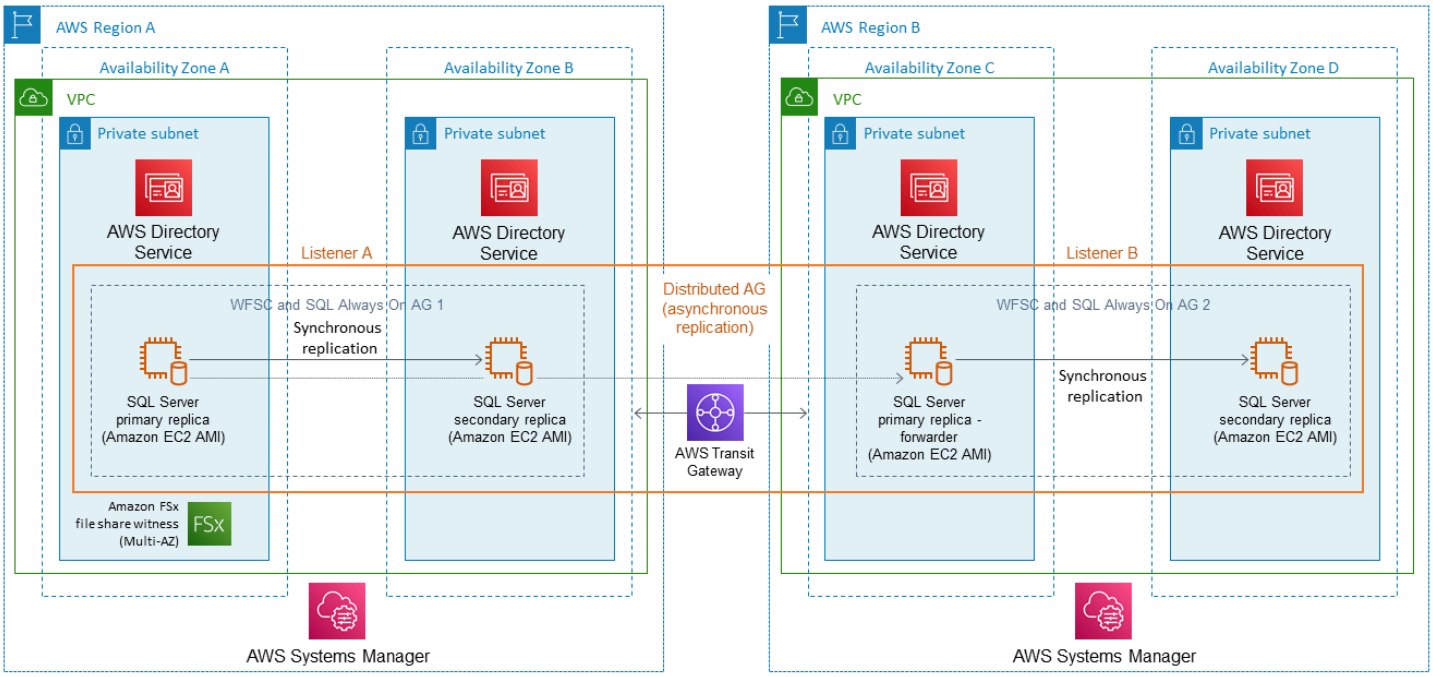
Architecture HA/DR à deux nœuds avec cluster de groupes de disponibilité Always On (région unique, multi-AZ)



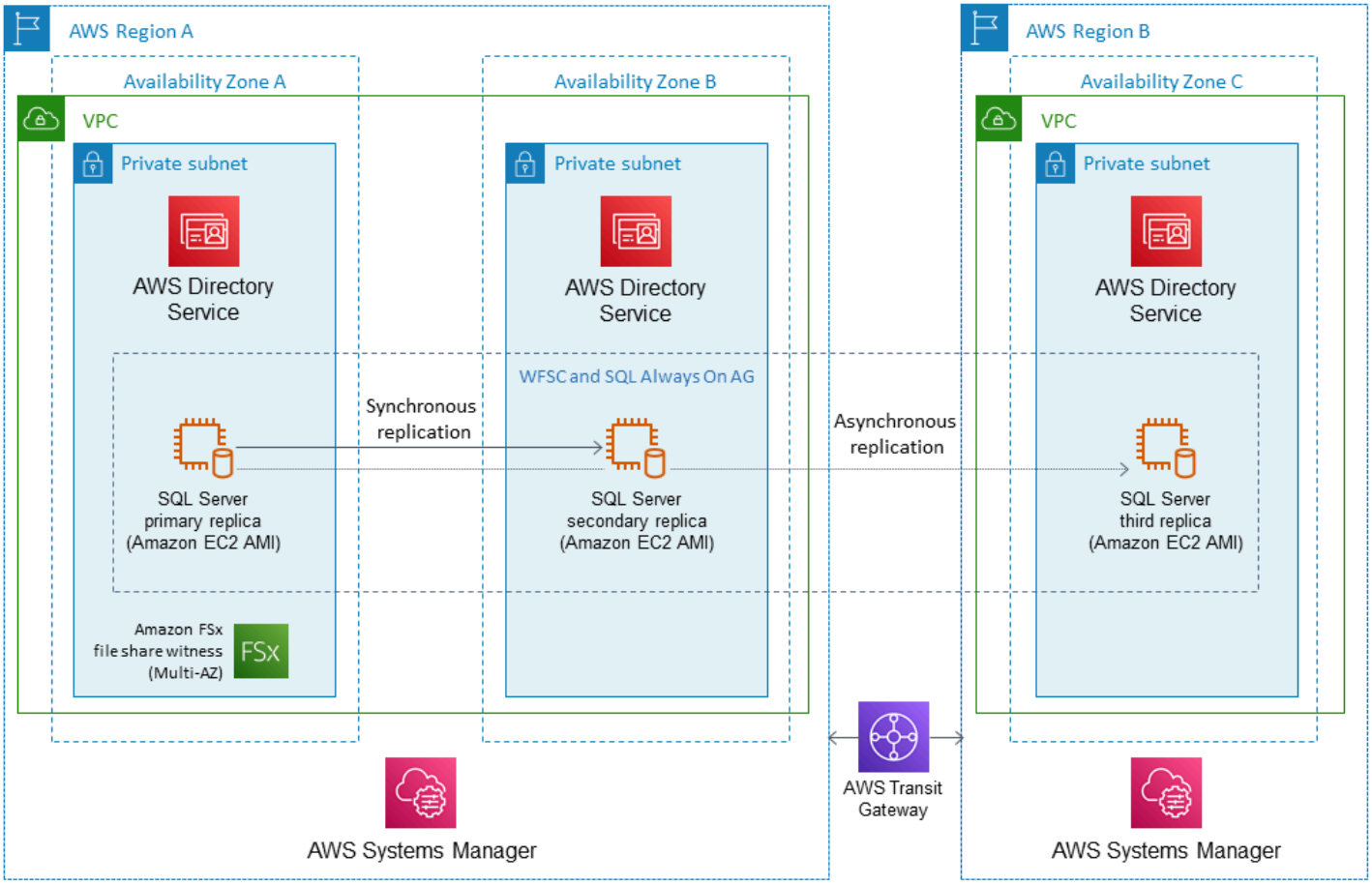
Architecture HA/DR à trois nœuds (région unique, multi-AZ)



