



Meilleures pratiques en matière de requêtes pour Amazon Redshift

AWS Conseils prescriptifs



AWS Conseils prescriptifs: Meilleures pratiques en matière de requêtes pour Amazon Redshift

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques commerciales et la présentation commerciale d'Amazon ne peuvent pas être utilisées en relation avec un produit ou un service extérieur à Amazon, d'une manière susceptible d'entraîner une confusion chez les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Introduction	1
Présentation	1
Public visé	1
Objectifs	1
Composants d'architecture	2
Facteurs de performance des requêtes	7
Propriétés de tableau	7
Clés de tri	7
Compression des données	8
Distribution de données	8
Maintenance des tables	8
Configuration de cluster	10
Type de nœud	10
Taille des nœuds, nombre de nœuds et tranches	10
Gestion de la charge de travail	10
Accélération des requêtes courtes	11
Requête SQL	11
Structure de requête	11
Compilation de code	12
Bonnes pratiques pour les tables	13
Comprendre le fonctionnement des clés de tri	13
Conseils pour le réglage des requêtes	13
Évaluer l'efficacité des clés de tri	14
Connaissez votre table	15
Choisissez le bon style de distribution des tables	15
Bonnes pratiques en matière de requêtes	17
Évitez d'utiliser l'instruction SELECT * FROM	17
Identifier les problèmes liés aux requêtes	17
Obtenez des informations récapitulatives sur votre requête	17
Évitez les jonctions croisées	18
Évitez les fonctions dans les prédicats de requête	18
Évitez les conversions de casting inutiles	19
Utiliser des expressions CASE pour des agrégations complexes	19
Utiliser des sous-requêtes	20

Utiliser des prédicats	20
Ajouter des prédicats pour filtrer les tables avec des jointures	21
Utilisez les opérateurs les moins chers pour les prédicats	21
Utiliser des clés de tri dans les clauses GROUP BY	21
Tirez parti des vues matérialisées	22
Faites attention aux colonnes des clauses GROUP BY et ORDER BY	22
Bonnes pratiques pour Redshift Spectrum	23
Pushdown des prédicats dans Redshift Spectrum	24
Conseils de réglage des requêtes pour Redshift Spectrum	25
Ressources	26
Historique du document	27
Glossaire	28
#	28
A	29
B	32
C	34
D	37
E	41
F	44
G	46
H	47
I	49
L	51
M	52
O	57
P	59
Q	63
R	63
S	66
T	70
U	72
V	72
W	73
Z	74
.....	lxxv

Meilleures pratiques en matière de requêtes pour Amazon Redshift

Ethan Stark, Amazon Web Services (AWS)

Juin 2024 ([historique du document](#))

Présentation

Ce guide fournit des recommandations et des bonnes pratiques pour optimiser les performances des requêtes et des tables dans [Amazon Redshift](#). Vous pouvez utiliser Amazon Redshift pour interroger des pétaoctets de données structurées et semi-structurées dans votre entrepôt de données et votre lac de données à l'aide du SQL standard. Ce guide fournit également une présentation des principaux composants de l'architecture d'un entrepôt de données Amazon Redshift. Ces connaissances, associées à la compréhension des facteurs de performance des requêtes tels que les propriétés des tables, la configuration du cluster et la structure des requêtes, peuvent vous aider à concevoir des tables et des requêtes efficaces pour votre entrepôt de données Amazon Redshift.

Public visé

Ce guide est destiné aux ingénieurs de données, aux architectes de données et aux analystes de données qui conçoivent ou utilisent des tables et des requêtes dans Amazon Redshift.

Objectifs

Ce guide peut vous aider, vous et votre organisation, à atteindre les objectifs suivants :

- Concevez des tables pour des opérations de stockage et de récupération de données optimales
- Requête de conception pour des performances optimales et des économies
- Optimisez les performances d'[Amazon Redshift Spectrum](#) pour interroger des données directement à partir de fichiers [sur Amazon Simple Storage Service \(Amazon S3\)](#)

Composants d'architecture d'un entrepôt de données Amazon Redshift

Nous vous recommandons d'avoir une connaissance de base des principaux composants de l'architecture d'un entrepôt de données Amazon Redshift. Ces connaissances peuvent vous aider à mieux comprendre comment concevoir vos requêtes et vos tables pour des performances optimales.

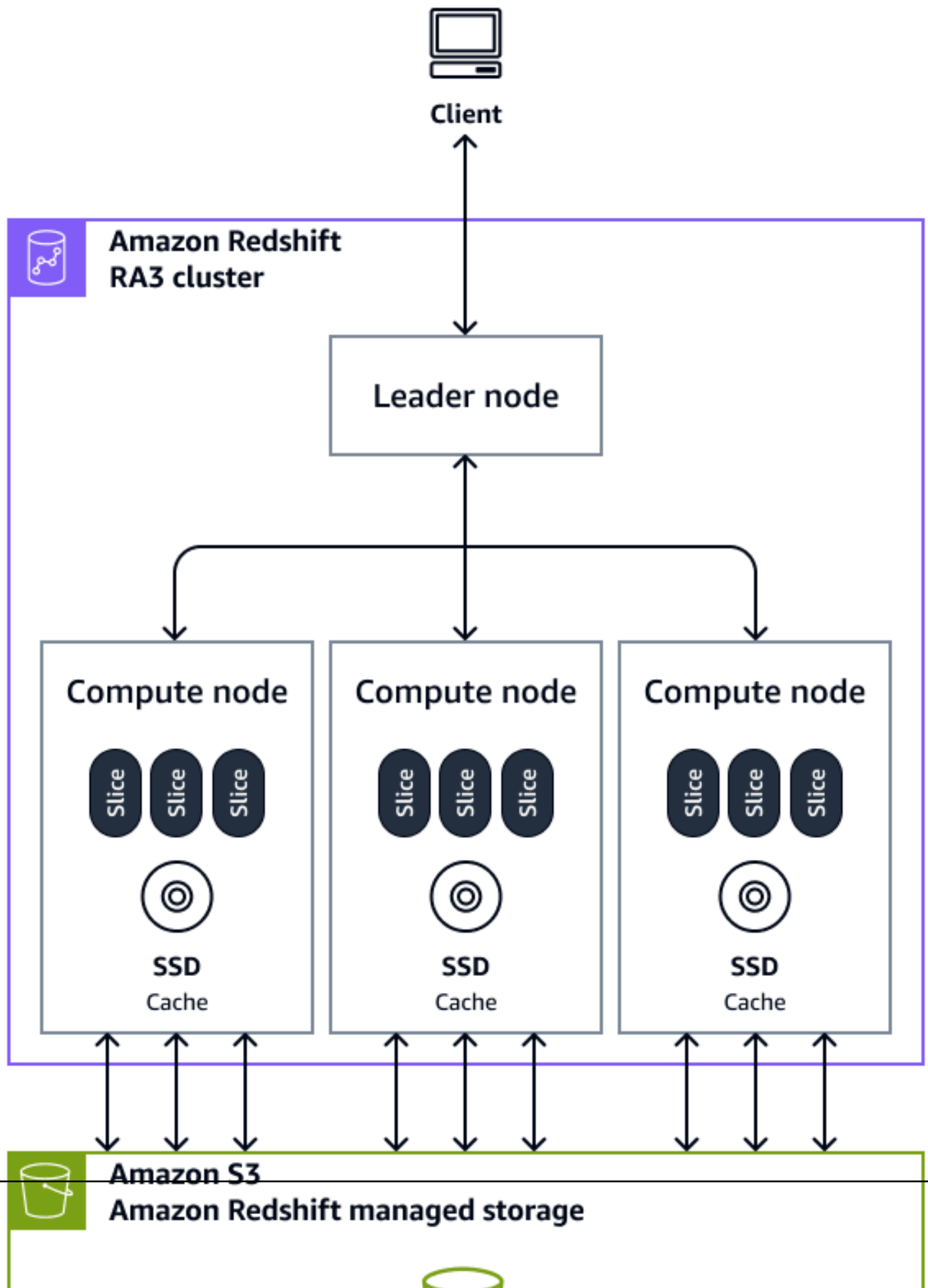
Un entrepôt de données dans Amazon Redshift comprend les principaux composants de l'architecture suivants :

