AWS Well-Architected Framework

Pilier de durabilité

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

Pilier de durabilité: AWS Well-Architected Framework

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

Résumé et introduction i
Introduction 1
Durabilité du cloud 3
Le modèle de responsabilité partagée 4
Durabilité dans le cloud 5
Durabilité dans le cloud 5
Durabilité via le cloud 5
Principes de conception pour la durabilité dans le cloud 6
Processus d'amélioration 8
Exemple de scénario 9
Identifier les objectifs d'amélioration 9
Ressources 10
Évaluer des améliorations spécifiques 10
Métriques de proxy 10
Métriques commerciales 11
Indicateurs de performance clés 11
Estimer l'amélioration 12
Évaluer les améliorations 13
Prioriser et planifier les améliorations 14
Tester et valider les améliorations 15
Déployer les modifications en production 16
Mesurer les résultats et reproduire les améliorations fructueuses 17
Durabilité en tant qu'exigence non fonctionnelle 19
Bonnes pratiques pour la durabilité dans le cloud 21
Sélection d'une région 21
SUS01-BP01 Choisir une région en fonction des exigences et des objectifs de durabilité de l'entreprise 21
Alignement sur la demande 23
SUS02-BP01 Mettre à l'échelle l'infrastructure de la charge de travail de façon dynamique ... 24
SUS02-BP02 S'aligner sur les objectifs SLAs de durabilité 28
SUS02-BP03 Arrêter la création et la maintenance des actifs inutilisés 30
SUS02-BP04 Optimiser le placement géographique des charges de travail en fonction de leurs exigences en matière de réseau 32

SUS02-BP05 Optimiser les ressources des membres de l'équipe pour les activités réalisées ... 35
SUS02-BP06 Mettre en œuvre la mise en mémoire tampon ou la régulation pour aplatir la courbe de demande ... 37
Logiciels et architecture ... 39
SUS03-BP01 Optimiser le logiciel et l'architecture pour les tâches asynchrones et planifiées ... 40
SUS03-BP02 Supprimer ou refactoriser les composants de charges de travail faiblement utilisés ou inutilisés ... 43
SUS03-BP03 Optimiser les sections de votre code qui consomment le plus de temps ou de ressources ... 46
SUS03-BP04 Optimiser l'impact sur les appareils et équipements ... 48
SUS03-BP05 Utiliser des modèles logiciels et des architectures qui soutiennent au mieux l’accès aux données et les modèles de stockage. ... 50
Gestion des données ... 53
SUS04-BP01 Mettre en œuvre une politique de classification des données ... 54
SUS04-BP02 Utiliser des technologies qui prennent en charge les modèles d'accès et de stockage des données ... 55
SUS04-BP03 Utiliser des politiques pour gérer le cycle de vie de vos ensembles de données ... 61
SUS04-BP04 Utiliser l'élasticité et l'automatisation pour étendre le stockage par blocs ou le système de fichiers ... 64
SUS04-BP05 Supprimer les données inutiles ou redondantes ... 65
SUS04-BP06 Utiliser des systèmes de fichiers partagés ou le stockage pour accéder aux données courantes ... 68
SUS04-BP07 Réduire le mouvement des données entre les réseaux ... 70
SUS04-BP08 Sauvegarder des données uniquement lorsqu’elles sont difficiles à recréer ... 73
Matériel et services ... 75
SUS05-BP01 Utilisez le minimum de matériel pour répondre à vos besoins ... 75
SUS05-BP02 Utiliser des types d’instance ayant le moins d’impact ... 77
SUS05-BP03 Utiliser des services gérés ... 81
SUS05-BP04 Optimiser votre utilisation des accélérateurs de calcul matériels ... 83
Processus et culture ... 85
SUS06-BP01 Communiquer et répercuter en cascade vos objectifs de durabilité ... 85
SUS06-BP02 Adopter des méthodes qui peuvent rapidement introduire des améliorations en matière de durabilité ... 89

SUS06-BP03 Maintenir à jour votre charge de travail 90
SUS06-BP04 Augmenter l'utilisation des environnements de génération 93
SUS06-BP05 Utiliser des tests Device Farms gérés 95
Conclusion 97
Collaborateurs 98
Suggestions de lecture 99
Révisions du document 100
Avis 102
AWS Glossaire 103

Pilier Durabilité – AWS Well-Architected Framework

Date de publication : 6 novembre 2024 (Révisions du document)

Le présent livre blanc porte sur le pilier Durabilité du cadre Amazon Web Services (AWS) Well-Architected Framework. Il présente des principes de conception, des conseils opérationnels, de bonnes pratiques, des compromis potentiels et des plans d'amélioration que vous pouvez utiliser pour atteindre les objectifs de durabilité pour vos charges de travail AWS.

Introduction

AWS Well-Architected Framework vous permet de mesurer le pour et le contre des options qui se présentent lors de la création de charges de travail sur AWS. Vous pouvez ainsi apprendre les bonnes pratiques architecturales en matière de conception et d'exploitation de charges de travail sécurisées, fiables, efficaces, économiques et durables dans le AWS Cloud. Well-Architected Framework vous permet également d'évaluer systématiquement vos architectures par rapport aux bonnes pratiques et d'identifier les domaines à améliorer. Disposer de charges de travail bien architecturées augmente considérablement votre capacité à atteindre vos résultats commerciaux.

Well-Architected Framework repose sur six piliers :

- Excellence opérationnelle
- Sécurité
- Fiabilité
- Efficacité des performances
- Optimisation des coûts
- Durabilité

Le présent document porte principalement sur le pilier Durabilité, et dans la même lignée, sur la durabilité environnementale. Il est destiné à ceux et celles qui occupent des postes technologiques, tels que les directeurs techniques (CTO), les architectes, les développeurs et les membres de l'équipe des opérations.

Après avoir lu ce document, vous comprendrez les recommandations et les stratégies AWS actuelles à utiliser lors de la conception d'architectures cloud axées sur la durabilité. En adoptant les pratiques

décrites dans le présent document, vous créerez des architectures qui optimisent l’efficacité tout en limitant le gaspillage.

Durabilité du cloud

La discipline de la durabilité traite des répercussions environnementales, économiques et sociétales à long terme de vos activités commerciales. La Commission mondiale sur l'environnement et le développement des Nations Unies définit le développement durable comme « un développement qui répond aux besoins du présent sans compromettre la capacité des générations futures à satisfaire les leurs ». Votre entreprise ou organisation peut avoir des impacts environnementaux négatifs, tels que des émissions carbone directes ou indirectes, des déchets non recyclables et des dommages causés aux ressources partagées comme l'eau potable.

Lorsque des charges de travail sont créées dans le cloud, la pratique de la durabilité consiste à comprendre les impacts des services utilisés, mesurer les impacts tout au long du cycle de vie de la charge de travail dans son ensemble et appliquer des principes de conception et de bonnes pratiques afin de réduire ces impacts. Ce document se concentre sur les impacts environnementaux, notamment la consommation et l'efficacité énergétiques, qui sont des leviers importants permettant aux architectes de recueillir des informations sur les actions directes afin d'utiliser moins de ressources.

Lorsque vous vous concentrez sur les impacts environnementaux, vous devez comprendre comment ils sont généralement pris en compte et les implications sur la comptabilisation des émissions de votre organisation. Le Greenhouse Gas Protocol organise les émissions de CO2 dans les champs d'application suivants, avec des exemples d'émissions pertinents dans chaque champ d'application pour un fournisseur de cloud comme AWS :

- Champs d'application 1 : toutes les émissions directes provenant des activités d'une organisation ou qui sont sous son contrôle. Par exemple, la combustion de carburant par les générateurs de secours du centre de données.
- Champs d'application 2 : émissions indirectes provenant de l'électricité achetée et utilisée pour alimenter les centres de données et d'autres installations. Par exemple, les émissions provenant de la production commerciale d'électricité.
- Champs d'application 3 : toutes les autres émissions indirectes des activités d'une organisation provenant de sources qu'elle ne contrôle pas. Les exemples AWS incluent les émissions liées à la construction des centres de données, ainsi qu'à la fabrication et au transport du matériel informatique déployé dans les centres de données.

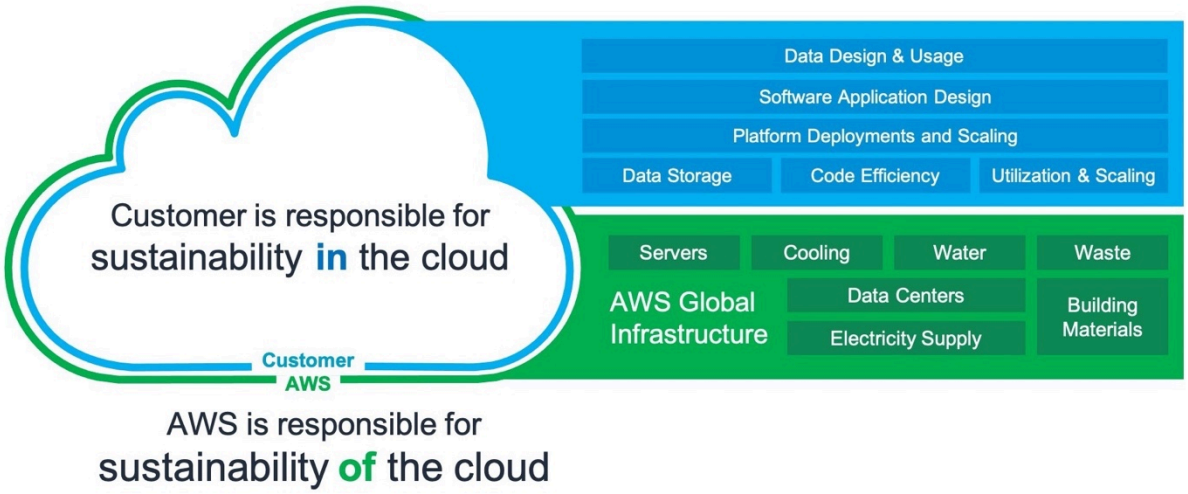
Du point de vue d'un client AWS, les émissions de vos charges de travail exécutées sur AWS sont comptabilisées comme des émissions indirectes et font partie des émissions du champ d'application 3. Chaque charge de travail déployée génère une fraction des émissions AWS totales issues de chacun des périmètres précédents. Le montant réel varie en fonction de la charge de travail et dépend de plusieurs facteurs, notamment les services AWS utilisés, l'énergie consommée par ces services, l'intensité carbone des réseaux électriques desservant les centres de données AWS où ils fonctionnent et les achats AWS d'énergie renouvelable.

Ce document décrit d'abord un modèle de responsabilité partagée pour la durabilité environnementale, puis fournit les bonnes pratiques architecturales à suivre afin que vous puissiez limiter l'impact de vos charges de travail en réduisant les ressources totales nécessaires à leur exécution dans les centres de données AWS.

Le modèle de responsabilité partagée

La durabilité environnementale est une responsabilité partagée entre les clients et AWS.

- AWS est responsable de l'optimisation et la durabilité du cloud, en fournissant une infrastructure efficace et partagée, l'intendance de l'eau et l'approvisionnement en énergie renouvelable.
- Les clients sont responsables de la durabilité dans le cloud, en optimisant les charges de travail et l'utilisation des ressources, et en limitant les ressources totales devant être déployées pour vos charges de travail.



Modèle de responsabilité partagée

Durabilité dans le cloud

Les fournisseurs de cloud ont une empreinte carbone plus faible et sont plus économes en énergie que les installations sur site classiques, car ils investissent dans une technologie d'alimentation et de refroidissement efficace, exploitent des populations de serveurs économes en énergie et atteignent des taux d'utilisation des serveurs élevés. Les charges de travail cloud réduisent l'impact en tirant parti des ressources partagées, telles que la mise en réseau, l'alimentation, le refroidissement et les installations physiques. Pour une meilleure durabilité, vous pouvez migrer vos charges de travail dans le cloud vers des technologies plus efficaces au fur et à mesure qu'elles deviennent disponibles, et utiliser des services basés sur le cloud pour transformer vos charges de travail.