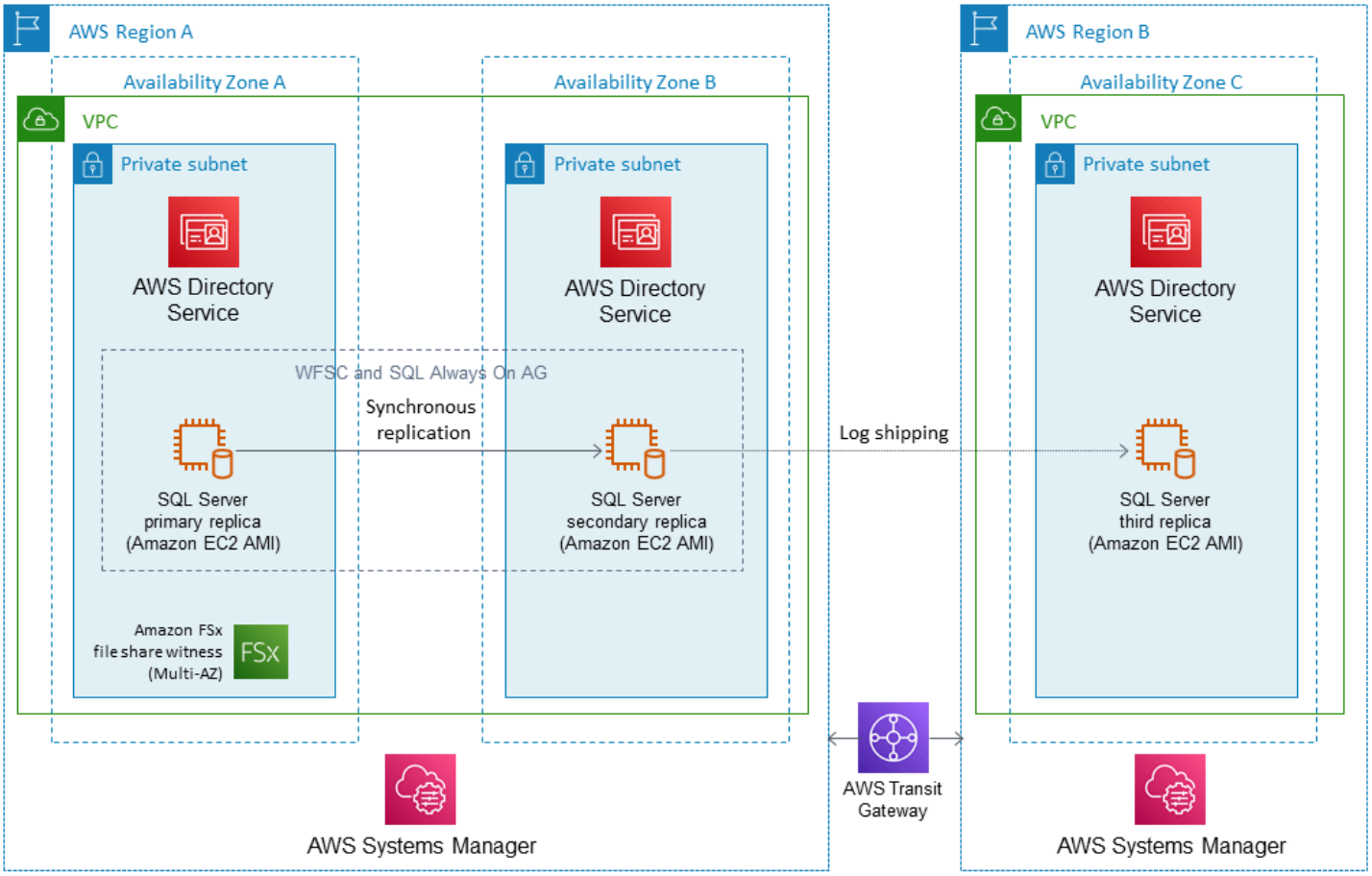
Architecture HA/DR à quatre nœuds avec cluster de groupes de disponibilité distribués Always On (multirégion, multi-AZ)



Architecture HA/DR à trois nœuds avec groupe de disponibilité unique (multirégion)



Architecture HA/DR à trois nœuds avec expédition de journaux (multirégion)