- **Clusters** : un cluster, composé d'un ou de plusieurs nœuds de calcul, est le composant d'infrastructure principal d'un entrepôt de données Amazon Redshift. Les nœuds de calcul sont transparents pour les applications externes, mais votre application cliente n'interagit directement qu'avec le nœud principal. Un cluster typique possède au moins deux nœuds de calcul. Les nœuds de calcul sont coordonnés par le biais du nœud principal.
- **Nœud principal** : un nœud principal gère les communications pour les programmes clients et tous les nœuds de calcul. Un nœud principal prépare également les plans d'exécution d'une requête chaque fois qu'une requête est soumise à un cluster. Lorsque les plans sont prêts, le nœud principal compile le code, le distribue aux nœuds de calcul, puis affecte des tranches de données à chaque nœud de calcul pour traiter les résultats de la requête.
- **Nœud de calcul** : un nœud de calcul exécute une requête. Le nœud principal compile le code pour les éléments individuels du plan afin d'exécuter la requête et attribue le code aux nœuds de calcul individuels. Les nœuds de calcul exécutent le code compilé et renvoient les résultats intermédiaires au nœud principal pour l'agrégation finale. Chaque nœud de calcul possède son propre processeur, sa propre mémoire et son propre stockage sur disque. Lorsque votre charge de travail augmente, vous pouvez augmenter la capacité de calcul et la capacité de stockage d'un cluster en augmentant le nombre de nœuds, en mettant à niveau le type de nœud, ou les deux.
- **Tranche de nœud** : un nœud de calcul est partitionné en unités appelées tranches. Chaque tranche d'un nœud de calcul reçoit une partie de la mémoire et de l'espace disque du nœud où elle traite une partie de la charge de travail assignée au nœud. Les tranches travaillent alors en parallèle pour terminer l'opération. Les données sont réparties entre les tranches en fonction du [style de distribution](#) et de la clé de distribution d'une table donnée. Une distribution uniforme des données permet à Amazon Redshift d'attribuer de manière uniforme les charges de travail aux tranches et de maximiser les avantages du traitement parallèle. Le nombre de tranches par nœud de calcul est

déterminé en fonction du type de nœud. Pour plus d'informations, consultez la section [Clusters et nœuds dans Amazon Redshift](#) dans la documentation Amazon Redshift.

- **Traitement massivement parallèle (MPP) :** Amazon Redshift utilise l'architecture MPP pour traiter rapidement les données, même les requêtes complexes et de grandes quantités de données. Plusieurs nœuds de calcul exécutent le même code de requête sur des portions de données afin d'optimiser le traitement parallèle.
- **Application client —** Amazon Redshift s'intègre à divers outils de chargement, d'extraction, de transformation et de chargement (ETL), de reporting de business intelligence (BI), d'exploration de données et d'analyse. Toutes les applications clientes communiquent avec le cluster uniquement via le nœud principal.

Le schéma suivant montre comment les composants de l'architecture d'un entrepôt de données Amazon Redshift fonctionnent ensemble pour accélérer les requêtes.



Le cycle de vie des requêtes comporte sept étapes :

1. Réception et analyse des requêtes :

- Le nœud principal reçoit la requête et analyse le SQL.
- L'analyseur produit un arbre de requête initial, qui représente la structure logique de la requête d'origine.
- Amazon Redshift introduit cet arbre de requêtes dans l'optimiseur de requêtes.

2. Optimisation des requêtes :

- L'optimiseur évalue la requête et, si nécessaire, la réécrit pour optimiser l'efficacité.
- Ce processus d'optimisation peut impliquer la création de plusieurs requêtes connexes pour en remplacer une seule.

3. Génération du plan de requêtes :

- L'optimiseur génère un plan de requête (ou plusieurs plans, si nécessaire) à exécuter.
- Le plan de requête spécifie les options d'exécution, telles que les types de jointure, l'ordre de jointure, les méthodes d'agrégation et les exigences de distribution des données.

4. Traduction du moteur d'exécution :

- Le moteur d'exécution traduit le plan de requête en étapes, segments et flux distincts :
 - Étape — Représente une opération individuelle requise lors de l'exécution de la requête. Les étapes peuvent être combinées pour permettre aux nœuds de calcul d'effectuer des requêtes, des jointures ou d'autres opérations de base de données.
 - Segment : combine plusieurs étapes qu'un seul processus peut exécuter. Il s'agit de la plus petite unité de compilation exécutable par une tranche de nœud de calcul. (Une tranche est l'unité de traitement parallèle dans Amazon Redshift.)
 - Stream : ensemble de segments répartis sur les tranches de nœuds de calcul disponibles.
- Le moteur d'exécution génère du code compilé en fonction de ces étapes, segments et flux. Le code compilé s'exécute plus rapidement que le code interprété et consomme moins de capacité de calcul.
- Le nœud principal diffuse le code compilé aux nœuds de calcul.

5. Exécution parallèle :

- Cette étape a lieu une fois pour chaque flux.
- Les tranches de nœuds de calcul exécutent des segments de requête en parallèle.