Ressources

- L'opportunité de réduire les émissions de carbone en passant à Amazon Web Services
- AWS active des solutions de durabilité

Durabilité dans le cloud

La durabilité dans le cloud est un effort continu axé principalement sur la réduction d'énergie et l'efficacité de tous les composants d'une charge de travail en tirant le meilleur parti possible des ressources allouées et en minimisant les ressources totales requises. Cet effort peut inclure la sélection initiale d'un langage de programmation efficace, l'adoption d'algorithmes modernes, l'utilisation de techniques de stockage de données performantes, le déploiement sur une infrastructure de calcul correctement dimensionnée et efficace, et la réduction des besoins en matériel de grande puissance pour les utilisateurs finaux.

Durabilité via le cloud

En plus de limiter l'impact des charges de travail que vous avez déployées, vous pouvez utiliser le AWS Cloud pour exécuter des charges de travail conçues pour vous aider à relever vos défis de durabilité à plus grande échelle. Parmi ces défis, citons par exemple la réduction des émissions de carbone, la réduction de la consommation d'énergie, le recyclage de l'eau ou la réduction des déchets dans d'autres domaines de votre entreprise ou organisation.

La durabilité à travers le cloud est lorsque vous utilisez la technologie AWS pour relever un défi de durabilité plus large. Par exemple, vous pouvez utiliser un service de machine learning pour détecter les comportements anormaux dans les machines industrielles. Grâce à ces données de

détection, vous pouvez effectuer une maintenance préventive pour réduire le risque d'incidents environnementaux causés par des pannes d'équipement inattendues et vous assurer que les machines continuent de fonctionner avec une efficacité maximale.

Principes de conception pour la durabilité dans le cloud

Appliquez ces principes de conception lors de la conception de vos charges de travail dans le cloud afin de maximiser la durabilité et de minimiser l'impact.

- Comprendre votre impact : mesurez l'impact de votre charge de travail sur le cloud et modélisez l'impact futur de votre charge de travail. Inclure toutes les sources d'impact, y compris les impacts résultant de l'utilisation que les clients font de vos produits, ainsi que les impacts résultant de leur déclassement et de leur mise hors service éventuels. Comparez le rendement productif à l'impact total de vos charges de travail sur le cloud en évaluant les ressources et les émissions nécessaires par unité de travail. Utilisez ces données pour établir des indicateurs clés de performance (KPI), évaluer des moyens d'améliorer la productivité tout en réduisant l'impact et estimer l'impact des changements proposés au fil du temps.
- Définir des objectifs de durabilité : pour chaque charge de travail dans le cloud, établissez des objectifs de durabilité à long terme, tels que la réduction des ressources de calcul et de stockage nécessaires par transaction. Modélisez le retour sur investissement des améliorations durables pour les charges de travail existantes et donnez aux propriétaires les ressources dont ils ont besoin pour investir dans leurs objectifs de durabilité. Planifiez en vue d'une croissance et concevez l'architecture de vos charges de travail afin que la croissance entraîne une intensité de l'impact moindre mesurée par rapport à une unité appropriée, par utilisateur ou par transaction par exemple. Les objectifs vous aident à soutenir les cibles de durabilité plus larges de votre entreprise ou organisation, identifier les régressions et privilégier les zones pouvant être améliorées.
- Optimiser l'utilisation : dimensionnez correctement les charges de travail et intégrez une conception efficace pour assurer une forte utilisation et optimiser l'efficacité énergétique du matériel sous-jacent. Deux hôtes s'exécutant à 30 % de leur utilisation sont moins efficaces qu'un seul hôte s'exécutant à 60 % du fait de la consommation énergétique de base par hôte. Éliminez ou minimisez également les ressources, le traitement et le stockage inactifs afin de réduire l'énergie totale nécessaire pour alimenter votre charge de travail.
- Anticiper et adopter de nouvelles offres matérielles et logicielles plus efficaces : soutenez les améliorations en amont de vos partenaires et fournisseurs pour permettre de réduire l'impact de vos charges de travail sur le cloud. Contrôlez et évaluez de façon continue des offres matérielles

et logicielles neuves et plus efficaces. Concevez de manière flexible afin de permettre l'adoption rapide de nouvelles technologies efficaces.

- Utiliser des services gérés : le partage des services auprès d'une clientèle importante permet de maximiser l'utilisation des ressources, ce qui réduit la quantité d'infrastructure nécessaire pour soutenir les charges de travail dans le cloud. Par exemple, les clients peuvent partager l'impact de composants courants du centre de données tels que l'énergie et les réseaux en migrant des charges de travail vers le AWS Cloud et en adoptant des services gérés comme AWS Fargate pour les conteneurs sans serveur. AWS y opère à grande échelle et est responsable de son fonctionnement efficace. Utilisez des services gérés qui peuvent permettre de réduire votre impact, tels que le déplacement automatique des données à accès peu fréquent vers un stockage froid avec les configurations du cycle de vie Amazon S3 ou Amazon EC2 Auto Scaling pour ajuster la capacité à répondre à la demande.
- Réduisez l'impact en aval de vos charges de travail dans le cloud : réduisez la quantité d'énergie ou de ressources nécessaires pour utiliser vos services. Réduisez ou supprimez la nécessité pour les clients d'avoir à mettre à niveau leurs appareils pour pouvoir utiliser vos services. Réalisez des tests à l'aide de Device Farms pour comprendre l'impact attendu et auprès de clients pour comprendre l'impact réel que représente l'utilisation de vos services.

Processus d'amélioration

Le processus d'amélioration de l'architecture implique de comprendre ce dont vous disposez et ce que vous pouvez faire pour vous améliorer : choisir des cibles à améliorer, tester les améliorations, adopter celles portant leurs fruits, mesurer votre réussite et partager ce que vous avez appris afin de le reproduire ailleurs, puis répéter le cycle.

Les objectifs de vos améliorations peuvent être les suivants :

- Éliminer le gaspillage, la faible utilisation et les ressources inutilisées ou inactives.
- Maximiser la valeur des ressources que vous consommez.

> **Note**
>
> Utiliser toutes les ressources que vous allouez et effectuer le même travail avec le minimum de ressources possible.

Aux premiers stades de l'optimisation, se concentrer d'abord sur les domaines engendrant du gaspillage ou une faible utilisation, puis passer à des optimisations plus ciblées correspondant à votre charge de travail spécifique.

Contrôler les changements dans la consommation de vos ressources au fil du temps. Identifier où les changements cumulés entraînent des augmentations inefficaces ou importantes dans la consommation des ressources. Déterminer le besoin d'apporter des améliorations afin de répondre aux changements dans la consommation et intégrer les améliorations prioritaires.

Les étapes suivantes sont conçues pour être un processus itératif qui évalue, hiérarchise, teste et déploie des améliorations axées sur la durabilité pour les charges de travail cloud.

1. Identifier les objectifs d'amélioration : examinez vos charges de travail par rapport aux bonnes pratiques de durabilité identifiées dans ce document et identifiez les objectifs d'amélioration.
2. Évaluer les améliorations spécifiques : évaluez les modifications spécifiques en vue d'améliorations potentielles, pour les coûts projetés et les risques métiers.
3. Établir des priorités et planifier les améliorations : priorisez les modifications qui apporteront les améliorations les plus importantes à moindre coût et au moindre risque, et définissez une stratégie de test et de mise en œuvre.

4. Tester et valider les améliorations : mettez en œuvre les modifications dans les environnements de test pour valider leur potentiel d'amélioration.
5. Déployer les modifications apportées à la production : implémentez les modifications dans les environnements de production.
6. Evaluez les résultats et reproduisez les succès : recherchez des opportunités de reproduire les améliorations fructueuses sur toutes les charges de travail et annulez les modifications dont les résultats ne conviennent pas.

Exemple de scénario

L'exemple de scénario suivant est référencé plus loin dans ce document pour illustrer chaque étape du processus d'amélioration.

Votre entreprise a une charge de travail qui effectue des manipulations d'images complexes sur des EC2 instances Amazon et stocke les fichiers modifiés et originaux pour que les utilisateurs puissent y accéder. Les activités de traitement sont CPU intensives et les fichiers de sortie sont extrêmement volumineux.

Identifier les objectifs d'amélioration

Identifiez les bonnes pratiques qui peuvent vous aider à atteindre vos objectifs de durabilité. Vous trouverez des descriptions détaillées de ces bonnes pratiques et des recommandations d'amélioration plus loin dans ce document.

Examinez vos charges de travail et les ressources utilisées. Identifier les points chauds tels que les grands déploiements et les ressources fréquemment utilisées. Évaluez ces points chauds pour déterminer les possibilités d'améliorer l'utilisation efficace de vos ressources et de réduire le total des ressources nécessaires pour atteindre vos résultats commerciaux.

Examinez votre charge de travail par rapport aux bonnes pratiques et identifiez les candidats à l'amélioration.

En appliquant cette étape au Exemple de scénario, vous identifierez les bonnes pratiques suivantes comme objectifs d'amélioration probables :

- Utiliser la quantité minimale de matériel pour répondre à vos besoins
- Utilisez les technologies qui prennent le mieux en charge vos modèles d'accès aux données et de stockage

Ressources

- Optimisation de votre AWS infrastructure pour la durabilité, partie I : calcul
- Optimisation de votre AWS infrastructure pour la durabilité, partie II : stockage
- Optimisation de votre AWS infrastructure pour la durabilité, partie III : mise en réseau

Évaluer des améliorations spécifiques

Identifiez les ressources allouées par votre charge de travail pour mener à bien une unité de travail. Évaluez les améliorations potentielles et estimez leur impact potentiel, le coût de mise en œuvre et les risques associés.

Pour mesurer les améliorations au fil du temps, déterminez d'abord ce que vous avez provisionné AWS et comment ces ressources sont consommées.

Commencez par un aperçu complet de votre AWS utilisation, puis utilisez les rapports sur les AWS coûts et l'utilisation pour identifier les points chauds. Utilisez cet exemple de code AWS pour vous aider à consulter et à analyser votre rapport à l'aide d'Amazon Athena.

Métriques de proxy

Lorsque vous évaluez des modifications spécifiques, vous devez également évaluer les métriques qui quantifient le mieux l'effet de cette modification sur la ressource associée. Ces métriques sont appelées métriques proxy. Sélectionnez les métriques de proxy qui reflètent le mieux le type d'amélioration que vous évaluez et les ressources ciblées par l'amélioration. Ces métriques peuvent évoluer avec le temps.

Les ressources allouées pour soutenir votre charge de travail incluent les ressources de calcul, de stockage et de réseau. Évaluez les ressources allouées à l'aide de vos métriques de proxy pour déterminer comment ces ressources sont consommées.

Utilisez vos métriques de proxy afin de mesurer les ressources allouées pour atteindre les résultats commerciaux.

Ressource	Exemple de métriques de proxy	Objectifs d'amélioration
Calcul	v CPU minutes	Optimiser l'utilisation des ressources allouées
Stockage	Go alloués	Réduire les ressources totales allouées
Réseau	Go transférés ou paquets transférés	Réduire la distance transférée et le total de Go transférés

Métriques commerciales

Sélectionnez des métriques commerciales pour quantifier la réalisation des résultats commerciaux. Les indicateurs de votre activité doivent refléter la valeur fournie par votre charge de travail, par exemple le nombre d'utilisateurs actifs simultanés, les API appels traités ou le nombre de transactions effectuées. Elles peuvent évoluer avec le temps. Soyez prudent lors de l'évaluation des métriques commerciales basées sur les coûts, car l'incohérence de la valeur des transactions invalide les comparaisons.