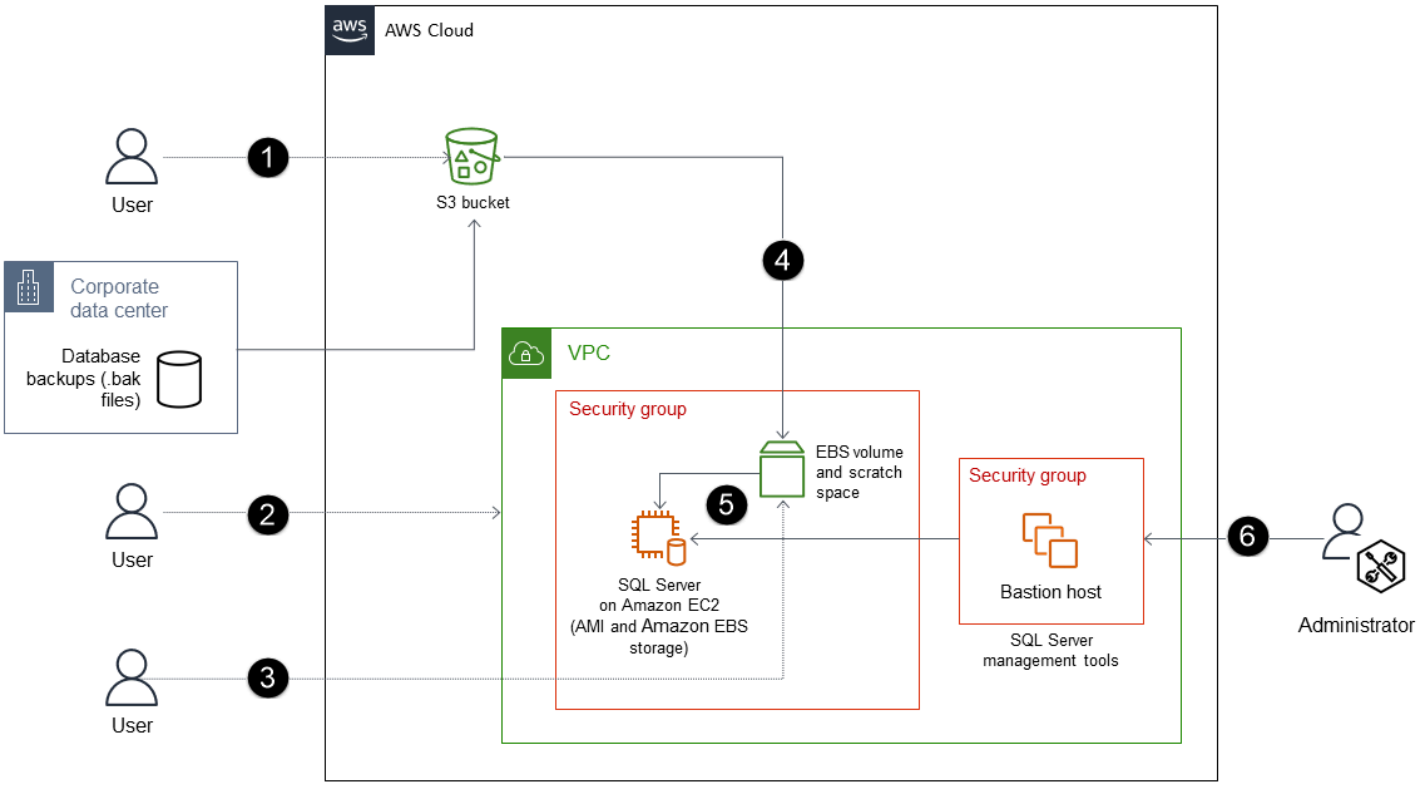


Options de restauration

Les sections suivantes proposent deux options de restauration de base de données pour SQL Server sur Amazon Elastic Compute Cloud (Amazon EC2), lorsque vos sauvegardes sont sur site.

Utilisation d'Amazon S3

Cette approche de restauration de base de données SQL Server utilise les commandes Amazon Simple Storage Service (Amazon S3) pour AWS Command Line Interface le AWS CLI() ou l'API Amazon S3 afin de télécharger les fichiers de sauvegarde directement dans un compartiment S3.



Le processus comprend les étapes suivantes :

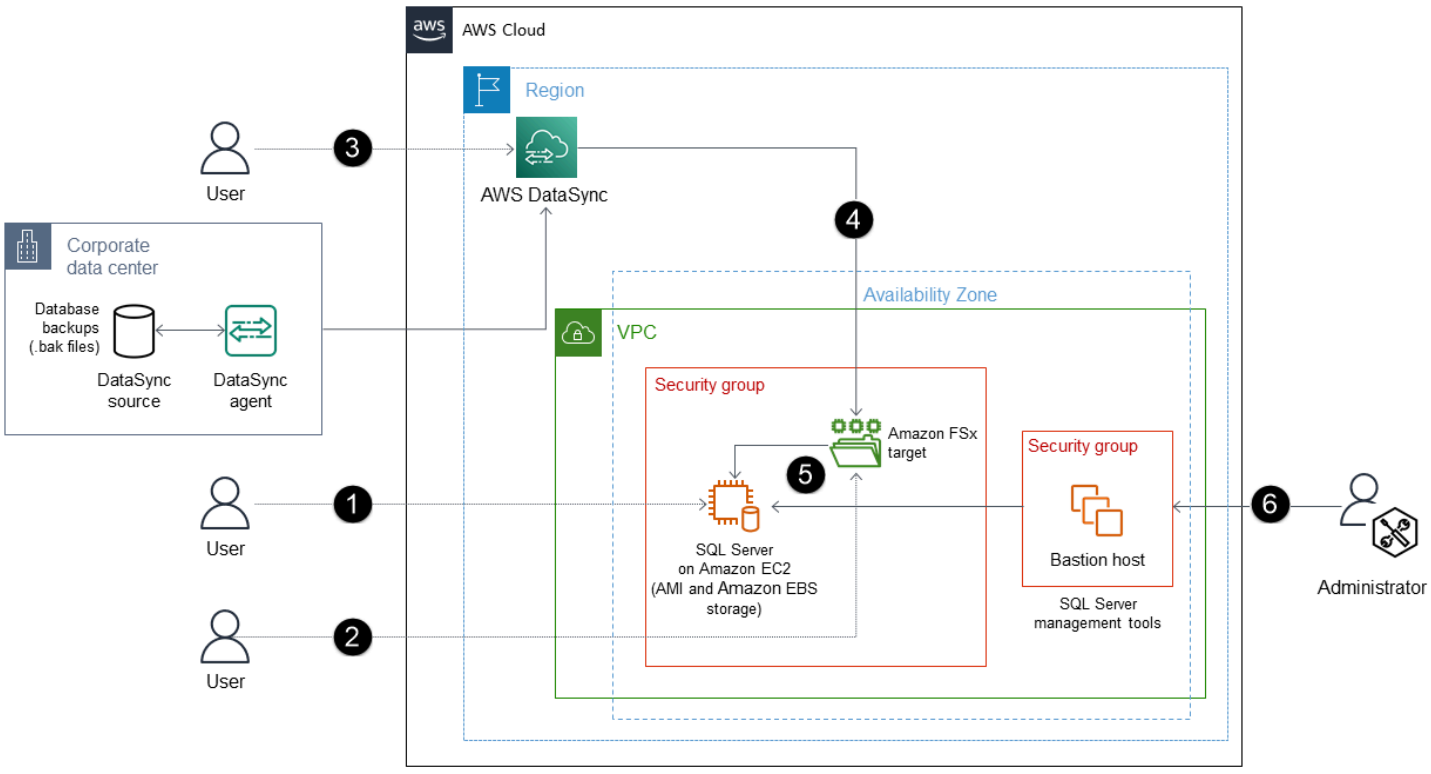
1. Créez un compartiment S3 (ou utilisez un compartiment existant) pour stocker les fichiers de sauvegarde et transférez les fichiers de sauvegarde (.bak) de votre base de données sur site vers le compartiment S3 à l'aide de la AWS CLI ou de l'API Amazon S3.
2. Déployez SQL Server sur une EC2 instance optimisée pour EBS, à l'aide d'une Amazon Machine Image (AMI) SQL Server. Cette AMI doit contenir des volumes EBS configurés avec une partition