- Au cours de ce processus, Amazon Redshift optimise les communications réseau, l'utilisation de la mémoire et la gestion des disques afin de transmettre les résultats intermédiaires d'une étape du plan de requêtes à l'autre.
- Cette optimisation contribue à accélérer l'exécution des requêtes.

6. Traitement des flux :

- Cette étape a lieu une fois pour chaque flux.
- Le moteur crée des segments exécutables pour chaque flux, pour un traitement parallèle efficace.

7. Tri final et agrégation :

- Le nœud principal prend en charge tout tri ou agrégation final requis par la requête.
- Une fois terminé, le nœud leader renvoie les résultats au client.

Pour plus d'informations sur les composants de l'architecture, consultez la section [Architecture du système d'entrepôt de données](#) dans la documentation Amazon Redshift.

Facteurs de performance des requêtes pour Amazon Redshift

Un certain nombre de facteurs peut affecter les performances des requêtes. Les aspects suivants des opérations de vos données, de votre cluster et de votre base de données jouent tous un rôle dans la rapidité avec laquelle vos requêtes sont traitées :

- [Propriétés de tableau](#)
 - [Clés de tri](#)(Amazon Redshift Advisor)
 - [Compression des données](#)(automatique)
 - [Distribution de données](#)(automatique)
 - [Maintenance des tables](#)(automatique)
- [Configuration de cluster](#)
 - [Type de nœud](#)
 - [Taille des nœuds, nombre de nœuds et tranches](#)
 - [Gestion de la charge de travail](#)(automatique)
 - [Accélération des requêtes courtes](#)(automatique)
- [Requête SQL](#)
 - [Structure de requête](#)
 - [Compilation de code](#)

Propriétés de tableau

Les tables Amazon Redshift sont les unités fondamentales pour le stockage des données dans Amazon Redshift, et chaque table possède un ensemble de propriétés qui déterminent son comportement et son accessibilité. Ces propriétés incluent le tri, le style de distribution, le codage de compression et bien d'autres. Il est essentiel de comprendre ces propriétés pour optimiser les performances, la sécurité et la rentabilité des tables Amazon Redshift.

Clés de tri

Amazon Redshift stocke les données sur disque dans un ordre trié en fonction des clés de tri d'une table. L'optimiseur de requêtes et le processeur de requêtes utilisent les informations relatives à

l'emplacement des données dans un nœud de calcul pour réduire le nombre de blocs à scanner. Cela améliore considérablement la vitesse des requêtes en réduisant la quantité de données à traiter. Nous vous recommandons d'utiliser des touches de tri pour faciliter les filtres dans la WHERE clause. Pour plus d'informations, consultez la section [Utilisation des clés de tri](#) dans la documentation Amazon Redshift.

Compression des données

La compression des données réduit les besoins en stockage, ce qui réduit les E/S du disque et améliore les performances des requêtes. Lorsque vous exécutez une requête, les données compressées sont lues en mémoire puis décompressées lors de l'exécution de la requête. En chargeant moins de données en mémoire, Amazon Redshift peut allouer plus de mémoire à l'analyse des données. Le stockage en colonnes stockant des données similaires de manière séquentielle, Amazon Redshift peut appliquer des codages de compression adaptatifs spécifiquement liés aux types de données en colonnes. Le meilleur moyen d'activer la compression des données sur les colonnes d'un tableau est d'utiliser l'AUTOoption d'Amazon Redshift qui permet d'appliquer des codages de compression optimaux lorsque vous chargez le tableau avec des données. Pour en savoir plus sur l'utilisation de la compression automatique des données, consultez la section [Chargement de tables avec compression automatique](#) dans la documentation Amazon Redshift.

Distribution de données

Amazon Redshift stocke les données sur les nœuds de calcul conformément au style de distribution d'une table. Lorsque vous exécutez une requête, l'optimiseur de requête redistribue les données sur les nœuds de calcul en fonction des besoins afin d'effectuer des jointures et des agrégations. Le choix du style de distribution adapté à une table permet de réduire l'impact de l'étape de la redistribution en plaçant les données à l'emplacement souhaité avant que les jointures soient effectuées. Nous vous recommandons d'utiliser des clés de distribution pour faciliter les jointures les plus courantes. Pour plus d'informations, consultez la section [Utilisation des styles de distribution de données](#) dans la documentation Amazon Redshift.

Maintenance des tables

Bien qu'Amazon Redshift fournisse des performances de pointe prêtes à l'emploi pour la plupart des charges de travail, le bon fonctionnement des clusters Amazon Redshift nécessite une maintenance. La mise à jour et la suppression de données créent des lignes mortes qui doivent être éliminées, et même les tables à ajout uniquement doivent être utilisées si l'ordre d'ajout n'est pas conforme à la clé de tri.

Vacuum

Le processus d'aspiration dans Amazon Redshift est essentiel pour le bon fonctionnement et la maintenance de votre cluster Amazon Redshift. Cela affecte également les performances des requêtes. Comme les suppressions et les mises à jour signalent les anciennes données mais ne les suppriment pas réellement, vous devez utiliser l'aspirateur pour récupérer l'espace disque occupé par les lignes du tableau marquées pour suppression par les opérations précédentes. UPDATE DELETE Amazon Redshift peut automatiquement trier et effectuer une VACUUM DELETE opération sur les tables en arrière-plan.

Pour nettoyer les tables après un chargement ou une série de mises à jour incrémentielles, vous pouvez également exécuter la commande VACUUM, sur toute la base de données ou sur chaque table individuelle. Si les tables sont dotées de touches de tri et que le chargement des tables n'est pas optimisé pour le tri au fur et à mesure de leur insertion, vous devez utiliser des aspirateurs pour récupérer les données (ce qui peut être crucial pour les performances). Pour plus d'informations, consultez les [tables d'aspiration](#) dans la documentation Amazon Redshift.

Analyser

L'ANALYZE opération met à jour les métadonnées statistiques sur les tables d'une base de données Amazon Redshift. Tenir les statistiques à jour améliore les performances des requêtes en permettant au planificateur de requête de choisir les plans optimaux. Amazon Redshift surveille en permanence votre base de données et effectue automatiquement des opérations d'analyse en arrière-plan. Pour minimiser l'impact sur les performances de votre système, l'ANALYZE opération s'exécute automatiquement pendant les périodes où les charges de travail sont faibles. Si vous choisissez d'exécuter explicitement ANALYZE, procédez comme suit :

- Exécutez la ANALYZE commande avant d'exécuter des requêtes.
- Exécutez régulièrement la ANALYZE commande sur la base de données à la fin de chaque cycle normal de chargement ou de mise à jour.
- Exécutez la ANALYZE commande sur les nouvelles tables que vous créez et sur les tables ou colonnes existantes qui subissent des modifications importantes.
- Envisagez d'exécuter ANALYZE des opérations selon des plannings différents pour différents types de tables et de colonnes, en fonction de leur utilisation dans les requêtes et de leur propension à changer.
- Pour économiser du temps et des ressources de cluster, utilisez la PREDICATE COLUMNS clause lorsque vous exécutez la ANALYZE commande.