Indicateurs de performance clés

À l'aide de la formule suivante, divisez les ressources allouées par les résultats commerciaux obtenus pour déterminer les ressources allouées par unité de travail.

$$Resources\ provisioned\ per\ unit\ of\ work = \frac{Proxy\ metric\ for\ provisioned\ resource}{Business\ metric\ for\ outcome}$$

KPIformule

Utilisez vos ressources par unité de travail comme les vôtresKPIs. Définissez des points de référence basés sur les ressources allouées comme base de comparaison.

Ressource	Exemple KPIs	Objectifs d'amélioration
Calcul	v CPU minutes par transaction	Optimiser l'utilisation des ressources allouées
Stockage	Go par transaction	Réduire les ressources totales allouées
Réseau	Go transférés par transacti on ou paquets transférés par transaction	Réduire la distance transférée et le total de Go transférés

Estimer l'amélioration

Estimez l'amélioration en tant que réduction quantitative des ressources allouées (comme indiqué par vos métriques indirectes) et en pourcentage de variation par rapport aux ressources de base allouées par unité de travail.

Ressource	Exemple KPIs	Objectifs d'amélioration
Calcul	% de réduction du nombre de vCPUs minutes par transacti on	Optimiser l'utilisation
Stockage	% de réduction des Go par transaction	Réduire les ressources totales allouées
Réseau	% de réduction des Go transférés par transaction ou des paquets transférés par transaction	Réduire la distance transférée et le total de Go transférés

Évaluer les améliorations

Évaluez les améliorations potentielles par rapport au bénéfice net attendu. Estimez le temps, le coût et le niveau d'effort de mise en œuvre et de maintenance, ainsi que les risques commerciaux tels que les impacts imprévus.

Les améliorations ciblées impliquent souvent des compromis entre les types de ressources consommées. Par exemple, pour réduire la consommation de calcul, vous pouvez stocker un résultat, ou pour limiter les données transférées, vous pouvez traiter les données avant d'envoyer le résultat à un client. Ces compromis sont abordés plus en détail ultérieurement.

Incluez les exigences non fonctionnelles lors de l'évaluation des risques pour votre charge de travail, y compris la sécurité, la fiabilité, l'efficacité des performances, l'optimisation des coûts et l'impact des améliorations sur votre capacité à exploiter votre charge de travail.

En appliquant cette étape au Exemple de scénario, vous évaluerez les améliorations cibles avec les résultats suivant :

Bonne pratique	Amélioration ciblée	Potential	Coût	Risque
Utiliser la quantité minimale de matériel pour répondre à vos besoins	Mettre en œuvre la mise à l'échelle prédictive pour réduire les périodes de faible utilisation	Moyen	Faible	Faible
Utilisez les technologies qui prennent le mieux en charge vos modèles d'accès aux données et de stockage	Mettre en œuvre des mécanismes de compression plus efficaces pour réduire le stockage total et le temps nécessaire pour y parvenir	Élevé	Faible	Faible

La mise en œuvre de la mise à l'échelle prédictive réduit les v CPU heures consommées par les instances sous-utilisées ou inutilisées, offrant ainsi des avantages modérés par rapport aux mécanismes de dimensionnement existants, avec une réduction estimée à 11 % des ressources consommées. Les coûts impliqués sont faibles et incluent la configuration des ressources du cloud et le fonctionnement du dimensionnement prédictif pour Amazon EC2 Auto Scaling. Le risque réside dans la limitation des performances lorsque la montée en puissance est effectuée de manière réactive en réponse à une demande dépassant les prévisions.

La mise en œuvre d'une compression plus efficace peut avoir un impact significatif avec des réductions importantes de la taille des fichiers sur toutes les images originales et les images manipulées, avec une réduction estimée de 25 % des besoins de stockage en production. La mise en œuvre du nouvel algorithme est une solution de substitution impliquant peu d'efforts et peu de risques.

Prioriser et planifier les améliorations

Hiérarchisez les améliorations identifiées en fonction de l'impact anticipé le plus important avec les coûts les plus bas et un risque acceptable.

Choisissez sur quelles améliorations vous concentrer initialement et incluez -les dans votre feuille de route de planification et de développement des ressources.

Appliquer cette étape au Exemple de scénario, vous hiérarchiserez les améliorations cibles comme suit :

Priorité	Amélioration	Potential	Coût	Risque
1	Mettre en œuvre des mécanismes de compression plus efficaces	Élevé	Faible	Faible
2	Utiliser la mise à l'échelle prédictive	Moyen	Faible	Faible

Le potentiel élevé, le faible coût et le risque de mise à jour de la compression des fichiers en font une cible de grande valeur pour votre entreprise et une priorité par rapport à la mise en œuvre de la mise

à l'échelle prédictive. Vous déterminez que la mise en œuvre de la mise à l'échelle prédictive avec son impact potentiel moyen, son faible coût et son faible risque doit être l'amélioration prioritaire une fois la compression des fichiers terminée.

Vous affectez un membre de l'équipe pour mettre en œuvre une compression de fichiers améliorée et ajouter la mise à l'échelle prédictive à votre backlog.

Tester et valider les améliorations

Réalisez de petits tests avec un investissement minimal pour limiter le risque que présenterait une initiative mise à grande échelle.

Implémentez une version représentative de votre charge de travail dans votre environnement de test pour limiter les coûts et les risques liés aux tests et à la validation. Effectuez un ensemble prédéfini de transactions de test, mesurez les ressources allouées et déterminez les ressources utilisées par unité de travail pour établir un point de référence pour les tests.

Mettez en œuvre votre objectif d'amélioration dans l'environnement de test et répétez le test en utilisant la même méthodologie dans les mêmes conditions. Mesurez ensuite les ressources allouées et les ressources utilisées par unité de travail lorsque votre amélioration est intégrée.

Calculez la variation en pourcentage par rapport au point de référence (ressources allouées par unité de travail) et déterminez la réduction quantitative attendue des ressources allouées dans votre environnement de production. Comparez ces valeurs aux valeurs anticipées. Déterminez si le résultat est un niveau d'amélioration acceptable. Évaluez si les compromis liés aux ressources supplémentaires consommées rendent le bénéfice net de l'amélioration inacceptable.

Déterminez si l'amélioration est un succès et si des ressources doivent être investies dans la mise en œuvre de ce changement en production. Si le changement est évalué comme infructueux à ce stade, redirigez vos ressources pour tester et valider votre prochaine cible et poursuivez le cycle d'amélioration.

% de réduction des ressources allouées par unité de travail	Réduction quantitative des ressources allouées	Action
Conforme aux attentes	Conforme aux attentes	Procéder à l'amélioration
Non conforme aux attentes	Conforme aux attentes	Procéder à l'amélioration

% de réduction des ressources allouées par unité de travail	Réduction quantitative des ressources allouées	Action
Conforme aux attentes	Non conforme aux attentes	Poursuivre une autre amélioration
Non conforme aux attentes	Non conforme aux attentes	Poursuivre une autre amélioration

En appliquant cette étape au Exemple de scénario, vous effectuez des tests pour valider le succès.

Après avoir effectué les tests sur l'algorithme de compression amélioré, le pourcentage de réduction des ressources allouées par unité de travail (le stockage requis à la fois pour l'image d'origine et l'image modifiée) est conforme aux attentes avec une réduction moyenne de 30 % du stockage alloué et une augmentation négligeable de la charge de calcul.

Vous déterminez que les ressources de calcul supplémentaires requises pour appliquer l'algorithme de compression amélioré aux fichiers existants en production sont insignifiantes par rapport à la réduction de stockage obtenue. Vous avez confirmé le succès de la réduction quantitative des ressources requises (TBsde stockage), et l'amélioration est approuvée pour le déploiement en production.

Déployer les modifications en production

Mettez en œuvre les améliorations testées, validées et approuvées en production. Effectuez la mise en œuvre à l'aide de déploiements limités, confirmez la fonctionnalité de votre charge de travail, testez la réduction réelle des ressources allouées et des ressources consommées par unité de travail dans le cadre du déploiement limité, puis identifiez les conséquences imprévues potentielles du changement. Procédez aux déploiements complets après des tests réussis.

Annulez les modifications si les tests échouent ou si vous identifiez des conséquences imprévues inacceptables pour cette modification.

En appliquant cette étape au Exemple de scénario, procédez comme suit :

Vous implémentez les modifications en production à l'aide d'un déploiement limité via une méthodologie de déploiement bleu-vert. Les tests de fonctionnalité sur les instances nouvellement déployées aboutissent. Vous constatez une réduction moyenne de 26 % du stockage alloué pour les

fichiers image originaux et les fichiers manipulés. Vous ne voyez aucune preuve d'augmentation de la charge de calcul lors de la compression de nouveaux fichiers.

Vous remarquez une diminution imprévue du temps écoulé pour compresser les fichiers image, et vous attribuez ce comportement au code hautement optimisé pour le nouvel algorithme de compression.

Vous procédez au déploiement complet de la nouvelle version.

Mesurer les résultats et reproduire les améliorations fructueuses

Mesurez les résultats et reproduisez les améliorations fructueuses de la manière suivante :

- Mesurez l'amélioration initiale des ressources allouées par unité de travail et la diminution quantitative des ressources allouées.
- Comparez les estimations initiales et les résultats des tests à vos mesures de production. Identifiez les facteurs qui pourraient avoir contribué aux différences et mettez à jour vos méthodologies d'estimation et de test, le cas échéant.
- Déterminez le succès, ainsi que le degré de succès, et partagez les résultats avec les parties prenantes.
- Si vous avez dû annuler des modifications en raison d'échecs de tests ou de conséquences négatives imprévues, identifiez les facteurs contributifs. Itérez là où cela est possible ou évaluez de nouvelles approches pour atteindre les objectifs liés à la modification.
- Exploitez les enseignements que vous avez tirés, définissez des normes et appliquez les améliorations fructueuses à d'autres systèmes qui pourraient en bénéficier de la même manière. Capturez et partagez votre méthodologie, les artefacts associés et les bénéfices nets entre les équipes et les organisations afin que d'autres puissent adopter votre norme et reproduire votre succès.
- Surveillez les ressources allouées par unité de travail et suivez les changements et l'impact total au fil du temps. Les modifications apportées à votre charge de travail, ou la façon dont vos clients consomment votre charge de travail, peuvent avoir un impact sur l'efficacité de votre amélioration. Réévaluez les possibilités d'amélioration si vous remarquez une diminution significative à court terme de l'efficacité de votre amélioration ou une réduction cumulée de l'efficacité au fil du temps.
- Quantifiez le bénéfice net de votre amélioration au fil du temps (y compris les bénéfices reçus par les autres équipes qui ont appliqué votre amélioration, le cas échéant) pour prouver le retour sur investissement de vos activités d'amélioration.

En appliquant cette étape au Exemple de scénario, vous mesurez les résultats suivants.

Votre charge de travail affiche une amélioration initiale de 23 % de réduction des besoins en stockage après le déploiement et l'application du nouvel algorithme de compression aux fichiers image existants.

La valeur mesurée est largement en accord avec les estimations initiales (25 %), et la différence significative par rapport aux tests (30 %) est déterminée comme étant le résultat du fait que les fichiers image utilisés dans les tests ne sont pas représentatifs des fichiers image présents en production. Vous modifiez le jeu d'images de test pour mieux refléter les images en production.

L'amélioration est couronnée de succès. La réduction totale du stockage alloué est inférieure de 2 % aux 25 % estimés, mais 23 % représente toujours une amélioration considérable de l'impact sur la durabilité et s'accompagne d'une réduction des coûts équivalente.