du système d'exploitation, une partition DATA, une partition LOG, un stockage tempdb (NVMe) et un espace de travail.

3. (Facultatif) Attachez un volume EBS non root à l' EC2 instance.
4. Copiez les fichiers de sauvegarde sur le volume EBS non root.
5. Restaurez les fichiers de sauvegarde du volume EBS vers SQL Server sur l' EC2 instance.
6. Utilisez les outils de gestion de SQL Server pour gérer votre base de données.

Utilisation AWS DataSync d'Amazon FSx

Cette approche de restauration de base de données SQL Server permet de AWS DataSync transférer les fichiers de sauvegarde vers Amazon FSx pour Windows File Server.



Le processus comprend les étapes suivantes :

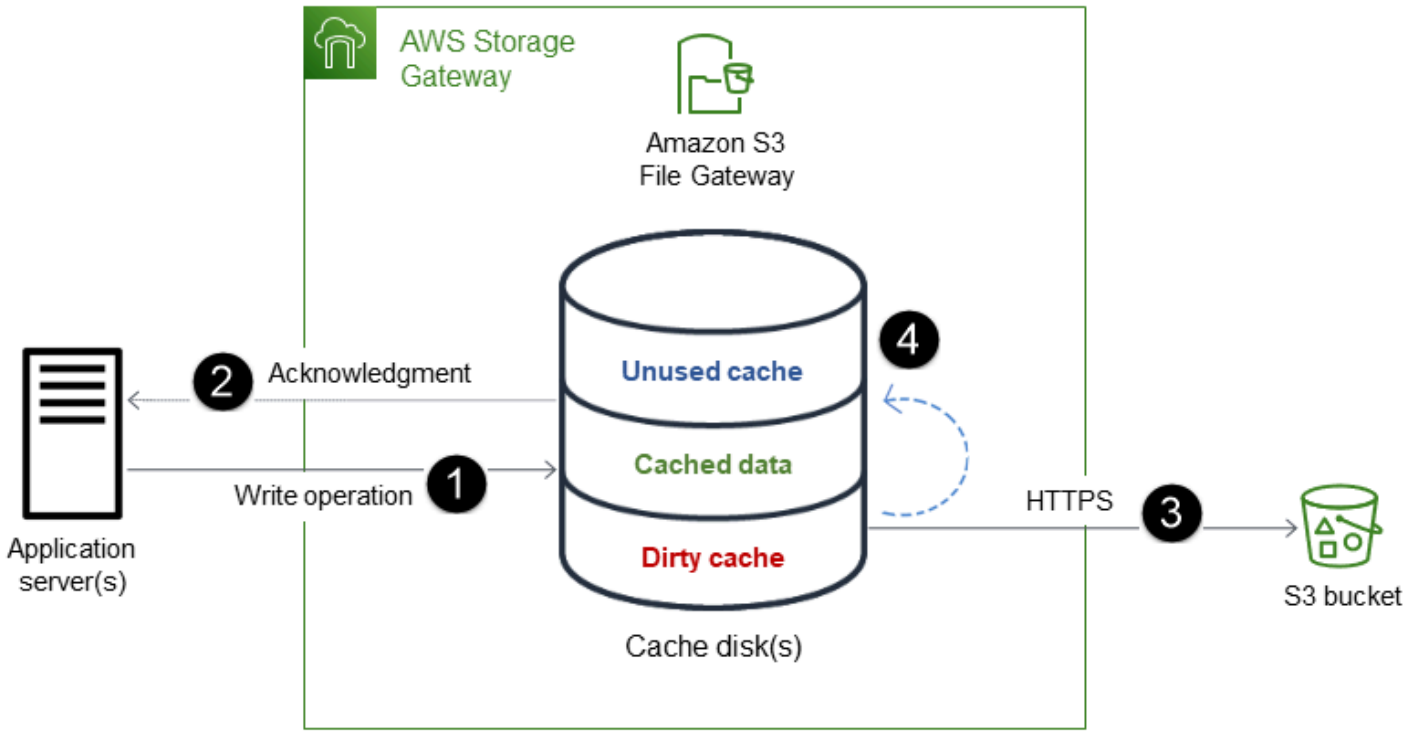
1. Déployez SQL Server sur une EC2 instance optimisée pour EBS avec pièce jointe NVMe, à l'aide d'une AMI contenant des volumes EBS configurés avec OS, DATA, LOG et tempdb. (Par exemple, vous pouvez utiliser la classe d'`r5d.large`instance optimisée pour la mémoire.)