Configuration de cluster

Un cluster est un ensemble de nœuds qui effectuent le stockage et le traitement proprement dits des données. Il est essentiel de configurer correctement votre cluster Amazon Redshift si vous souhaitez atteindre les objectifs suivants :

- Évolutivité et simultanété élevées
- Utilisation efficace d'Amazon Redshift
- De meilleures performances
- moindre coût

Type de nœud

Un cluster Amazon Redshift peut utiliser l'un des différents types de nœuds (RA3 DC2, et DS2). Chaque type de nœud propose différentes tailles et limites vous permettant de mettre à l'échelle votre cluster de façon appropriée. La taille de nœud détermine la capacité de stockage, la mémoire, le processeur et le prix de chaque nœud du cluster. L'optimisation des coûts et des performances commence par le choix du type et de la taille de nœud appropriés. Pour plus d'informations sur les types de nœuds, consultez la section [Présentation des clusters Amazon Redshift](#) dans la documentation Amazon Redshift.

Taille des nœuds, nombre de nœuds et tranches

Un nœud de calcul est divisé en tranches. Un plus grand nombre de nœuds signifie davantage de processeurs et de tranches, ce qui permet à vos requêtes de traiter plus rapidement en exécutant des parties de la requête simultanément sur les tranches. Cependant, un plus grand nombre de nœuds signifie également des dépenses plus importantes. Cela signifie que vous devez trouver l'équilibre entre le coût et les performances adapté à votre système. Pour plus d'informations sur l'architecture du cluster Amazon Redshift, consultez la section [Architecture du système d'entrepôt de données](#) dans la documentation Amazon Redshift.

Gestion de la charge de travail

La gestion de la charge de travail Amazon Redshift (WLM) permet aux utilisateurs de gérer de manière flexible les files d'attente avec des priorités afin que les requêtes courtes et rapides ne restent pas bloquées dans des files d'attente derrière des requêtes de longue durée. Le WLM

automatique utilise des algorithmes d'apprentissage automatique (ML) pour profiler les requêtes et les placer dans la file d'attente appropriée avec les ressources appropriées, tout en gérant la simultanéité des requêtes et l'allocation de mémoire. Pour plus d'informations sur le WLM, consultez [Implémentation de la gestion de charge de travail](#) dans la documentation Amazon Redshift.

Accélération des requêtes courtes

L'accélération des requêtes courtes (SQA) donne la priorité aux requêtes de courte durée par rapport aux requêtes de longue durée. SQA exécute les requêtes dans un espace dédié afin que les requêtes SQA ne soient pas obligées d'attendre dans des files d'attente derrière des requêtes plus longues. SQA priorise uniquement les requêtes de courte durée qui sont placées dans une file d'attente définie par l'utilisateur. Si vous utilisez la SQA, les requêtes de courte durée commencent à s'exécuter plus rapidement et vous pouvez obtenir des résultats plus rapidement. Si vous activez SQA, vous pouvez réduire ou éliminer les files d'attente WLM dédiées aux requêtes de courte durée. En outre, les requêtes de longue durée n'ont pas besoin de se disputer des emplacements dans une file d'attente WLM. Cela signifie que vous pouvez configurer vos files d'attente WLM pour utiliser moins de slots de requêtes. Si vous réduisez la simultanéité, le débit des requêtes est augmenté et les performances globales du système sont améliorées pour la plupart des charges de travail. Pour plus d'informations sur la SQA, consultez la section [Utilisation de l'accélération des requêtes courtes](#) dans la documentation Amazon Redshift.

Requête SQL

Une requête de base de données est une demande de données provenant d'une base de données. La demande doit être envoyée dans un cluster Amazon Redshift à l'aide de SQL. Amazon Redshift prend en charge les outils clients SQL qui se connectent via Java Database Connectivity (JDBC) et Open Database Connectivity (ODBC). Vous pouvez utiliser la plupart des outils clients SQL qui prennent en charge les pilotes ODBC ou JDBC.

Structure de requête

La façon dont votre requête est écrite influe considérablement sur ses performances. Nous vous recommandons d'écrire des requêtes pour traiter et renvoyer le moins de données nécessaire pour répondre à vos besoins. Pour plus d'informations sur la manière de structurer vos requêtes, consultez la section [Meilleures pratiques pour la conception de requêtes Amazon Redshift](#) de ce guide.

Compilation de code

Amazon Redshift génère et compile le code pour chaque plan d'exécution des requêtes. Le code compilé s'exécute plus rapidement car il supprime la surcharge liée à l'utilisation d'un interpréteur. Vous avez généralement des frais généraux la première fois que le code est généré et compilé. Par conséquent, la première fois que vous exécutez une requête, ses performances peuvent prêter à confusion. Les frais généraux peuvent être particulièrement visibles lorsque vous exécutez des requêtes ponctuelles. Nous vous recommandons d'exécuter la requête une seconde fois afin de déterminer ses performances habituelles.

Amazon Redshift utilise un service de compilation sans serveur pour dimensionner les compilations de requêtes au-delà des ressources de calcul d'un cluster Amazon Redshift. Les segments de code compilés sont mis en cache localement sur le cluster et dans un cache pratiquement illimité. Ce cache persiste après le redémarrage du cluster. Les appels ultérieurs de la même requête s'exécutent plus rapidement car ils peuvent ignorer la phase de compilation. Le cache n'est pas compatible entre les versions d'Amazon Redshift. Par conséquent, le code est recompilé lorsque les requêtes s'exécutent après une mise à niveau de la version. En utilisant un service de compilation évolutif, Amazon Redshift peut compiler du code en parallèle pour offrir des performances rapides. L'ampleur de l'accélération de la charge de travail dépend de la complexité et de la simultanéité des requêtes.

Bonnes pratiques pour la conception de tables Amazon Redshift

Cette section fournit une vue d'ensemble des meilleures pratiques en matière de conception de tables de base de données. Nous vous recommandons de suivre ces bonnes pratiques pour optimiser les performances et l'efficacité des requêtes.