Les seules conséquences imprévues du changement sont la réduction bénéfique du temps écoulé pour effectuer la compression et une réduction équivalente v consommée. CPU Ces améliorations sont attribuées au code hautement optimisé.

Vous définissez un projet open source interne dans lequel vous partagez votre code, les artefacts associés, des conseils sur la façon d'implémenter la modification et les résultats de votre implémentation. Le projet open source interne permet à vos équipes d'adopter facilement le code pour tous leurs cas d'utilisation de stockage de fichiers persistants. Vos équipes adoptent l'amélioration comme standard. Les avantages secondaires du projet open source interne sont que tous ceux qui adoptent la solution bénéficient des améliorations apportées à la solution, et n'importe qui peut apporter des améliorations au projet.

Vous publiez votre amélioration fructueuse et partagez le projet open source au sein de votre organisation. Chaque équipe qui adopte la solution bénéficie des mêmes avantages avec un investissement minimum, et contribue au bénéfice net reçu de votre investissement. Vous publiez ces données comme un témoignage de réussite continu.

Vous continuez à surveiller l'impact de l'amélioration au fil du temps et apporterez des modifications au projet open source interne si nécessaire.

Durabilité en tant qu’exigence non fonctionnelle

L’ajout de la durabilité à votre liste d’exigences métiers peut générer des solutions plus rentables. Le fait de s'efforcer de tirer le meilleur parti des ressources que vous utilisez et d'en utiliser moins se traduit directement par des économies, AWS car vous ne payez que pour ce que vous utilisez.

La réalisation des objectifs de durabilité peut ne pas nécessiter de compromis équivalents dans une ou plusieurs autres métriques traditionnelles telles que le temps de fonctionnement, la disponibilité ou le temps de réponse. Vous pouvez réaliser des gains significatifs en matière de durabilité sans impact mesurable sur les résultats des services. Lorsque des compromis mineurs sont nécessaires, les améliorations de la durabilité obtenues grâce à ces compromis peuvent l’emporter sur la modification de la qualité du service.

Encouragez les membres de votre équipe à expérimenter continuellement des améliorations en matière de durabilité lors de l’élaboration des exigences fonctionnelles. Les équipes doivent également intégrer des métriques de substitution lors de la fixation des objectifs afin de s’assurer qu’elles évaluent l’intensité des ressources lors de l’élaboration de leurs charges de travail.

Voici des exemples de compromis permettant de réduire les ressources cloud que vous consommez :

Ajuster la qualité du résultat : vous pouvez effectuer un compromis sur la qualité des résultats (QoR) contre une réduction de l’intensité de la charge de travail grâce au calcul approximatif. La pratique du calcul approximatif recherche des opportunités d’exploitation de l’écart entre ce dont les clients ont besoin et ce que vous produisez réellement. Par exemple, si vous placez vos données dans une structure de données définie, vous pouvez ajouter l'opérateur ORDER BY SQL pour supprimer tout traitement inutile, économisant ainsi des ressources tout en fournissant une réponse acceptable.

Ajuster le temps de réponse : un temps de réponse plus lent peut permettre d’économiser du carbone en minimisant les frais généraux partagés. Le traitement des tâches éphémères ad hoc peut entraîner des frais généraux de démarrage. Groupez les tâches et traitez-les par lots au lieu de payer les frais généraux chaque fois qu’une tâche arrive. Le traitement par lots échange une augmentation du temps de réponse contre une réduction des frais généraux partagés liés au lancement d’une instance, au téléchargement du code source et à l’exécution du processus.

Ajuster la disponibilité : avec AWS, vous pouvez ajouter de la redondance et atteindre les objectifs de haute disponibilité en quelques clics. Vous pouvez augmenter la redondance grâce à des techniques comme la stabilité statique en provisionnant des ressources inactives qui entraînent toujours une diminution de l’utilisation. Évaluez les besoins de l’entreprise lors de la fixation des objectifs. Des

compromis relativement mineurs en matière de disponibilité peuvent entraîner des améliorations beaucoup plus importantes en matière d'utilisation. Par exemple, le modèle d'architecture de stabilité statique consiste à mettre en service une capacité de basculement inutilisée pour une prise en charge immédiate après une défaillance d'un composant. Un relâchement des exigences de disponibilité peut supprimer le besoin de capacité en ligne inutilisée en laissant le temps à l'automatisation de déployer des ressources de remplacement. L'ajout de capacité de basculement à la demande entraîne une utilisation globale plus élevée, sans impact sur l'entreprise pendant les opérations normales, et présente l'avantage secondaire de réduire les coûts.

Bonnes pratiques pour la durabilité dans le cloud

Optimisez l'emplacement de la charge de travail, ainsi que votre architecture en fonction de la demande, des logiciels, des données, du matériel et des processus afin d'accroître l'efficacité énergétique. Chacun de ces domaines représente des opportunités d'utiliser les bonnes pratiques visant à réduire l'impact en matière de durabilité de votre charge de travail dans le cloud. Il s'agit là d'optimiser l'utilisation et de réduire le gaspillage ainsi que l'ensemble des ressources déployées et alimentées pour assurer votre charge de travail.

Rubriques

- Sélection d'une région
- Alignement sur la demande
- Logiciels et architecture
- Gestion des données
- Matériel et services
- Processus et culture

Sélection d'une région

Le choix de la région en fonction de votre charge de travail influe considérablement sur ses indicateurs de performance clés, y compris les performances, les coûts et l'empreinte carbone. Pour améliorer efficacement ces indicateurs de performance clés, vous devez choisir les régions pour votre charge de travail en fonction des exigences et des objectifs de durabilité de votre entreprise.

Bonnes pratiques

- SUS01-BP01 Choisir une région en fonction des exigences et des objectifs de durabilité de l'entreprise

SUS01-BP01 Choisir une région en fonction des exigences et des objectifs de durabilité de l'entreprise

Choisissez une région pour votre charge de travail en fonction des exigences et des objectifs de durabilité de votre entreprise afin d'optimiser ses KPI, dont les performances, les coûts et l'empreinte carbone.

Anti-modèles courants :

- Vous sélectionnez la région de la charge de travail en fonction de votre propre emplacement.
- Vous regroupez toutes les ressources de charge de travail dans un seul emplacement géographique.

Avantages liés au respect de cette bonne pratique : le fait de placer une charge de travail à proximité des projets d'énergie renouvelable d'Amazon ou des régions affichant une faible intensité de carbone publiée peut contribuer à réduire l'empreinte carbone d'une charge de travail dans le cloud.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Le AWS Cloud est un réseau en constante expansion de régions et de points de présence (PoP), avec une infrastructure de réseau mondiale les reliant entre eux. Le choix de la région en fonction de votre charge de travail influe considérablement sur ses indicateurs de performance clés, y compris les performances, les coûts et l'empreinte carbone. Pour améliorer efficacement ces indicateurs de performance clés, vous devez choisir les régions pour votre charge de travail en fonction des exigences et des objectifs de durabilité de votre entreprise.

Étapes d'implémentation

- Présélectionner les régions potentielles : suivez ces étapes pour évaluer et présélectionner les régions potentielles pour votre charge de travail en fonction des exigences de votre entreprise, notamment de la conformité, des fonctionnalités disponibles, du coût et de la latence :
  - Confirmez que ces régions sont conformes aux réglementations locales en vigueur (par exemple, la souveraineté des données).
  - Utilisez les listes de services régionaux AWS pour vérifier si les régions disposent des services et des fonctionnalités dont vous avez besoin pour gérer votre charge de travail.
  - Calculez le coût de la charge de travail pour chaque région à l'aide du Calculateur de tarification AWS.
  - Testez la latence du réseau entre les emplacements des utilisateurs finaux et chaque Région AWS.
- Choisir des régions : choisissez des régions proches de projets Amazon d'énergie renouvelable et des régions où le réseau a une intensité de carbone publiée inférieure à celle d'autres sites (ou régions).

- Identifiez vos directives de durabilité pertinentes pour suivre et comparer les émissions de carbone d'une année à l'autre sur la base du protocole sur les gaz à effet de serre (méthodes basées sur le marché et basées sur la localisation).
- Choisissez une région en fonction de la méthode que vous utilisez pour suivre les émissions de carbone. Pour plus de détails sur le choix d'une région en fonction de vos directives en matière de développement durable, consultez Comment sélectionner une région pour votre charge de travail en fonction des objectifs de durabilité.

Ressources

Documents connexes :

- Comprendre les estimations de vos émissions de carbone
- Amazon à travers le monde
- Méthodologie de l'énergie renouvelable
- Éléments à prendre en compte lors de la sélection d'une région pour vos charges de travail

Vidéos connexes :

- AWS re:Invent 2023 - Sustainability innovation in AWS Global Infrastructure
- AWS re:Invent 2023 – Architecture durable : passé, présent et future
- AWS re:Invent 2022 – Fournir des architectures durables et performantes
- AWSre:Invent 2022 – L'architecture de manière durable et réduisez votre AWSempreinte carbone
- AWSre:Invent 2022 – Durabilité dans AWS les infrastructures mondiales

Alignement sur la demande

La façon dont les utilisateurs et les applications consomment vos charges de travail et d'autres ressources peut vous aider à identifier les améliorations nécessaires pour atteindre vos objectifs de durabilité. Mettez à l'échelle l'infrastructure pour répondre en permanence à la demande et vérifiez que vous n'utilisez que les ressources minimales requises pour prendre en charge vos utilisateurs. Alignez les niveaux de service sur les besoins des clients. Positionnez des ressources afin de limiter le réseau nécessaire aux utilisateurs et aux applications pour les consommer. Supprimez les ressources inutilisées. Fournissez aux membres de votre équipe des appareils qui répondent à leurs besoins et minimisent leur impact en matière de durabilité.

Bonnes pratiques

- SUS02-BP01 Mettre à l’échelle l’infrastructure de la charge de travail de façon dynamique
- SUS02-BP02 S'aligner sur les objectifs SLAs de durabilité
- SUS02-BP03 Arrêter la création et la maintenance des actifs inutilisés
- SUS02-BP04 Optimiser le placement géographique des charges de travail en fonction de leurs exigences en matière de réseau
- SUS02-BP05 Optimiser les ressources des membres de l'équipe pour les activités réalisées
- SUS02-BP06 Mettre en œuvre la mise en mémoire tampon ou la régulation pour aplatir la courbe de demande

SUS02-BP01 Mettre à l’échelle l’infrastructure de la charge de travail de façon dynamique

Utilisez l’élasticité du cloud et mettez à l’échelle votre infrastructure de façon dynamique afin de rapprocher l’offre de ressources cloud de la demande et d’éviter de sur provisionner une capacité dans votre charge de travail.

Anti-modèles courants :

- Vous ne mettez pas à l’échelle votre infrastructure avec la charge de l’utilisateur.
- Vous mettez à l’échelle manuellement votre infrastructure en permanence.
- Vous conservez une capacité accrue après un événement de mise à l’échelle au lieu de la réduire.

Avantages liés au respect de cette bonne pratique : la configuration et le test de l’élasticité de la charge de travail permettent de faire correspondre efficacement l’offre de ressources cloud à la demande et d’éviter le sur provisionnement de capacité. Vous pouvez profiter de l’élasticité du cloud pour mettre à l’échelle automatiquement la capacité pendant et après les pics de demande, afin d’utiliser uniquement le bon nombre de ressources nécessaires pour répondre aux exigences de votre entreprise.

Niveau de risque encouru si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

Le cloud vous apporte la flexibilité dont vous avez besoin pour développer ou réduire vos ressources de manière dynamique via une grande variété de mécanismes afin de répondre aux fluctuations

de la demande. Rapprocher de façon optimale l'offre de la demande a le plus faible impact environnemental pour une charge de travail.