2. FSx À utiliser pour Windows File Server pour créer un serveur de fichiers. Il peut être utilisé comme emplacement de stockage temporaire pour télécharger des fichiers de sauvegarde SQL Server (.bak) depuis votre environnement local.
3. Créez un DataSync point de terminaison et un agent pour le serveur de FSx fichiers Amazon.
4. DataSync automatise la synchronisation des données entre votre stockage sur site et le serveur de FSx fichiers Amazon sans avoir besoin d'Amazon S3.
5. Restaurez les fichiers de sauvegarde du serveur de FSx fichiers Amazon vers SQL Server sur l' EC2instance.
6. Utilisez les outils de gestion de SQL Server pour gérer votre base de données.

> **Note**
>
> Amazon EC2 propose Microsoft SQL Server sur Microsoft Windows Server AMIs pour plusieurs éditions de SQL Server.

Utilisation d'Amazon S3 File Gateway

Vous pouvez utiliser Amazon S3 File Gateway pour stocker des sauvegardes SQL Server natives sur Amazon S3, comme illustré dans le schéma suivant. Il existe également des outils tels que Commvault LiteSpeedqui vous aident à gérer les sauvegardes au niveau des fichiers à grande échelle et à les stocker directement dans Amazon S3. Vous pouvez également utiliser un outil tel que SIOS DataKeeper pour la sauvegarde/restauration et la configuration de la reprise après sinistre.



Le processus comprend les étapes suivantes :

1. Les données sont écrites sur le disque de cache local de la passerelle de fichiers.
2. Une fois que les données sont conservées en toute sécurité dans le cache local, la passerelle de fichiers confirme la fin de l'opération d'écriture à l'application cliente.
3. La passerelle de fichiers transfère les données vers le compartiment S3 de manière asynchrone. Il optimise le transfert de données et utilise le protocole HTTPS pour chiffrer les données en transit.
4. Une fois les données téléchargées dans le compartiment S3, elles restent dans le cache local de la passerelle de fichiers jusqu'à ce qu'elles soient expulsées.

Prochaines étapes et ressources

Ce guide décrit les meilleures pratiques en matière de restauration rapide après sinistre des bases de données SQL Server. Les recommandations incluent l'utilisation d'images pour restaurer l'instance de l'application et l'utilisation de méthodes SQL natives pour restaurer la base de données ou, de préférence, le basculement de la base de données. Contrairement aux restaurations de bases de données volumineuses qui peuvent prendre des heures, l'utilisation des sauvegardes Amazon Machine Image (AMI EC2) d'Amazon Elastic Compute Cloud (Amazon) en combinaison avec les journaux de transactions les plus récents vous aide à atteindre vos objectifs de point de reprise (RPO) et de temps de restauration (RTO) tout en réduisant vos coûts globaux. L'approche optimale dépend de la taille de votre base de données, du nombre et de la nature des sauvegardes, ainsi que de la fréquence des sauvegardes du journal des transactions pour lesquelles une stratégie de reprise après sinistre doit être conçue. Consultez les liens suivants pour plus d'informations, les meilleures pratiques, les guides de démarrage rapide et les conseils prescriptifs sur la migration et l'hébergement de SQL Server sur Amazon. EC2

Documentation

- Bonnes pratiques et recommandations pour le clustering SQL Server sur Amazon EC2 (EC2documentation Amazon)
- Boutique d' EC2instances Amazon (EC2 documentation Amazon)
- Réplication d'objets (documentation Amazon S3)
- Restauration rapide des instantanés Amazon EBS (EC2 documentation Amazon)
- SQL Server avec réplication Always On sur le AWS Cloud (déploiement de référence Quick Start)
- Types de volumes Amazon EBS (EC2 documentation Amazon)
- Utilisation FSx d'un serveur de fichiers Windows avec Microsoft SQL Server (FSx documentation Amazon)
- Qu'est-ce que c'est AWS Backup ? (AWS Backup documentation)
- AWS Windows AMIs (EC2 documentation Amazon)

AWS Conseils prescriptifs

- Bonnes pratiques pour le déploiement de Microsoft SQL Server sur Amazon EC2
- EC2 Sauvegarde et restauration Amazon avec instantanés et AMIs

- Placer tempdb dans un magasin d'instances

Articles de blog et actualités

- Stockez facilement vos sauvegardes SQL Server dans Amazon S3 à l'aide de File Gateway
- Surveillez les coûts de transfert de données liés à Amazon S3 Replication
- Déploiement multirégional de SQL Server à l'aide de groupes de disponibilité distribués
- Notes de terrain : Création d'une architecture multirégionale pour SQL Server à l'aide de FCI et de groupes de disponibilité distribués
- Amazon propose EC2 désormais Microsoft SQL Server sur Microsoft Windows Server 2022 AMIs

Documentation SQL Server

- Éditions et fonctionnalités prises en charge de SQL Server

Annexe : Types de stockage SSD Amazon EBS

Amazon Elastic Block Store (Amazon EBS) fournit les volumes sauvegardés sur disque SSD (Solid State Drive) suivants. Pour obtenir les informations les plus récentes, consultez les types de volumes Amazon EBS dans la EC2 documentation Amazon.