Comprendre le fonctionnement des clés de tri

Amazon Redshift stocke vos données sur le disque dans un ordre trié selon la clé de tri. L'optimiseur de requête Amazon Redshift utilise l'ordre de tri lorsqu'il détermine les plans de requête optimaux. Pour utiliser efficacement les clés de tri, nous vous recommandons de procéder comme suit :

- Gardez le tableau rangé autant que possible.
- Utilisez le VACUUM tri pour rétablir des performances optimales.
- Évitez de compresser la colonne clé de tri.
- Si la clé de tri est compressée et si le `sortkey1_skew` ratio est significativement élevé, recréez le tableau sans activer la compression sur la clé de tri.
- Évitez d'appliquer une fonction aux colonnes clés de tri. Par exemple, dans la requête suivante, la colonne clé de `trans_dt` : `TIMESTAMPTZ` tri n'est pas utilisée si vous la convertissez en `DATE` :

```
select order_id, order_amt
from sales
where trans_dt::date = '2021-01-08'::date
```

- Effectuez INSERT les opérations dans l'ordre des clés de tri.
- Utilisez les clés de tri dans la GROUP BY clause lorsque cela est possible.

Conseils pour le réglage des requêtes

Nous vous recommandons de procéder comme suit pour affiner vos requêtes :

- Pour une efficacité optimale, commandez toujours les clés de tri composées de la cardinalité la plus faible à la plus haute cardinalité.

- Si la clé principale d'une clé de tri composée est relativement unique (c'est-à-dire qu'elle a une cardinalité élevée), évitez d'ajouter des colonnes supplémentaires à votre clé de tri. L'ajout de colonnes supplémentaires a peu d'impact sur les performances des requêtes, mais augmente les coûts de maintenance.

Évaluer l'efficacité des clés de tri

Pour optimiser vos requêtes, vous devez être en mesure d'évaluer l'efficacité de vos requêtes. Nous vous recommandons d'utiliser la vue [SVL_QUERY_SUMMARY](#) pour trouver des informations générales sur l'exécution d'une requête. Dans cette vue, vous pouvez utiliser l'attribut IS_RRSCAN pour déterminer si une étape EXPLAIN du plan utilise un scan à plage restreinte. Vous pouvez également utiliser l'attribut rows_pre_filter pour déterminer la sélectivité d'une clé de tri.

Vous pouvez également utiliser une vue d'administration GitHub appelée [v_my_last_query_summary](#). La vue affiche les informations relatives à la dernière requête exécutée.

L'instruction suivante indique comment trouver des informations générales sur l'exécution d'une requête.

```
select lpad(' ',stm+seg+step) || label as label,
       rows,
       bytes,
       is_diskbased,
       is_rrscan,
       rows_pre_filter
from svl_query_summary
where query = pg_last_query_id()
order by stm, seg, step;
```

La requête précédente renvoie l'exemple de sortie suivant.

label	☐ rows	bytes	is_diskbased	is_rrscan	rows_pre_filter
scan tbl=163860 name=orders	☐ 1500000	24000000	f	f	1500000
project	☐ 1500000	0	f	f	0
project	☐ 1500000	0	f	f	0
hash tbl=968	☐ 1500000	24000000	f	f	0
scan tbl=163852 name=lineitem	☐ 6001215	144029160	f	t	6001215
project	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0
hjoin tbl=968	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0
project	☐ 6001215	0	f	f	0

Connaissez votre table

Il est important de comprendre les propriétés critiques de votre table. Pour en savoir plus sur votre table, procédez comme suit :

- Utilisez [PG_TABLE_DEF](#) pour afficher les informations relatives aux colonnes des tables.
- Utilisez [SVV_TABLE_INFO](#) pour afficher des informations plus complètes sur une table, notamment le biais de distribution des données, le biais de distribution des clés, la taille du tableau et les statistiques.

Choisissez le bon style de distribution des tables

Lorsque vous exécutez une requête, l'optimiseur de requête redistribue les lignes sur les nœuds de calcul en fonction des besoins afin d'effectuer les jointures et les agrégations. L'objectif de la sélection d'un style de distribution de table est de minimiser l'impact de l'étape de redistribution en localisant les données là où elles doivent se trouver avant d'exécuter la requête.

Nous recommandons l'approche suivante pour choisir le bon style de distribution de table :

- Évitez la diffusion et la redistribution dans un plan d'exécution des requêtes en colocalisant les lignes dans le même nœud. Par exemple, en sélectionnant un `DISTKEY`, vous pouvez répartir la table de faits et la table unidimensionnelle sur leurs colonnes communes. Choisissez la plus grande dimension en fonction de la taille de l'ensemble de données filtré. Comme seules les lignes utilisées dans la jointure doivent être distribuées, prenez en compte la taille du jeu de données après filtrage, et non la taille de la table.

- Assurez-vous que la colonne dans laquelle la clé de distribution est créée n'est pas asymétrique. Dans le cas contraire, un nœud de calcul pourrait effectuer des tâches plus lourdes que les autres. Si vous remarquez une asymétrie, pensez à modifier la colonne de la clé de distribution. Une colonne peut être considérée comme candidate pour une clé de distribution si ses valeurs sont distribuées uniformément ou si ses valeurs cardinales sont élevées.
- Si la table utilisée dans la condition de jointure est petite (moins de 1 Go), considérez le style de distribution ALL.
- Vous pouvez compresser la clé de distribution, mais vous devez éviter de compresser la colonne de la clé de tri (en particulier la première colonne de la clé de tri).

 Note

Si vous utilisez l'optimisation automatique des tables, il n'est pas nécessaire de choisir le style de distribution de votre table. Pour plus d'informations, consultez la section [Utilisation de l'optimisation automatique des tables](#) dans la documentation Amazon Redshift. Pour demander à Amazon Redshift de choisir le style de distribution approprié, spécifiez AUTO pour le style de distribution.

Bonnes pratiques pour la conception de requêtes Amazon Redshift

Cette section fournit une vue d'ensemble des meilleures pratiques en matière de conception de requêtes. Nous vous recommandons de suivre les meilleures pratiques décrites dans cette section pour optimiser les performances et l'efficacité des requêtes.

Évitez d'utiliser l'instruction SELECT * FROM

Nous vous recommandons d'éviter d'utiliser `SELECT * FROM` cette déclaration. Au lieu de cela, listez toujours les colonnes à analyser. Cela réduit le temps d'exécution des requêtes et analyse les coûts des requêtes Amazon Redshift Spectrum.

Exemple de ce qu'il faut éviter

```
select *  
from sales;
```

Exemple de bonne pratique

```
select sales_date, sales_amt  
from sales;
```

Identifier les problèmes liés aux requêtes

Nous vous recommandons de consulter la vue [STL_ALERT_EVENT_LOG](#) pour identifier et corriger les éventuels problèmes liés à votre requête.