La demande peut être fixe ou variable, ce qui nécessite des métriques et une automatisation pour que la gestion ne devienne pas contraignante. Les applications peuvent se mettre à l'échelle de façon verticale (dans les deux sens) en modifiant la taille de l'instance, de façon horizontale (dans les deux sens) en modifiant le nombre d'instances, ou une combinaison des deux.

Vous pouvez utiliser plusieurs approches pour rapprocher l'offre de ressources de la demande.

- Approche de suivi des objectifs : surveillez votre métrique de capacité de mise à l'échelle et augmentez ou réduisez automatiquement votre capacité selon vos besoins.
- Mise à l'échelle prédictive : mettez à l'échelle en prévision des tendances quotidiennes et hebdomadaires.
- Approche basée sur le calendrier : définissez votre propre calendrier de mise à l'échelle en fonction de changements de charge prévisibles.
- Mise à l'échelle des services : choisissez des services (tels que les services sans serveur) qui sont évolutifs de manière native dès leur conception ou qui proposent une mise à l'échelle automatique en tant que fonctionnalité.

Identifiez les périodes d'utilisation faible ou nulle, et mettez vos ressources à l'échelle afin de supprimer toute capacité excédentaire et améliorer l'efficacité.

Étapes d'implémentation

- L'élasticité correspond à l'offre de ressources dont vous disposez et à la demande pour ces ressources. Les instances, les conteneurs et les fonctions fournissent les mécanismes pour l'élasticité, soit en combinaison avec la mise à l'échelle automatique, soit en tant que fonction du service. AWS fournit une gamme de mécanismes de mise à l'échelle automatique pour veiller à ce que les charges de travail puissent réduire verticalement avec rapidité et facilité pendant les périodes de faible charge utilisateur. Voici des exemples de mécanismes de mise à l'échelle automatique :

Mécanisme de mise à l'échelle automatique	Où utiliser
Amazon EC2 Auto Scaling	Utilisez-le pour vous assurer que vous disposez du nombre adéquat d'instances

Mécanisme de mise à l'échelle automatique	Où utiliser
	Amazon EC2 disponibles pour gérer la charge utilisateur de votre application.
Application Autoscaling	Permet de mettre à l'échelle automatiquement les ressources pour des services AWS individuels au-delà d'Amazon EC2, tels que les fonctions Lambda ou les services Amazon Elastic Container Service (Amazon ECS).
Outil Cluster Autoscaler de Kubernetes	Utilisez-le pour mettre à l'échelle automatiq uement les clusters Kubernetes sur AWS.

- La mise à l'échelle est souvent abordée pour les services de calcul, tels que les instances Amazon EC2 ou les fonctions AWS Lambda. Envisagez la configuration de services non informatiques tels que les unités de capacité de lecture et d'écriture Amazon DynamoDB ou les partitions Amazon Kinesis Data Streams pour répondre à la demande.
- Vérifiez que les métriques de l'augmentation ou de la diminution sont validées par rapport au type de charge de travail déployée. Si vous déployez une application de transcodage vidéo, une utilisation de 100 % du processeur est attendue. N'en faites pas votre métrique principale. Le cas échéant, vous pouvez utiliser une métrique personnalisée (telle que l'utilisation de la mémoire) pour votre politique de mise à l'échelle. Pour choisir les bonnes métriques, tenez compte des conseils suivants pour Amazon EC2 :
  - La métrique doit être une métrique d'utilisation valide et décrire à quel point l'instance est occupée.
  - La valeur de métrique doit augmenter ou diminuer en proportion du nombre d'instances présentes dans le groupe Auto Scaling.
- Utilisez une mise à l'échelle dynamique plutôt qu'une mise à l'échelle manuelle pour votre groupe Auto Scaling. Nous vous recommandons également d'utiliser des politiques de dimensionnement pour le suivi des cibles dans votre dimensionnement dynamique.
- Vérifiez que les déploiements de charges de travail peuvent gérer à la fois les événements d'augmentation et de diminution des charges de travail. Créez des scénarios de test pour les événements de diminution afin de vérifier que la charge de travail se comporte comme prévu et n'a aucun impact sur l'expérience utilisateur (comme la perte de sessions permanentes). Vous pouvez utiliser l'historique des activités pour vérifier une activité de mise à l'échelle dans un groupe Auto Scaling.

- Évaluez votre charge de travail pour les modèles prédictifs et mettez-la à l'échelle de manière proactive pour anticiper les changements prévisibles et prévus de la demande. Avec la mise à l'échelle prédictive, vous pouvez supprimer le besoin de surprovisionner de la capacité. Pour en savoir plus, reportez-vous à Mise à l'échelle prédictive avec Amazon EC2 Auto Scaling.

Ressources

Documents connexes :

- Mise en route avec Amazon EC2 Auto Scaling
- Mise à l'échelle prédictive pour EC2 alimentée par le machine learning
- Analyser le comportement des utilisateurs avec Amazon OpenSearch Service, Amazon Data Firehose et Kibana
- Qu'est-ce qu'Amazon CloudWatch ?
- Surveillance de la charge de la base de données avec Performance Insights sur Amazon RDS
- Présentation de la prise en charge native pour la mise à l'échelle prédictive avec Amazon EC2 Auto Scaling
- Présentation de Karpenter, un Cluster Autoscaler de Kubernetes hautement performant et open source
- Présentation approfondie d'Amazon ECS Cluster Auto Scaling

Vidéos connexes :

- AWSre:Invent 2023 - Mise à l'échelle AWS des 10 premiers millions d'utilisateurs
- AWS re:Invent 2023 - Architecture durable : passé, présent et future
- AWS re:Invent 2022 - Build a cost-, energy-, and resource-efficient compute environment
- AWS re:Invent 2022 - Scaling containers from one user to millions
- AWS re:Invent 2#023 - Scaling FM inference to hundreds of models with Amazon SageMaker AI
- AWS re:Invent 2023 - Harness the power of Karpenter to scale, optimize & upgrade Kubernetes

Exemples connexes :

- Autoscaling

SUS02-BP02 S'aligner sur les objectifs SLAs de durabilité

Passez en revue et optimisez les accords de niveau de service relatifs à la charge de travail (SLA) en fonction de vos objectifs de durabilité afin de minimiser les ressources nécessaires pour soutenir votre charge de travail tout en continuant à répondre aux besoins de l'entreprise.

Anti-modèles courants :

- La charge SLAs de travail est inconnue ou ambiguë.
- Vous définissez votre objectif SLA uniquement en fonction de la disponibilité et des performances.
- Vous utilisez le même modèle de conception (comme une architecture multi-AZ) pour toutes vos charges de travail.

Avantages de l'établissement de cette meilleure pratique : L'alignement sur les objectifs SLAs de durabilité permet une utilisation optimale des ressources tout en répondant aux besoins de l'entreprise.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : faible

Directives d’implémentation

SLAsdéfinissez le niveau de service attendu d'une charge de travail dans le cloud, tel que le temps de réponse, la disponibilité et la conservation des données. Ils influent sur l’architecture, l’utilisation des ressources et l’impact environnemental d’une charge de travail sur le cloud. À un rythme régulier, passez en revue SLAs et faites des compromis qui réduisent considérablement l'utilisation des ressources en échange de baisses acceptables des niveaux de service.

Étapes d’implémentation

- Comprenez les objectifs de durabilité : identifiez les objectifs de durabilité de votre organisation, tels que la réduction des émissions de carbone ou l’amélioration de l’utilisation des ressources.
- Révision SLAs : Évaluez vos SLAs produits pour déterminer s'ils répondent aux besoins de votre entreprise. Si vous dépassezSLAs, effectuez un examen plus approfondi.
- Comprenez les compromis : déterminez les compromis entre la complexité de votre charge de travail (comme le volume élevé d’utilisateurs simultanés), les performances (comme la latence) et l’impact sur le développement durable (comme les ressources requises). Généralement, la priorisation de deux des facteurs se fait au détriment du troisième.

- Ajustez SLAs : ajustez le vôtre SLAs en faisant des compromis qui réduisent de manière significative les impacts sur le développement durable en échange de baisses acceptables des niveaux de service.
  - Durabilité et fiabilité : les charges de travail hautement disponibles ont tendance à consommer davantage de ressources.
  - Durabilité et performance : l'utilisation de plus de ressources pour améliorer les performances pourrait avoir un impact environnemental plus important.
  - Durabilité et sécurité : des charges de travail trop sécurisées peuvent avoir un impact environnemental plus important.
- Définissez la durabilité SLAs si possible : incluez la durabilité SLAs pour votre charge de travail. Par exemple, définissez un niveau d'utilisation minimal afin de garantir la durabilité SLA de vos instances de calcul.
- Utilisez des modèles de conception efficaces : utilisez des modèles de conception tels que les microservices AWS qui donnent la priorité aux fonctions critiques pour l'entreprise et permettent de réduire les niveaux de service (tels que les objectifs de temps de réponse ou de temps de reprise) pour les fonctions non critiques.
- Communiquez et responsabilisez : partagez-les SLAs avec toutes les parties prenantes concernées, y compris votre équipe de développement et vos clients. Utilisez les rapports pour suivre et surveiller lesSLAs. Attribuez la responsabilité d'atteindre les objectifs de durabilité de votre entrepriseSLAs.
- Utilisez des incitations et des récompenses : utilisez des incitations et des récompenses pour atteindre ou SLAs dépasser les objectifs de durabilité.
- Révisez et répétez : révisez et ajustez régulièrement vos objectifs SLAs pour vous assurer qu'ils correspondent à l'évolution des objectifs de durabilité et de performance.

Ressources

Documents connexes :

- Comprenez les modèles de résilience et les compromis pour concevoir une architecture efficace dans le cloud
- Importance du contrat de niveau de service pour les fournisseurs de SaaS

Vidéos connexes :

- AWS re:Invent 2023 - Capacité, disponibilité, rentabilité : choisissez trois
- AWS re:INVENT 2023 - Architecture durable : passé, présent et futur
- AWS re:Invent 2023 - Modèles d'intégration avancés et compromis pour les systèmes faiblement couplés
- AWS re:Invent 2022 - Fournir des architectures durables et performantes
- AWS re:Invent 2022 - Créez un environnement informatique économe en termes de coûts, d'énergie et de ressources

SUS02-BP03 Arrêter la création et la maintenance des actifs inutilisés

Mettez hors service les ressources inutilisées de votre charge de travail afin de réduire le nombre de ressources cloud nécessaires pour répondre à votre demande et minimiser le gaspillage.

Anti-modèles courants :

- Vous n’analysez pas votre application pour détecter les ressources redondantes ou qui ne sont plus nécessaires.
- Vous ne supprimez pas les ressources redondantes ou qui ne sont plus nécessaires.

Avantages de l’établissement de cette meilleure pratique : la suppression des actifs inutilisés libère des ressources et améliore l’efficacité globale de la charge de travail.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : faible

Directives d’implémentation

Les ressources inutilisées consomment les ressources du cloud telles que l’espace de stockage et la puissance de calcul. En identifiant et en éliminant ces ressources, vous pouvez les libérer, ce qui se traduit par une architecture cloud plus efficace. Analysez régulièrement les ressources de l’application telles que les rapports pré-compilés, les jeux de données, les images statiques et les modèles d’accès aux ressources pour identifier des redondances, une sous-utilisation et d’éventuelles cibles de mise hors service. Supprimez ces ressources redondantes pour réduire le gaspillage de ressources dans votre charge de travail.

Étapes d’implémentation

- Réaliser un inventaire : réalisez un inventaire complet pour identifier tous les actifs relevant de votre charge de travail.