	General Purpose SSD		Provisioned IOPS SSD		
Type de volume	`gp3`	`gp2`	`io2 Block Express`[1]	`io2`	`io1`
Durabilité	Durabilité de 99,8 % à 99,9 % (taux de défaillance annuel de 0,1 % à 0,2 %)	Durabilité de 99,8 % à 99,9 % (taux de défaillance annuel de 0,1 % à 0,2 %)	Durabilité de 99,999 % (taux de défaillance annuel de 0,001 %)	Durabilité de 99,999 % (taux de défaillance annuel de 0,001 %)	Durabilité de 99,8 % à 99,9 % (taux de défaillance annuel de 0,1 % à 0,2 %)
Cas d'utilisation	• Applications interactives à faible latence • Environnements de développement et de test	• Applications interactives à faible latence • Environnements de développement et de test	Charges de travail nécessitant : • Une latence moyenne inférieure à la milliseconde • Performance IOPS soutenue • Plus de 64 000 IOPS ou 1 000	• Charges de travail nécessitant des performances d'IOPS soutenues ou supérieures à 16 000 IOPs • Charges de travail de base de données à fort taux d'I/O.	• Charges de travail nécessitant des performances d'IOPS soutenues ou supérieures à 16 000 IOPs • Charges de travail de base de données à fort taux d'I/O.

	General Purpose SSD		Provisioned IOPS SSD		
			Mbits/s de débit		
Taille du volume	1 Gio – 16 Tio	1 Gio – 16 Tio	4 GiB — 64 TiB	4 GiB — 16 TiB	4 GiB — 16 TiB
Nombre maximal d'IOPS par volume (16 Kio d'E/S)	16,000	16,000	256 000	64 000 [2]	64 000 [2]
Débit maximal par volume	1,000 Mio/s	250 Mbits/s [3]	4 000 Mio/s	1 000 Mbits/s [2]	1 000 Mbits/s [2]
Multi-Attach Amazon EBS	Non pris en charge	Non pris en charge	Pris en charge	Pris en charge	Pris en charge
Volume de démarrage	Pris en charge	Pris en charge	Pris en charge	Pris en charge	Pris en charge

[1] Les `io2 Block Express` volumes sont pris en charge uniquement avec les instances R5b. `io2`les volumes attachés à une `R5b` instance pendant ou après le lancement s'exécutent automatiquement sur Block Express. Pour plus d'informations, consultez les volumes io2 Block Express dans la EC2 documentation Amazon.

[2] Les IOPS et le débit maximaux ne sont garantis que sur les instances basées sur le système Nitro dotées de plus de 32 000 IOPS. D'autres instances garantissent jusqu'à 32,000 IOPS et 500 Mio/s. Les volumes `io1` créés avant le 6 décembre 2017 et qui n'ont pas été modifié depuis leur création peuvent ne pas atteindre des performances optimales à moins que vous ne modifiez le volume.

[3] La limite de débit est comprise entre 128, MiB/s and 250 MiB/s, depending on the volume size. Volumes smaller than or equal to 170 GiB deliver a maximum throughput of 128 MiB/s. Volumes larger than 170 GiB but smaller than 334 GiB deliver a maximum throughput of 250 MiB/s if burst credits are available. Volumes larger than or equal to 334 GiB deliver 250 MiB/s quels que soient

les crédits de rafale. gp2les volumes créés avant le 3 décembre 2018 et qui n'ont pas été modifiés depuis leur création risquent de ne pas atteindre leurs performances optimales, sauf si vous modifiez le volume.

Historique du document

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

Modification	Description	Date
Publication initiale	—	28 février 2022

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 l'édition compatible avec Amazon Aurora 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 dans le. AWS Cloud
- Racheter (rachat) : optez pour un autre produit, généralement en passant d'une licence traditionnelle à un modèle SaaS. Exemple : migrez votre système de gestion de la relation client (CRM) vers Salesforce.com.
- Réhéberger (lift and shift) : transférez une application vers le cloud sans apporter de modifications pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Oracle sur une instance EC2 dans le. AWS Cloud
- Relocaliser (lift and shift au niveau de l'hyperviseur) : transférez l'infrastructure vers le cloud sans acheter de nouveau matériel, réécrire des applications ou modifier vos opérations existantes. Vous migrez des serveurs d'une plateforme sur site vers un service cloud pour la même plateforme. Exemple : migrer une Microsoft Hyper-V application vers AWS.
- Retenir : conservez les applications dans votre environnement source. Il peut s'agir d'applications nécessitant une refactorisation majeure, que vous souhaitez retarder, et

d'applications existantes que vous souhaitez retenir, car rien ne justifie leur migration sur le plan commercial.

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

A

ABAC

Voir contrôle d'accès basé sur les attributs.

services abstraits

Consultez la section Services gérés.

ACIDE

Voir atomicité, consistance, isolation, durabilité.

migration active-active

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

migration active-passive

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

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 son AIOps utilisation dans la stratégie de AWS migration, consultez 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.

CCo E

Voir le Centre d'excellence du cloud.

CDC

Voir 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 du 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 CCoarticles électroniques du blog sur la stratégie AWS Cloud d'entreprise.

cloud computing

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

modèle d'exploitation du cloud

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

étapes d'adoption du cloud

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

- Projet : exécution de quelques projets liés au cloud à des fins de preuve de concept et d'apprentissage
- Base : réaliser des investissements fondamentaux pour accélérer votre adoption du cloud (par exemple, créer une zone de landing zone, définir un CCo E, établir 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 publié sur le blog AWS Cloud Enterprise Strategy. Pour plus d'informations sur leur lien avec la stratégie de AWS migration, consultez le guide de préparation à la migration.

CMDB

Consultez la base de données de gestion des configurations.

référentiel de code

Emplacement où le code source et d'autres ressources, comme la documentation, les exemples et les scripts, sont stockés et mis à jour par le biais de processus de contrôle de version. Les référentiels cloud courants incluent GitHub ouBitbucket Cloud. Chaque version du code est appelée branche. Dans une structure de microservice, chaque référentiel est consacré à une seule fonctionnalité. Un 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, Amazon SageMaker AI fournit des algorithmes de traitement d'image pour les CV.

dérive de configuration

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

base de données de gestion des configurations (CMDB)

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

pack de conformité

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

intégration continue et livraison continue (CI/CD)

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

CV

Voir vision par ordinateur.