Obtenez des informations récapitulatives sur votre requête

Nous vous recommandons d'utiliser les vues `SVL_QUERY_SUMMARY` et [SVL_QUERY_REPORT](#) pour obtenir des informations récapitulatives sur vos requêtes. Vous pouvez utiliser ces informations pour optimiser vos requêtes.

Évitez les jonctions croisées

Nous vous recommandons d'éviter d'utiliser des jonctions croisées, sauf en cas de nécessité absolue. Sans condition de jointure, les jointures croisées produisent le produit cartésien de deux tables. Les jointures croisées sont généralement exécutées sous forme de boucles imbriquées (le type de jointure le plus lent possible).

Exemple de ce qu'il faut éviter

```
select c.c_name,  
       n.n_name  
from tpch.customer c,  
     tpch.nation n;
```

Exemple de bonne pratique

```
select c.c_name,  
       n.n_name  
from tpch.customer c,  
join tpch.nation n  
  on n.n_nationkey = c.c_nationkey;
```

Évitez les fonctions dans les prédicats de requête

Nous vous recommandons d'éviter d'utiliser des fonctions dans les prédicats de requête. L'utilisation de fonctions dans les prédicats de requête peut avoir un impact négatif sur les performances, car les fonctions ajoutent généralement une charge de traitement supplémentaire à chaque ligne et ralentissent l'exécution globale de la requête.

Exemple de ce qu'il faut éviter

```
select sum(o_totalprice)  
from tpch.orders  
where datepart(year, o_orderdate) = 1992;
```

Exemple de bonne pratique

```
select sum(o_totalprice)  
from tpch.orders
```

```
where o_orderdate between '1992-01-01' and '1992-12-31';
```

Évitez les conversions de casting inutiles

Nous vous recommandons d'éviter d'utiliser une conversion par diffusion inutile sur les requêtes, car le casting des types de données prend du temps et des ressources et ralentit l'exécution des requêtes.

Exemple de ce qu'il faut éviter

```
select sum(o_totalprice)
from tpch.orders
where o_ordertime::date = '1992-01-01';
```

Exemple de bonne pratique

```
select sum(o_totalprice)
from tpch.orders
where o_ordertime between '1992-01-01 00:00:00' and '1992-12-31 23:59:59';
```

Utiliser des expressions CASE pour des agrégations complexes

Nous vous recommandons d'utiliser une [expression CASE](#) pour effectuer des agrégations complexes au lieu de sélectionner plusieurs fois dans la même table.

Exemple de ce qu'il faut éviter

```
select sum(sales_amt) as us_sales
from sales
where country = 'US';

select sum(sales_amt) as ca_sales
from sales
where country = 'CA';
```

Exemple de bonne pratique

```
select sum(case when country = 'US' then sales_amt end) as us_sales,
```

```
sum(case when country = 'CA' then sales_amt end) as ca_sales
from sales;
```

Utiliser des sous-requêtes

Nous vous recommandons d'utiliser des sous-requêtes lorsqu'une table de la requête est utilisée uniquement pour les conditions de prédicat et que la sous-requête renvoie un petit nombre de lignes (moins de 200 environ).

Exemple de ce qu'il faut éviter

Si une sous-requête renvoie moins de 200 lignes :

```
select sum(order_amt) as total_sales
from sales
where region_key IN
      (select region_key
       from regions
       where state = 'CA');
```

Exemple de bonne pratique

Si une sous-requête renvoie une valeur supérieure ou égale à 200 lignes :

```
select sum(o.order_amt) as total_sales
from sales o
join regions r
  on r.region_key = o.region_key
  and r.state = 'CA';
```

Utiliser des prédicats

Nous vous recommandons d'utiliser des prédicats pour restreindre le jeu de données autant que possible. Les prédicats sont utilisés en SQL pour filtrer et restreindre les données renvoyées dans une requête. En spécifiant des conditions dans un prédicat, vous pouvez spécifier les lignes qui doivent être incluses dans les résultats de la requête en fonction de conditions spécifiées. Cela vous permet de récupérer uniquement les données qui vous intéressent et d'améliorer l'efficacité et la précision de vos requêtes. Pour plus d'informations, consultez la section [Conditions](#) dans la documentation Amazon Redshift.

Ajouter des prédicats pour filtrer les tables avec des jointures

Nous vous recommandons d'ajouter des prédicats pour filtrer les tables qui participent aux jointures, même si les prédicats appliquent les mêmes filtres. L'utilisation de prédicats pour filtrer les tables comportant des jointures dans SQL peut améliorer les performances des requêtes en réduisant la quantité de données à traiter et en réduisant la taille du jeu de résultats intermédiaires. En spécifiant les conditions de l'opération de jointure dans la WHERE clause, le moteur d'exécution des requêtes peut éliminer les lignes qui ne répondent pas aux conditions avant leur jointure. Cela se traduit par un ensemble de résultats réduit et une exécution plus rapide des requêtes.

Exemple de ce qu'il faut éviter

```
select p.product_name, sum(o.order_amt)
from sales o
join product p
    on r.product_key = o.product_key
where o.order_date > '2022-01-01';
```

Exemple de bonne pratique

```
select p.product_name, sum(o.order_amt)
from sales o
join product p
    on p.product_key = o.product_key
    and p.added_date > '2022-01-01'
where o.order_date > '2022-01-01';
```

Utilisez les opérateurs les moins chers pour les prédicats

Dans le prédicat, utilisez les opérateurs les moins chers possibles. Les opérateurs de [conditions de comparaison](#) sont préférables aux opérateurs [LIKE](#). Les opérateurs sont toujours préférables aux opérateurs [SIMILAR TO](#) ou [POSIX](#).

Utiliser des clés de tri dans les clauses GROUP BY

Utilisez des clés de tri dans la GROUP BY clause afin que le planificateur de requêtes puisse utiliser une agrégation plus efficace. Une requête peut être éligible à une agrégation monophasée lorsque sa GROUP BY liste ne contient que des colonnes de clé de tri, dont l'une est également la clé de

distribution. Les colonnes de clé de tri de la GROUP BY liste doivent inclure la première clé de tri, suivie des autres clés de tri que vous souhaitez utiliser dans l'ordre des clés de tri.

Tirez parti des vues matérialisées

Si possible, réécrivez la requête en remplaçant le code complexe par une vue matérialisée, ce qui améliorera considérablement les performances de la requête. Pour plus d'informations, consultez la section [Création de vues matérialisées dans Amazon Redshift](#) dans la documentation Amazon Redshift.

Faites attention aux colonnes des clauses GROUP BY et ORDER BY

Si vous utilisez les deux ORDER BY clauses GROUP BY et, assurez-vous de placer les colonnes dans le même ordre dans GROUP BY les deux ORDER BY clauses. GROUP BY exige implicitement que les données soient triées. Si votre ORDER BY clause est différente, les données doivent être triées deux fois.