- Analyser l'utilisation : la surveillance continue vous permet d'identifier les actifs statiques qui ne sont plus nécessaires.
- Supprimer les actifs inutilisés : élaborez un plan pour supprimer les actifs dont vous n'avez plus besoin.
  - Avant de supprimer une ressource, évaluez l'impact de sa suppression sur l'architecture.
  - Consolidez les ressources générées qui se chevauchent afin de supprimer tout traitement redondant.
  - Mettez à jour vos applications pour ne plus produire et stocker les ressources qui ne sont pas nécessaires.
- Communiquer avec des tiers : demandez aux tiers d'arrêter de produire et de stocker les ressources gérées en votre nom qui ne sont plus nécessaires. Demandez à consolider les actifs redondants.
- Utilisez des politiques de cycle de vie : utilisez des politiques de cycle de vie pour supprimer automatiquement les actifs inutilisés.
  - Vous pouvez utiliser Amazon S3 Lifecycle afin de gérer vos objets au cours de leur cycle de vie.
  - Vous pouvez utiliser Amazon Data Lifecycle Manager pour automatiser la création, la conservation et la suppression des EBS instantanés Amazon et des produits EBS sauvegardés par AMIs Amazon.
- Examiner et optimiser : examinez régulièrement votre charge de travail pour identifier et supprimer les ressources inutilisées.

Ressources

Documents connexes :

- Optimisation de votre AWS infrastructure pour la durabilité, partie II : stockage
- Comment puis-je résilier les ressources actives dont je n'ai plus besoin sur mon ordinateur Compte AWS ?

Vidéos connexes :

- AWS re:INVENT 2023 - Architecture durable : passé, présent et futur
- AWS re:Invent 2022 - Préserver et optimiser la valeur des actifs multimédias numériques à l'aide d'Amazon S3

- AWS re:Invent 2023 - Optimisez les coûts dans vos environnements multi-comptes

SUS02-BP04 Optimiser le placement géographique des charges de travail en fonction de leurs exigences en matière de réseau

Pour votre charge de travail, sélectionnez un emplacement et des services cloud qui réduisent la distance que le trafic réseau doit parcourir et diminuent les ressources réseau totales requises pour gérer votre charge de travail.

Anti-modèles courants :

- Vous sélectionnez la région de la charge de travail en fonction de votre propre emplacement.
- Vous regroupez toutes les ressources de charge de travail dans un seul emplacement géographique.
- Tout le trafic passe par vos centres de données existants.

Avantages liés au respect de cette bonne pratique : placer une charge de travail à proximité de ses clients fournit une faible latence, tout en réduisant les mouvements de données sur le réseau ainsi que l'impact sur l'environnement.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

L' AWS Cloud infrastructure est construite autour d'options de localisation telles que les régions, les zones de disponibilité, les groupes de placement et les emplacements périphériques tels que AWS Outpostsles zones AWS locales. Ces options d'emplacement sont responsables de la connectivité entre les composants d'application, les services cloud, les réseaux périphériques et les centres de données sur site.

Analysez les modèles d'accès au réseau dans votre charge de travail pour identifier comment utiliser ces options de localisation dans le cloud et réduire la distance que le trafic réseau doit parcourir.

Étapes d'implémentation

- Analysez les modèles d'accès au réseau dans votre charge de travail afin d'identifier comment les utilisateurs utilisent votre application.

  - Utilisez des outils de surveillance, tels qu'Amazon CloudWatch AWS CloudTrail, pour recueillir des données sur les activités du réseau.
  - Analysez les données pour identifier le modèle d’accès au réseau.
- Choisissez les régions pour votre déploiement de charge de travail en fonction des éléments clés suivants :
  - Votre objectif en matière de développement durable : comme expliqué dans Sélection de la région.
  - Lieu de stockage de vos données : pour les applications utilisant de grandes quantités de données (telles que le big data et le machine learning). Le code de l’application doit s’exécuter aussi près que possible des données.
  - Lieu de stockage de vos données : pour les applications orientées utilisateur, choisissez une région (ou des régions) proche des utilisateurs de votre charge de travail.
  - Autres contraintes : tenez compte des contraintes telles que le coût et la conformité, comme expliqué dans la section Éléments à prendre en compte lors de la sélection d’une région pour vos charges de travail.
- Utilisez des solutions de mise en cache locale ou proposées par AWS pour les ressuorces fréquemment utilisées afin d’améliorer les performances, de réduire les déplacements de données et de diminuer l’impact environnemental.

Service	Utilisation
Amazon CloudFront	Utilisez-le pour mettre en cache du contenu statique tel que des images, des scripts et des vidéos, ainsi que du contenu dynamique tel que API des réponses ou des applications Web.
Amazon ElastiCache	Permet de mettre en cache du contenu pour les applications Web.
DynamoDB Accelerator	Permet d’ajouter une accélération en mémoire à vos tables DynamoDB.

- Utilisez des services capables de vous aider à exécuter du code plus proche des utilisateurs de votre charge de travail :

Service	Utilisation
Lambda@Edge	Destiné aux opérations exigeantes en puissance de calcul qui sont lancées lorsque des objets ne sont pas dans le cache.
CloudFront Fonctions Amazon	À utiliser pour des cas d'utilisation simples tels que HTTP les manipulations de demandes ou de réponses qui peuvent être initiées par des fonctions de courte durée.
AWS IoT Greengrass	Permet d'exécuter du calcul local, une messagerie et une mise en cache de données pour les appareils connectés.

- Utilisez le regroupement de connexions afin de pouvoir réutiliser les connexions et réduire les ressources nécessaires.
- Utilisez des magasins de données distribués qui ne s'appuient pas sur des connexions persistantes ni sur des mises à jour synchrones pour des raisons de cohérence afin de servir les populations régionales.
- Remplacez la capacité du réseau statique pré-allouée par une capacité dynamique partagée, et partagez l'impact en matière de durabilité de la capacité du réseau avec d'autres abonnés.

Ressources

Documents connexes :

- Optimisation de votre AWS infrastructure pour la durabilité, partie III : mise en réseau
- ElastiCache Documentation Amazon
- Qu'est-ce qu'Amazon CloudFront ?
- CloudFront Principales fonctionnalités d'Amazon
- AWS Infrastructure mondiale
- AWS Zones locales et AWS Outposts choix de la technologie adaptée à votre charge de travail périphérique
- Groupes de placement

- AWS Zones Locales
- AWS Outposts

Vidéos connexes :

- Démystifier le transfert de données sur AWS
- Augmenter les performances du réseau sur les instances Amazon EC2 de nouvelle génération
- AWS Vidéo explicative sur les Zones Locales
- AWS Outposts: Overview and How it Works
- AWS re:Invent 2023 - Une stratégie de migration pour les charges de travail en périphérie et sur site
- AWS re:INVENT 2021 - AWS Outposts : Apporter l' AWS expérience sur site
- AWS re:Invent 2020 - AWS Wavelength : Exécutez des applications avec une latence extrêmement faible à la périphérie de la 5G
- AWS re:Invent 2022 - Zones AWS locales : création d'applications pour une périphérie distribuée
- AWS re:Invent 2021 - Création de sites Web à faible latence avec Amazon CloudFront
- AWS re:Invent 2022 - Améliorez les performances et la disponibilité avec AWS Global Accelerator
- AWS re:Invent 2022 - Construisez votre réseau étendu mondial en utilisant AWS
- AWS re:Invent 2020 : gestion du trafic mondial avec Amazon Route 53

Exemples connexes :

- AWS Ateliers de réseautage
- Architecting for sustainability - Minimize data movement across networks

SUS02-BP05 Optimiser les ressources des membres de l'équipe pour les activités réalisées

Optimisez les ressources fournies aux membres de l’équipe pour réduire l’impact sur la durabilité environnementale tout en répondant à leurs besoins.

Anti-modèles courants :

- Vous ignorez l’impact des appareils utilisés par les membres de votre équipe sur l’efficacité globale de votre application cloud.
- Vous gérez et mettez à jour manuellement les ressources utilisées par les membres de l’équipe.

Avantages liés au respect de cette bonne pratique : l'optimisation des ressources des membres de l'équipe améliore l'efficacité globale des applications compatibles avec le cloud.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : faible

Directives d’implémentation

Comprenez les ressources que les membres de votre équipe utilisent pour consommer vos services, leur cycle de vie prévu et l’impact financier et sur la durabilité. Mettez en œuvre des stratégies pour optimiser ces ressources. Par exemple, effectuez des opérations complexes, telles que le rendu et la compilation, sur une infrastructure évolutive hautement utilisée plutôt que sur des systèmes mono-utilisateurs puissants et sous-utilisés.

Étapes d’implémentation

- Utilisez des postes de travail économes en énergie : fournissez aux membres de l'équipe des postes de travail et des périphériques économes en énergie. Utilisez des fonctionnalités de gestion de l’alimentation efficaces (comme le mode faible consommation) sur ces appareils afin de réduire leur consommation d’énergie.
- Adopter la virtualisation : utilisez des bureaux virtuels et le streaming d'applications pour limiter les exigences liées aux mises à niveau et aux appareils.
- Encouragez la collaboration à distance : encouragez les membres de l'équipe à utiliser des outils de collaboration à distance tels qu'Amazon Chime ou AWS Wickr à réduire les déplacements et les émissions de carbone associées.
- Utilisez des logiciels économes en énergie : fournissez aux membres de l'équipe des logiciels économes en énergie en supprimant ou en désactivant les fonctionnalités et les processus inutiles.
- Gérer les cycles de vie : évaluez l'impact des processus et des systèmes sur le cycle de vie de votre appareil et choisissez des solutions qui réduisent au minimum le besoin de remplacer celui-ci tout en répondant aux exigences de l'entreprise. Entretenez et mettez à jour régulièrement les postes de travail et les logiciels afin de maintenir et d’améliorer l’efficacité.
- Gestion des appareils à distance : intégrez la gestion à distance des appareils afin de réduire les déplacements professionnels nécessaires.

- AWS Systems Manager Fleet Manager est une interface utilisateur (UI) unifiée qui vous permet de gérer à distance vos nœuds exécutés sur site AWS ou sur site.

Ressources

Documents connexes :

- Qu'est-ce qu'Amazon WorkSpaces ?
- Optimiseur de coûts pour Amazon WorkSpaces
- Documentation Amazon AppStream 2.0
- NICE DCV

Vidéos connexes :

- Gestion des coûts pour Amazon WorkSpaces sur AWS

SUS02-BP06 Mettre en œuvre la mise en mémoire tampon ou la régulation pour aplatir la courbe de demande

La mise en mémoire tampon et la limitation aplatissent la courbe de la demande et réduisent la capacité provisionnée requise pour votre charge de travail.

Anti-modèles courants :

- Vous traitez les demandes des clients immédiatement alors que ce n’est pas nécessaire.
- Vous n’analysez pas les exigences des demandes des clients.

Avantages liés au respect de cette bonne pratique : l’aplatissement de la courbe de demande réduit la capacité provisionnée requise pour la charge de travail. En réduisant la capacité provisionnée, on réduit la consommation d’énergie et l’impact environnemental.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : faible

Directives d’implémentation

L’aplatissement de la courbe de demande de la charge de travail peut vous aider à réduire la capacité provisionnée pour une charge de travail et à réduire son impact environnemental. Prenons

l'exemple une charge de travail dont la courbe de demande est représentée dans la figure ci-dessous. Cette charge de travail a deux pics, et pour gérer ces pics, la capacité des ressources comme indiqué par la ligne orange est provisionnée. Les ressources et l'énergie utilisées pour cette charge de travail ne sont pas indiquées par la zone sous la courbe de la demande, mais par la zone sous la ligne de la capacité provisionnée, car la capacité provisionnée est nécessaire pour gérer ces deux pics.