D

données au repos

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

classification des données

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

Voir reprise après sinistre.

détection de dérive

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

DVSM

Voir la cartographie de la chaîne de valeur du développement.

E

EDA

Voir analyse exploratoire des données.

EDI

Voir échange de données informatisé.

informatique de périphérie

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

échange de données informatisé (EDI)

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

chiffrement

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

clé de chiffrement

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

endianisme

Ordre 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 CI/CD pipeline, l'environnement de production est le dernier environnement de déploiement.
- Environnements supérieurs : tous les environnements accessibles aux utilisateurs autres que l'équipe de développement principale. Ils peuvent inclure un environnement de production, des environnements de préproduction et des environnements pour les tests d'acceptation par les utilisateurs.

épopée

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

ERP

Voir Planification des ressources d'entreprise.

analyse exploratoire des données (EDA)

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

F

tableau des faits

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

échouer rapidement

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

limite d'isolation des défauts

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

branche de fonctionnalités

Voir succursale.

fonctionnalités

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

importance des fonctionnalités

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

transformation de fonctionnalité

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

invitation en quelques coups

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

FGAC

Découvrez le contrôle d'accès détaillé.

contrôle d'accès détaillé (FGAC)

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

migration instantanée (flash-cut)

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

FM

Voir le modèle de fondation.

modèle de fondation (FM)

Un vaste réseau neuronal d'apprentissage profond qui s'est entraîné sur d'énormes ensembles de données généralisées et non étiquetées. FMs sont capables d'effectuer une grande variété de tâches générales, telles que comprendre le langage, générer du texte et des images et converser en langage naturel. Pour plus d'informations, voir Que sont les modèles de base ?

G

IA générative

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

blocage géographique

Voir les restrictions géographiques.

restrictions géographiques (blocage géographique)

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

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

Flux de travail Gitflow

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

image dorée

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

stratégie inédite

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

barrière de protection

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

H

HA

Découvrez la haute disponibilité.

migration de base de données hétérogène

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

haute disponibilité (HA)

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

modernisation des historiens

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

données de rétention

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

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.

IIo T

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 AWS de référence de sécurité recommande de configurer votre compte réseau avec les fonctions entrantes, sortantes et d'inspection VPCs afin de protéger l'interface bidirectionnelle entre votre application et l'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

Un 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, voir Élaboration d'une stratégie de transformation numérique de l'Internet des objets (IIoT) industriel.

VPC d'inspection

Dans une architecture AWS multi-comptes, un VPC centralisé qui gère les inspections du trafic réseau VPCs entre (identique ou Régions AWS différent), Internet et les réseaux locaux. L'architecture AWS de référence de sécurité recommande de configurer votre compte réseau avec les fonctions entrantes, sortantes et d'inspection VPCs afin de protéger l'interface bidirectionnelle entre votre application et l'Internet en général.

Internet des objets (IoT)

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

interprétabilité

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

IoT

Voir Internet des objets.

Bibliothèque d'informations informatiques (ITIL)

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

gestion des services informatiques (ITSM)

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

ITIL

Consultez la bibliothèque d'informations informatiques.

ITSM

Voir Gestion des services informatiques.

L

contrôle d'accès basé sur des étiquettes (LBAC)

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

zone de destination

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

grand modèle de langage (LLM)

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

migration de grande envergure

Migration de 300 serveurs ou plus.

LBAC

Voir contrôle d'accès basé sur des étiquettes.

principe de moindre privilège

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

lift and shift

Voir 7 Rs.

système de poids faible

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

LLM

Voir le grand modèle de langage.

environnements inférieurs

Voir environnement.

M

machine learning (ML)

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

branche principale

Voir succursale.

malware

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

services gérés

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

système d'exécution de la fabrication (MES)

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

MAP

Voir Migration Acceleration Program.

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 lorsqu'il fonctionne. 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/d'abonnement, pour les appareils IoT aux ressources limitées.

microservice

Un petit service indépendant qui communique via un réseau bien défini APIs 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 en utilisant Lightweight. APIs 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 des responsables des opérations, des analystes commerciaux et des propriétaires, des ingénieurs de migration, des développeurs et DevOps des professionnels travaillant dans le cadre de sprints. Entre 20 et 50 % du portefeuille d'applications d'entreprise est constitué de modèles répétés qui peuvent être optimisés par une approche d'usine. Pour plus d'informations, veuillez consulter la rubrique discussion of migration factories et le guide Cloud Migration Factory dans cet ensemble de contenus.

métadonnées de migration

Informations relatives à l'application et au serveur nécessaires pour finaliser la migration. Chaque modèle de migration nécessite un ensemble de métadonnées de migration différent. Les exemples de métadonnées de migration incluent le sous-réseau cible, le groupe de sécurité et le AWS compte.

modèle de migration

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

Évaluation du portefeuille de migration (MPA)

Outil en ligne qui fournit des informations pour valider l'analyse de rentabilisation en faveur de la migration vers le. AWS Cloud La MPA propose une évaluation détaillée du portefeuille (dimensionnement approprié des serveurs, tarification, comparaison du coût total de possession,

analyse des coûts de migration), ainsi que la planification de la migration (analyse et collecte des données d'applications, regroupement des applications, priorisation des migrations et planification des vagues). L'outil MPA (connexion requise) est disponible gratuitement pour tous les AWS consultants et consultants APN Partner.

Évaluation de la préparation à la migration (MRA)

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

stratégie de migration

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

ML

Voir apprentissage automatique.

modernisation

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

évaluation de la préparation à la modernisation

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

applications monolithiques (monolithes)

Applications qui s'exécutent en tant que service unique avec des processus étroitement couplés. Les applications monolithiques ont plusieurs inconvénients. Si une fonctionnalité de

l'application connaît un pic de demande, l'architecture entière doit être mise à l'échelle. L'ajout ou l'amélioration des fonctionnalités d'une application monolithique devient également plus complexe lorsque la base de code s'élargit. Pour résoudre ces problèmes, vous pouvez utiliser une architecture de microservices. Pour plus d'informations, veuillez consulter Decomposing monoliths into microservices.