Exemple de ce qu'il faut éviter

```
select a, b, c, sum(d)
from a_table
group by b, c, a
order by a, b, c
```

Exemple de bonne pratique

```
select a, b, c, sum(d)
from a_table
group by a, b, c
order by a, b, c
```

Bonnes pratiques d'utilisation d'Amazon Redshift Spectrum

Cette section fournit une vue d'ensemble des meilleures pratiques relatives à l'utilisation d'[Amazon Redshift Spectrum](#). Nous vous recommandons de suivre les meilleures pratiques suivantes pour obtenir des performances optimales lorsque vous utilisez Redshift Spectrum :

- Sachez que les types de fichiers ont une influence significative sur les performances des requêtes Redshift Spectrum. Pour améliorer les performances, utilisez des fichiers codés en colonnes tels que ORC ou Parquet, et utilisez le format CSV uniquement pour les tableaux de très petites dimensions.
- Utilisez le partitionnement basé sur des préfixes pour tirer parti de l'élagage des partitions. Cela implique d'utiliser des filtres attachés aux partitions de votre lac de données.
- Redshift Spectrum s'adapte automatiquement pour traiter les demandes volumineuses, alors faites tout ce qui est possible dans Redshift Spectrum (par exemple, le pushdown des prédicats).
- Faites attention à partitionner les fichiers sur des colonnes fréquemment filtrées. Si les données sont partitionnées par une ou plusieurs colonnes filtrées, Redshift Spectrum peut tirer parti de l'élagage des partitions et éviter de scanner les partitions et les fichiers inutiles. Il est courant de les partitionner selon des critères temporels.
- Vous pouvez vérifier l'efficacité de vos partitions et l'efficacité de votre requête Redshift Spectrum à l'aide de la requête suivante.

```
Select query,
       segment,
       max(assigned_partitions) as total_partitions,
       max(qualified_partitions) as qualified_partitions
From svl_s3partition
Where query=pg_last_query_id()
Group by 1,2;
```

La requête précédente montre ce qui suit :

- `total_partitions` — Le nombre de partitions reconnues par AWS Glue Data Catalog
- `qualified_partitions` — Le nombre de préfixes sur Amazon Simple Storage Service (Amazon S3) auxquels on accède pour la requête Redshift Spectrum
- Vous pouvez également consulter la table `SVL_S3QUERY_SUMMARY` système pour en savoir plus sur l'efficacité de vos partitions et l'efficacité de votre requête Redshift Spectrum. Pour cela, utilisez l'instruction suivante.

```
Select *  
From svl_s3query_summary  
Where query=pg_last_query_id();
```

La requête précédente renvoie encore plus d'informations, notamment `is_partitioneds3_scanned_rows/bytes`, et des `s3_returned_rows/bytes` valeurs en plus des fichiers montrant l'efficacité de l'élagage des partitions.

Pushdown des prédicats dans Redshift Spectrum

L'utilisation du pushdown des prédicats permet d'éviter de consommer des ressources dans le cluster Amazon Redshift. Vous pouvez transférer de nombreuses opérations SQL vers la couche Redshift Spectrum. Nous vous recommandons d'en profiter dans la mesure du possible.

Gardez à l'esprit les points suivants :

- Vous pouvez évaluer complètement certains types d'opérations SQL au sein de la couche Redshift Spectrum, notamment les suivantes :
 - GROUP BY clauses
 - Conditions de comparaison et de correspondance de modèles (par exemple,) LIKE
 - Fonctions d'agrégation (par exemple COUNT, SUM, AVG, MIN, et MAX)
 - `regex_replace`, `to_upperdate_trunc`, et autres fonctions
- Vous ne pouvez pas transférer certaines opérations vers la couche Redshift Spectrum, notamment DISTINCT et ORDER BY. Effectuez cette opération ORDER BY uniquement au niveau supérieur de la requête si possible, car le tri est effectué dans le nœud principal.
- Examinez votre EXPLAIN plan de requête pour vérifier si le transfert des prédicats est efficace. Pour rechercher des portions de Redshift Spectrum dans une EXPLAIN commande, suivez les étapes suivantes :
 - S3 Seq Scan
 - S3 HashAggregate
 - S3 Query Scan
 - Scan Seq PartitionInfo
 - Partition Loop

- Utilisez le plus petit nombre de colonnes dans votre requête. Redshift Spectrum peut éliminer les colonnes à scanner si les données sont au format Parquet ou ORC.
- Utilisez largement les partitions pour le traitement parallèle et l'élimination des partitions, et maintenez la taille des fichiers à au moins 64 Mo si possible.
- Réglez le `TABLE PROPERTIES 'numRows' = 'nnn'` si vous utilisez `CREATE EXTERNAL TABLE` ou `ALTER TABLE`. Amazon Redshift n'analyse pas les tables externes pour générer des statistiques de table que l'optimiseur de requêtes utilise pour générer un plan de requêtes. Si les statistiques ne sont pas définies, Amazon Redshift part du principe que les tables externes sont les plus grandes et que les tables locales sont les plus petites.

Conseils de réglage des requêtes pour Redshift Spectrum

Nous vous recommandons de garder les points suivants à l'esprit lorsque vous ajustez vos requêtes :

- Le nombre de nœuds Redshift Spectrum que votre cluster Amazon Redshift peut engager pour une requête est lié au nombre de tranches de votre cluster.
- Le redimensionnement de votre cluster peut bénéficier aux profils de calcul locaux, aux profils de stockage et aux capacités de requête de la requête du lac de données Amazon S3.
- Le planificateur de requêtes Amazon Redshift pousse les prédicats et les agrégations vers la couche de requêtes Redshift Spectrum lorsque cela est possible.
- Lorsque de grandes quantités de données sont renvoyées depuis Amazon S3, le traitement est limité par les ressources de votre cluster.
- Comme Redshift Spectrum évolue automatiquement pour traiter les demandes volumineuses, vos performances globales s'améliorent chaque fois que vous pouvez transférer le traitement vers la couche Redshift Spectrum.