Courbe de demande avec deux pics distincts nécessitant une capacité allouée élevée

Vous pouvez utiliser la mise en mémoire tampon ou la limitation pour modifier la courbe de la demande et lisser les pics, ce qui signifie moins de capacité provisionnée et moins d'énergie consommée. Mettez en œuvre la limitation lorsque vos clients peuvent effectuer de nouvelles tentatives. Mettez en œuvre une mémoire tampon pour stocker la demande et reporter le traitement.

Effet de la limitation sur la courbe de demande et la capacité provisionnée.

Étapes d'implémentation

- Analysez les demandes des clients pour déterminer comment y répondre. Les questions à se poser sont les suivantes :
  - Cette demande peut-elle être traitée de manière asynchrone ?
  - Le client a-t-il la possibilité de lancer de nouvelles tentatives ?
- Si le client a la possibilité de lancer de nouvelles tentatives, vous pouvez mettre en œuvre un système de limitation, qui indique à la source que si elle ne peut pas répondre à la demande au moment même, elle doit réessayer plus tard.
  - Vous pouvez utiliser Amazon API Gateway pour implémenter la régulation.
- Pour les clients qui ne peuvent pas effectuer de nouvelles tentatives, il faut mettre en œuvre un tampon pour aplanir la courbe de demande. Un tampon diffère le traitement des demandes, ce qui permet aux applications qui s'exécutent à différents débits de communiquer efficacement. Une approche basée sur la mémoire tampon utilise une file d'attente ou un flux pour accepter les messages des producteurs. Les messages sont lus par les consommateurs et traités, ce qui permet aux messages de fonctionner au rythme qui répond aux besoins des entreprises.
  - Amazon Simple Queue Service (AmazonSQS) est un service géré qui fournit des files d'attente permettant à un seul consommateur de lire des messages individuels.

  - Amazon Kinesis fournit un flux de données qui permet à de nombreux consommateurs de lire les mêmes messages.
- Analysez la demande globale, le taux de variation et le temps de réponse requis pour dimensionner correctement la limitation ou le tampon nécessaire.

Ressources

Documents connexes :

- Commencer à utiliser Amazon SQS
- Intégration d'applications à l'aide de files d'attente et de messages
- Gestion et surveillance de la API régulation de vos charges de travail
- Limiter à grande échelle un système multi-tenant hiérarchisé à l'aide REST API de Gateway API
- Intégration d'applications à l'aide de files d'attente et de messages

Vidéos connexes :

- AWS re:Invent 2022 - Modèles d'intégration d'applications pour les microservices
- AWS re:Invent 2023 - Économies intelligentes : stratégies d'optimisation des coûts d'Amazon EC2
- AWS re:Invent 2023 - Modèles d'intégration avancés et compromis pour les systèmes faiblement couplés

Logiciels et architecture

Mettez en œuvre des modèles permettant de lisser les charges et de conserver une haute utilisation constante des ressources déployées afin de réduire les ressources consommées. Les composants peuvent devenir inactifs s'ils ne sont pas utilisés à la suite de changements de comportement des utilisateurs dans le temps. Révisez les modèles et l'architecture afin de consolider les composants sous-utilisés et d'augmenter l'utilisation globale. Mettez hors service les composants qui ne sont plus nécessaires. Soyez sensibilisé aux performances des composants de vos charges de travail et optimisez les composants qui consomment le plus de ressources. Soyez au courant des appareils que vos clients utilisent pour accéder à vos services et mettez en œuvre des modèles qui réduisent le besoin de mettre à niveau les appareils.

Bonnes pratiques

- SUS03-BP01 Optimiser le logiciel et l'architecture pour les tâches asynchrones et planifiées
- SUS03-BP02 Supprimer ou refactoriser les composants de charges de travail faiblement utilisés ou inutilisés
- SUS03-BP03 Optimiser les sections de votre code qui consomment le plus de temps ou de ressources
- SUS03-BP04 Optimiser l'impact sur les appareils et équipements
- SUS03-BP05 Utiliser des modèles logiciels et des architectures qui soutiennent au mieux l’accès aux données et les modèles de stockage.

SUS03-BP01 Optimiser le logiciel et l'architecture pour les tâches asynchrones et planifiées

Utilisez des modèles d’architecture et de logiciels efficaces comme ceux axés sur les files d’attente afin de maintenir une utilisation élevée et constante des ressources déployées.

Anti-modèles courants :

- Vous mettez en service trop de ressources dans votre charge de travail cloud pour répondre aux pics imprévus de la demande.
- Votre architecture ne découple pas les expéditeurs et les destinataires de messages asynchrones par un composant de messagerie.

Avantages liés au respect de cette bonne pratique :

- Des modèles de logiciels et d’architecture efficaces réduisent les ressources inutilisées dans votre charge de travail et améliorent l’efficacité globale.
- Vous pouvez mettre à l’échelle le traitement indépendamment de la réception de messages asynchrones.
- Par le biais d’un composant de messagerie, vous avez assoupli les exigences de disponibilité auxquelles vous pouvez répondre avec moins de ressources.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

Utilisez des modèles d'architecture efficaces, tels que l’architecture axée sur les événements, qui permettent une utilisation uniforme des composants et minimisent le surprovisionnement de votre charge de travail. L’utilisation de modèles d’architecture efficaces réduit au minimum les ressources inutilisées en raison des changements de la demande au fil du temps.

Comprenez les exigences des composants de votre charge de travail et adoptez des modèles d’architecture qui augmentent l’utilisation globale des ressources. Mettez hors service les composants qui ne sont plus nécessaires.

Étapes d’implémentation

- Analysez la demande pour votre charge de travail afin de déterminer comment y répondre.
- Pour les demandes ou les tâches qui ne nécessitent pas de réponses synchrones, utilisez des architectures axées sur les files d’attente et des agents de travail de mise à l’échelle automatique afin de maximiser l’utilisation. Voici quelques exemples de situations où vous pourriez envisager une architecture axée sur les files d’attente :

Mécanismes de mise en file d'attente	Description
AWS Batch files d'attente pour les emplois	AWS Batch les tâches sont soumises à une file d'attente où elles résident jusqu'à ce qu'elles puissent être planifiées pour être exécutées dans un environnement informatique.
Amazon Simple Queue Service et instances Amazon EC2 Spot	Associer des instances Amazon SQS et Spot pour créer une architecture efficace et tolérante aux pannes.

- Pour les demandes ou les tâches qui peuvent être traitées à tout moment, utilisez les mécanismes de planification afin de traiter les tâches par lots pour plus d’efficacité. Voici quelques exemples de mécanismes de planification sur AWS :

Mécanismes de planification	Description
Amazon EventBridge Scheduler	Une fonctionnalité d'Amazon EventBridge qui vous permet de créer, d'exécuter et de gérer des tâches planifiées à grande échelle.
AWS Glue calendrier basé sur le temps	Définissez un calendrier basé sur le temps pour vos robots d'exploration et vos tâches dans. AWS Glue
Tâches planifiées d'Amazon Elastic Container Service (AmazonECS)	Amazon ECS prend en charge la création de tâches planifiées. Les tâches planifiées utilisent EventBridge les règles d'Amazon pour exécuter des tâches selon un calendrier ou en réponse à un EventBridge événement.
Instance Scheduler	Configurez les plannings de démarrage et d'arrêt pour vos EC2 instances Amazon et Amazon Relational Database Service.

- Si vous utilisez des mécanismes d’interrogation et de webhooks dans votre architecture, remplacez-les par des événements. Utilisez des architectures axées sur les événements pour créer des charges de travail hautement efficaces.
- Tirez parti du mode sans serveur AWS pour éliminer l'infrastructure surprovisionnée.
- Dimensionnez les composants individuels afin d’éviter les ressources inactives attendant une entrée.
  - Vous pouvez utiliser les recommandations de redimensionnement dans AWS Cost Explorer ou AWS Compute Optimizer pour identifier les opportunités de redimensionnement.
  - Pour en savoir plus, se reporter à Dimensionnement adéquat : dimensionnement des instances en fonction des charges de travail.

Ressources

Documents connexes :

- Qu'est-ce qu'Amazon Simple Queue Service ?

- Qu'est-ce qu'Amazon MQ ?
- Mise à l'échelle basée sur Amazon SQS
- Qu'est-ce que c'est AWS Step Functions ?
- Qu'est-ce que c'est AWS Lambda ?
- Utilisation AWS Lambda avec Amazon SQS
- Qu'est-ce qu'Amazon EventBridge ?
- Gestion des flux de travail asynchrones à l'aide d'un REST API

Vidéos connexes :

- AWS re:Invent 2023 - Sur la voie de l'architecture événementielle sans serveur
- AWS re:Invent 2023 - Utilisation du mode sans serveur pour l'architecture axée sur les événements et la conception axée sur le domaine
- AWS re:Invent 2023 - Modèles avancés basés sur les événements avec Amazon EventBridge
- AWS re:INVENT 2023 - Architecture durable : passé, présent et futur
- Modèles de messages asynchrones | Événements AWS

Exemples connexes :

- Architecture axée sur les événements avec processeurs AWS Graviton et instances Amazon Spot EC2

SUS03-BP02 Supprimer ou refactoriser les composants de charges de travail faiblement utilisés ou inutilisés

Supprimez les composants inutilisés et devenus inutiles, et refactorisez les composants peu utilisés afin de minimiser le gaspillage dans votre charge de travail.

Anti-modèles courants :

- Vous ne vérifiez pas régulièrement le niveau d’utilisation des différents composants de votre charge de travail.
- Vous ne vérifiez pas les recommandations des outils de redimensionnement AWS tels que AWS Compute Optimizer.

Avantages liés au respect de cette bonne pratique : la suppression des composants inutilisés minimise le gaspillage et améliore l'efficacité globale de votre charge de travail dans le cloud.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Les composants inutilisés ou sous-utilisés d'une charge de travail dans le cloud consomment des ressources de calcul, de stockage ou de réseau inutiles. Supprimez ou refactorisez ces composants pour réduire directement le gaspillage et améliorer l'efficacité globale d'une charge de travail dans le cloud. Il s'agit d'un processus d'amélioration itératif qui peut être lancé par l'évolution de la demande ou la publication d'un nouveau service cloud. Par exemple, une baisse significative de la durée d'exécution d'une fonction AWS Lambda peut indiquer la nécessité de réduire la taille de la mémoire. De plus, à mesure que AWS publie de nouveaux services et de nouvelles fonctionnalités, les services et l'architecture optimaux pour votre charge de travail peuvent changer.

Surveillez en permanence l'activité de la charge de travail et recherchez les possibilités d'améliorer le niveau d'utilisation des différents composants. En supprimant les composants inutilisés et en effectuant des activités de redimensionnement, vous répondez aux besoins de votre entreprise avec le moins de ressources cloud possible.

Étapes d'implémentation

- Inventorier vos ressources AWS : créez un inventaire de vos ressources AWS. Dans AWS, vous pouvez activer Explorateur de ressources AWS pour explorer et organiser vos ressources AWS. Pour en savoir plus, consultez AWSre:Invent 2022 – Comment gérer les ressources et les applications à grande échelle AWS.
- Surveiller l'utilisation : surveillez et capturez les métriques d'utilisation des composants critiques de votre charge de travail (comme l'utilisation du processeur, l'utilisation de la mémoire ou le débit du réseau dans les métriques Amazon CloudWatch).
- Identifier les composants inutilisés : identifiez les composants inutilisés ou sous-utilisés dans votre architecture.
  - Pour des charges de travail stables, vérifiez les outils de redimensionnement AWS, par exemple AWS Compute Optimizer à intervalles réguliers, pour identifier les composants inactifs, inutilisés ou sous-utilisés.
  - Pour les charges de travail éphémères, évaluez les métriques d'utilisation pour identifier les composants inactifs, inutilisés ou sous-utilisés.