MPA

Voir Évaluation du portefeuille de migration.

MQTT

Voir Message Queuing Telemetry Transport.

classification multi-classes

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

infrastructure mutable

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

O

OAC

Voir Contrôle d'accès à l'origine.

OAI

Voir l'identité d'accès à l'origine.

OCM

Voir gestion du changement organisationnel.

migration hors ligne

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

OI

Consultez la section Intégration des opérations.

OLA

Voir l'accord au niveau opérationnel.

migration en ligne

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

OPC-UA

Voir Open Process Communications - Architecture unifiée.

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

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 peuvent accéder au contenu d'un compartiment S3 uniquement via une distribution spécifique CloudFront . Voir également OAC, qui fournit un contrôle d'accès plus précis et amélioré.

ORR

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

DE

Voir technologie opérationnelle.

VPC sortant (de sortie)

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

P

limite des autorisations

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

informations personnelles identifiables (PII)

Informations qui, lorsqu'elles sont consultées directement ou associées à d'autres données connexes, peuvent être utilisées pour déduire raisonnablement l'identité d'une personne. Les exemples d'informations personnelles incluent les noms, les adresses et les informations de contact.

PII

Voir les informations personnelles identifiables.

manuel stratégique

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

PLC

Voir contrôleur logique programmable.

PLM

Consultez la section Gestion du cycle de vie des produits.

policy

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

persistance polyglotte

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

évaluation du portefeuille

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

predicate

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

prédicat pushdown

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

contrôle préventif

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

principal

Entité AWS capable d'effectuer des actions et d'accéder aux ressources. Cette entité est généralement un utilisateur root pour un Compte AWS rôle IAM ou un utilisateur. Pour plus

d'informations, veuillez consulter la rubrique Principal dans Termes et concepts relatifs aux rôles, dans la documentation IAM.

confidentialité dès la conception

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

zones hébergées privées

Conteneur contenant des informations sur la manière dont vous souhaitez qu'Amazon Route 53 réponde aux requêtes DNS pour un domaine et ses sous-domaines au sein d'un ou de plusieurs VPCs domaines. Pour plus d'informations, veuillez consulter Working with private hosted zones dans la documentation Route 53.

contrôle proactif

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

gestion du cycle de vie des produits (PLM)

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

environnement de production

Voir environnement.

contrôleur logique programmable (PLC)

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

chaînage rapide

Utiliser le résultat d'une invite LLM comme entrée pour l'invite suivante afin de générer de meilleures réponses. Cette technique est utilisée pour décomposer une tâche complexe en sous-tâches ou pour affiner ou développer de manière itérative une réponse préliminaire. Cela

permet d'améliorer la précision et la pertinence des réponses d'un modèle et permet d'obtenir des résultats plus précis et personnalisés.

pseudonymisation

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

publish/subscribe (pub/sub)

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

Q

plan de requête

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

régression du plan de requêtes

Le cas où un optimiseur de service de base de données choisit un plan moins optimal qu'avant une modification donnée de l'environnement de base de données. Cela peut être dû à des changements en termes de statistiques, de contraintes, de paramètres d'environnement, de liaisons de paramètres de requêtes et de mises à jour du moteur de base de données.

R

Matrice RACI

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

RAG

Voir Retrieval Augmented Generation.

rançongiciel

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

Matrice RASCI

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

RCAC

Voir contrôle d'accès aux lignes et aux colonnes.

réplica en lecture

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

réarchitecte

Voir 7 Rs.

objectif de point de récupération (RPO)

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

objectif de temps de récupération (RTO)

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

refactoriser

Voir 7 Rs.

Région

Un ensemble de AWS ressources dans une zone géographique. Chacun Région AWS est isolé et indépendant 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 AWS Cloud Résilience.

politique basée sur les ressources

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

matrice responsable, redevable, consulté et informé (RACI)

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

contrôle réactif

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

retain

Voir 7 Rs.

se retirer

Voir 7 Rs.

Génération augmentée de récupération (RAG)

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

rotation

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

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

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

RPO

Voir l'objectif du point de récupération.

RTO

Voir l'objectif en matière de temps de rétablissement.

runbook

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

S

SAML 2.0

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

SCADA

Voir Contrôle de supervision et acquisition de données.

SCP

Voir la politique de contrôle des services.

secret

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

sécurité dès la conception

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

contrôle de sécurité

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

renforcement de la sécurité

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

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

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

automatisation des réponses de sécurité

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

chiffrement côté serveur

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

Politique de contrôle des services (SCP)

Politique qui fournit un contrôle centralisé des autorisations pour tous les comptes d'une organisation dans AWS Organizations. SCPs définissez des garde-fous ou des limites aux actions qu'un administrateur peut déléguer à des utilisateurs ou à des rôles. Vous pouvez les utiliser SCPs comme listes d'autorisation ou de refus pour spécifier 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, voir 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.

invite du système

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

T

tags

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

Un hub de transit réseau que vous pouvez utiliser pour interconnecter vos réseaux VPCs et ceux 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

Une connexion entre deux VPCs qui vous permet d'acheminer le trafic en utilisant des adresses IP privées. Pour plus d'informations, veuillez consulter la rubrique Qu'est-ce que l'appairage de VPC ? dans la documentation Amazon VPC.

vulnérabilités

Défaut logiciel ou matériel qui compromet la sécurité du système.

W

cache actif

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

données chaudes

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

fonction de fenêtre

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

charge de travail

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

flux de travail

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

VER

Voir écrire une fois, lire plusieurs.

WQF

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

écrire une fois, lire plusieurs (WORM)

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

Z

exploit Zero-Day

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

vulnérabilité « jour zéro »

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

invite Zero-Shot

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

application zombie

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

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