Ressources

- [Bonnes pratiques d'Amazon Redshift \(documentation Amazon Redshift\)](#)
- [Meilleures pratiques d'Amazon Redshift pour la conception de requêtes \(documentation Amazon Redshift\)](#)
- [Optimisation des performances des requêtes](#) (documentation Amazon Redshift)
- [Plan de requête](#) (documentation Amazon Redshift)
- [Améliorer les performances des requêtes Amazon Redshift Spectrum](#) (documentation Amazon Redshift)
- [Comprendre le cycle de vie des requêtes dans Amazon Redshift](#) (directives AWS prescriptives)

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
AQUA supprimé	Nous avons supprimé les informations relatives à Advanced Query Accelerator (AQUA).	14 juin 2024
Publication initiale	—	3 février 2023

AWS Glossaire des directives prescriptives

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

Nombres

7 R

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

- **Refactorisation/réarchitecture** : transférez une application et modifiez son architecture en tirant pleinement parti des fonctionnalités natives cloud pour améliorer l'agilité, les performances et la capacité de mise à l'échelle. Cela implique généralement le transfert du système d'exploitation et de la base de données. Exemple : migrez votre base de données Oracle sur site vers 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 locale vers Oracle sur une EC2 instance du 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 la synchronisation des bases de données source et cible est maintenue, mais seule la base de données source gère les transactions provenant de la connexion d'applications pendant que les données sont répliquées vers la base de données cible. La base de données cible n'accepte aucune transaction pendant la migration.

fonction d'agrégation

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

AI

Voir [intelligence artificielle](#).

AIOps

Voir les [opérations d'intelligence artificielle](#).

anonymisation

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

anti-motif

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

contrôle des applications

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

portefeuille d'applications

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

intelligence artificielle (IA)

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

opérations d'intelligence artificielle (AIOps)

Processus consistant à utiliser des techniques de machine learning pour résoudre les problèmes opérationnels, réduire les incidents opérationnels et les interventions humaines, mais aussi améliorer la qualité du service. Pour plus d'informations sur 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

Voir base de [données de gestion de configuration](#).

référentiel de code

Emplacement où le code source et d'autres ressources, comme la documentation, les exemples et les scripts, sont stockés et mis à jour par le biais de processus de contrôle de version. Les référentiels cloud courants incluent GitHub ou Bitbucket 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 is commonly described as a 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).

environnement

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

- Environnement de développement : instance d'une application en cours d'exécution à laquelle seule l'équipe principale chargée de la maintenance de l'application peut accéder. Les environnements de développement sont utilisés pour tester les modifications avant de les promouvoir dans les environnements supérieurs. Ce type d'environnement est parfois appelé environnement de test.
- Environnements inférieurs : tous les environnements de développement d'une application, tels que ceux utilisés pour les générations et les tests initiaux.
- Environnement de production : instance d'une application en cours d'exécution à laquelle les utilisateurs finaux peuvent accéder. Dans un pipeline CI/CD, l'environnement de production est le dernier environnement de déploiement.
- Environnements supérieurs : tous les environnements accessibles aux utilisateurs autres que l'équipe de développement principale. Ils peuvent inclure un environnement de production, des environnements de préproduction et des environnements pour les tests d'acceptation par les utilisateurs.

épopée

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

ERP

Voir [Planification des ressources d'entreprise](#).

analyse exploratoire des données (EDA)

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

F

tableau des faits

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

échouer rapidement

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

limite d'isolation des défauts

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

branche de fonctionnalités

Voir [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 clics 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 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

laC

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.

Ilo 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

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 au fur et à mesure de son fonctionnement. Pour plus d'informations, voir [Création de mécanismes](#) dans le cadre AWS Well-Architected.

compte membre

Tous, à l'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 téléométrique 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 avec le service de migration AWS 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

Voir [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, évaluer, prévenir ou réduire l'ampleur des incidents et des défaillances possibles. Pour plus d'informations, voir [Operational Readiness Reviews \(ORR\)](#) dans le AWS Well-Architected Framework.

technologie opérationnelle (OT)

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

intégration des opérations (OI)

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

journal de suivi d'organisation

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

gestion du changement organisationnel (OCM)

Cadre pour gérer les transformations métier majeures et perturbatrices du point de vue des personnes, de la culture et du leadership. L'OCM aide les organisations à se préparer et à effectuer la transition vers de nouveaux systèmes et de nouvelles politiques en accélérant

l'adoption des changements, en abordant les problèmes de transition et en favorisant des changements culturels et organisationnels. Dans la stratégie de AWS migration, ce cadre est appelé accélération du personnel, en raison de la rapidité du changement requise dans les projets d'adoption du cloud. Pour plus d'informations, veuillez consulter le [guide OCM](#).

contrôle d'accès d'origine (OAC)

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

identité d'accès d'origine (OAI)

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

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.

politique

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

persistance polyglotte

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

évaluation du portefeuille

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

predicate

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

prédicat pushdown

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

contrôle préventif

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

principal

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

confidentialité dès la conception

Une approche 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 \(RACI\)](#).

CHIFFON

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 la section [AWS Cloud Résilience](#).

politique basée sur les ressources

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

matrice responsable, redevable, consulté et informé (RACI)

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

contrôle réactif

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

retain

Voir [7 Rs](#).

se retirer

Voir [7 Rs](#).

Génération augmentée de récupération (RAG)

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

rotation

Processus de mise à jour périodique d'un [secret](#) pour empêcher un attaquant d'accéder aux informations d'identification.

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

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

RPO

Voir l'[objectif du point de récupération](#).

RTO

Voir l'[objectif relatif au 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 AWS de sécurité. Parmi les actions de réponse automatique, citons la modification d'un groupe de sécurité VPC, l'application de correctifs à une EC2 instance Amazon 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 définissent 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, consultez la section [Approche progressive de la modernisation des applications dans le AWS Cloud](#)

SPOF

Voir [point de défaillance unique](#).

schéma en étoile

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

modèle de figuier étrangleur

Approche de modernisation des systèmes monolithiques en réécrivant et en remplaçant progressivement les fonctionnalités du système jusqu'à ce que le système hérité puisse être mis hors service. Ce modèle utilise l'analogie d'un figuier de vigne qui se développe dans un arbre existant et qui finit par supplanter son hôte. Le schéma a été [présenté par Martin Fowler](#) comme un moyen de gérer les risques lors de la réécriture de systèmes monolithiques. Pour obtenir un exemple d'application de ce modèle, veuillez consulter [Modernizing legacy Microsoft ASP.NET \(ASMX\) web services incrementally by using containers and Amazon API Gateway](#).

sous-réseau

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

contrôle de supervision et acquisition de données (SCADA)

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

chiffrement symétrique

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

tests synthétiques

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

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

balises

Des paires clé-valeur qui agissent comme des métadonnées pour organiser vos AWS ressources. Les balises peuvent vous aider à gérer, identifier, organiser, rechercher et filtrer des ressources. Pour plus d'informations, veuillez consulter la rubrique [Balisage de vos AWS ressources](#).

variable cible

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

liste de tâches

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

environnement de test

Voir [environnement](#).

entraînement

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

passerelle de transit

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.