- Supprimer les composants inutilisés : retirez les composants et les ressources associées (comme les images Amazon ECR) qui ne sont plus nécessaires.
  - Nettoyage automatisé des images inutilisées dans Amazon ECR
  - Supprimez des volumes Amazon Elastic Block Store (Amazon EBS) inutilisés en utilisant AWS Config et AWS Systems Manager
- Refactoriser les composants sous-utilisés : refactorisez ou consolidez les composants sous-utilisés avec d'autres ressources pour améliorer l'efficacité de l'utilisation. Par exemple, vous pouvez provisionner plusieurs petites bases de données sur une seule instance de base de données Amazon RDS au lieu d'exécuter des bases de données sur des instances sous-utilisées individuelles.
- Évaluer les améliorations : identifiez les ressources provisionnées par votre charge de travail pour mener à bien une unité de travail. Utilisez ces informations pour évaluer les améliorations obtenues en supprimant ou en refactorisant des composants.
  - Mesure et suivi de l'efficacité du cloud à l'aide de métriques proxy de durabilité, partie I : Que sont les métriques proxy ?
  - Mesure et suivi de l'efficacité du cloud à l'aide de métriques proxy de durabilité, partie II : Établissement d'un pipeline de métriques

Ressources

Documents connexes :

- AWS Trusted Advisor
- Qu'est-ce qu'Amazon CloudWatch ?
- Dimensionnement adéquat : dimensionnement des instances en fonction des charges de travail
- Optimisation de vos coûts avec les recommandations de dimensionnement adéquat

Vidéos connexes :

- AWSre:Invent 2023 - Capacity, availability, cost efficiency: Pick three

SUS03-BP03 Optimiser les sections de votre code qui consomment le plus de temps ou de ressources

Optimisez votre code qui s'exécute dans les différents composants de votre architecture afin de minimiser l'utilisation des ressources tout en maximisant les performances.

Anti-modèles courants :

- Vous ignorez l'optimisation de votre code pour l'utilisation des ressources.
- Vous répondez généralement aux problèmes de performance en augmentant les ressources.
- Votre processus de révision et de développement du code ne permet pas de suivre les variations de performance.

Avantages liés au respect de cette bonne pratique : l'utilisation d'un code efficace minimise l'utilisation des ressources et améliore les performances.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Il est essentiel d'examiner chaque domaine fonctionnel, y compris le code d'une application conçue dans le cloud, pour optimiser l'utilisation des ressources et les performances. Surveillez en permanence les performances de votre charge de travail dans les environnements de construction et de production et identifiez les possibilités d'améliorer les extraits de code qui utilisent particulièrement bien les ressources. Adoptez un processus de révision régulier pour identifier les bogues ou les anti-modèles dans votre code qui utilisent les ressources de manière inefficace. Exploitez des algorithmes simples et efficaces qui produisent les mêmes résultats pour votre cas d'utilisation.

Étapes d'implémentation

- Utiliser un langage de programmation efficace : utilisez un système d'exploitation et un langage de programmation efficaces pour la charge de travail. Pour en savoir plus sur les langages de programmation économes en énergie (y compris Rust), reportez-vous à Sustainability with Rust.
- Utiliser un compagnon de codage basé sur l'IA : envisagez d'utiliser un compagnon de codage basé sur l'IA tel qu'Amazon Q Developer pour écrire efficacement du code.
- Automatiser les révisions de code : pendant le développement de vos charges de travail, adoptez un processus de révision automatique du code pour améliorer la qualité et identifier les bogues et les anti-modèles.

    - Automatisez les révisions de code avec Amazon CodeGuru Reviewer
    - Détecter les bogues de concurrence avec Amazon CodeGuru
    - Améliorer la qualité du code des applications Python grâce à Amazon CodeGuru
- Utiliser un profileur de code : utilisez un profileur de code pour identifier les sections du code les plus longues ou qui consomment le plus de ressources dans le but de les optimiser.
    - Réduire l'empreinte carbone de votre organisation avec Amazon CodeGuru Profiler
    - Comprendre l'utilisation de la mémoire dans votre application Java avec Amazon CodeGuru Profiler
    - Améliorer l'expérience client et réduire les coûts avec Amazon CodeGuru Profiler
- Surveiller et optimiser : utilisez des ressources de surveillance continue pour identifier les composants nécessitant des ressources élevées ou présentant une configuration sous-optimale.
    - Remplacez les algorithmes à forte intensité de calcul par des versions plus simples et plus efficaces qui produisent le même résultat.
    - Supprimez le code inutile tel que le tri et le formatage.
- Utilisez la refactorisation ou la transformation du code : découvrez les possibilités de transformation du code Amazon Q pour la maintenance et les mises à niveau des applications.
    - Mettez à niveau les versions linguistiques avec Amazon Q Code Transformation
    - AWS re:Invent 2023 - Automate app upgrades & maintenance using Amazon Q Code Transformation

Ressources

Documents connexes :

- Présentation d'Amazon CodeGuru Profiler ?
- Instances FPGA
- Les kits de développement logiciel (SDK) AWS sur les outils pour créer sur AWS

Vidéos connexes :

- Améliorez l'efficacité du code à l'aide d'Amazon CodeGuru Profiler
- Automate Code Reviews and Application Performance Recommendations with Amazon CodeGuru

SUS03-BP04 Optimiser l'impact sur les appareils et équipements

Comprenez les appareils et les équipements utilisés dans votre architecture et employez des stratégies pour réduire leur utilisation. Cela peut minimiser l’impact environnemental global de votre charge de travail dans le cloud.

Anti-modèles courants :

- Vous ignorez l’impact environnemental des appareils utilisés par vos clients.
- Vous gérez et mettez à jour manuellement les ressources utilisées par les clients.

Avantages liés au respect de cette bonne pratique : la mise en œuvre de modèles logiciels et de fonctionnalités optimisés pour les appareils du client peut réduire l’impact environnemental global de la charge de travail dans le cloud.

Niveau de risque encouru si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

La mise en œuvre de modèles et de fonctionnalités logicielles optimisés pour les appareils des clients peut réduire l’impact environnemental de plusieurs façons :

- La mise en œuvre de nouvelles fonctionnalités qui sont rétrocompatibles peut réduire le nombre de remplacements de matériel.
- L’optimisation d’une application pour qu’elle fonctionne efficacement sur les appareils peut contribuer à réduire leur consommation d’énergie et à prolonger leur durée de vie (s’ils sont alimentés par une batterie).
- L’optimisation d’une application pour les appareils peut également réduire le transfert de données sur le réseau.

Comprenez les appareils et les équipements utilisés dans votre architecture, leur cycle de vie prévu et l’impact du remplacement de ces composants. Mettez en œuvre des modèles et des fonctionnalités logicielles qui minimisent la consommation d’énergie de l’appareil, réduisent la nécessité pour les clients de remplacer l’appareil et aussi de le mettre à niveau manuellement.

Étapes d'implémentation

- Réaliser un inventaire : dressez l'inventaire des appareils utilisés dans votre architecture. Les appareils peuvent être des mobiles, des tablettes, IOT des appareils, des lampes intelligentes ou même des appareils intelligents en usine.
- Utilisez des appareils économes en énergie : envisagez d'utiliser des appareils économes en énergie dans votre architecture. Utilisez les configurations de gestion de l'alimentation sur les appareils pour passer en mode faible consommation lorsqu'ils ne sont pas utilisés.
- Exécutez des applications efficaces : optimisez l'exécution de l'application sur les appareils :
  - utilisez des stratégies telles que l'exécution de tâches en arrière-plan pour réduire leur consommation d'énergie.
  - Prenez en compte la bande passante et la latence du réseau lorsque vous créez des charges utiles et intégrez des capacités qui aident vos applications à fonctionner correctement sur des liens à faible bande passante et à latence élevée.
  - Convertissez les charges utiles et les fichiers dans les formats optimisés requis par les appareils. Par exemple, vous pouvez utiliser Amazon Elastic Transcoder ou AWS Elemental MediaConvert pour convertir des fichiers multimédias numériques volumineux haute qualité en formats que les utilisateurs peuvent lire sur des appareils mobiles, des tablettes, des navigateurs Web et des téléviseurs connectés.
  - Réalisez des activités gourmandes en calcul côté serveur (comme le rendu d'images) ou utilisez le streaming d'applications pour améliorer l'expérience utilisateur sur des appareils plus anciens.
  - Segmentez et paginez la sortie, en particulier, pour les séances interactives, afin de gérer les charges utiles et limiter les exigences en matière de stockage local.
- Impliquer les fournisseurs : collaborez avec des fournisseurs d'appareils qui utilisent des matériaux durables et assurent la transparence de leurs chaînes d'approvisionnement et de leurs certifications environnementales.
- Utiliser les mises à jour over-the-air (OTA) : utilisez le mécanisme automatisé over-the-air (OTA) pour déployer les mises à jour sur un ou plusieurs appareils.
  - Vous pouvez utiliser un pipeline CI/CD pour mettre à jour les applications mobiles.
  - Vous pouvez utiliser AWS IoT Device Management pour gérer à distance les appareils connectés à grande échelle.
- Utiliser des parcs d'appareils gérés : pour tester les nouvelles fonctionnalités et les mises à jour, utilisez Device Farm avec des ensembles représentatifs de matériel et itérez le développement

pour maximiser les dispositifs pris en charge. Pour en savoir plus, consultez SUS06-BP05 Utiliser des tests Device Farms gérés.

- Continuer à surveiller et à améliorer : suivez la consommation d'énergie des appareils pour identifier les domaines à améliorer. Utilisez les nouvelles technologies ou les bonnes pratiques pour améliorer les impacts environnementaux de ces appareils.

Ressources

Documents connexes :

- Qu'est-ce que c'est AWS Device Farm ?
- AppStream 2.0 Documentation
- NICE DCV
- OTAtutoriel pour la mise à jour du firmware sur les appareils fonctionnant sous Free RTOS
- Optimisation de vos appareils IoT pour la durabilité environnementale

Vidéos connexes :

- AWS re:Invent 2023 - Améliorez la qualité de vos applications mobiles et Web en utilisant AWS Device Farm

SUS03-BP05 Utiliser des modèles logiciels et des architectures qui soutiennent au mieux l'accès aux données et les modèles de stockage.

Comprenez comment les données sont utilisées au sein de votre charge de travail, et comment elles sont consommées par vos utilisateurs, transférées et stockées. Utilisez des modèles et des architectures logicielles qui prennent le mieux en charge l'accès et le stockage des données afin de minimiser les ressources de calcul, de mise en réseau et de stockage nécessaires pour supporter la charge de travail.

Anti-modèles courants :

- Vous partez du principe que toutes les charges de travail ont des modèles de stockage de données et d'accès similaires.
- Vous n'utilisez qu'un seul niveau de stockage, partant du principe que toutes les charges de travail s'intègrent dans ce niveau.

- Vous partez du principe que les modèles d'accès aux données n'évolueront pas dans le temps.
- Votre architecture prend en charge un potentiel pic important d'accès aux données, ce qui fait que les ressources restent inactives la plupart du temps.

Avantages liés au respect de cette bonne pratique : la sélection et l'optimisation de votre architecture en fonction des modèles d'accès aux données et de stockage contribueront à réduire la complexité du développement et à augmenter l'utilisation globale. Savoir quand utiliser les tables globales, le partitionnement des données et la mise en cache vous aidera à réduire les frais généraux opérationnels et à vous mettre à l'échelle en fonction des besoins de votre charge de travail.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Pour améliorer la durabilité à long terme de la charge de travail, utilisez des modèles d'architecture qui prennent en charge les caractéristiques d'accès aux données et de stockage pour votre charge de travail. Ces modèles vous aident à récupérer et à traiter efficacement les données. Par exemple, vous pouvez utiliser une architecture de données moderne sur AWS avec des services spécifiques optimisés pour vos cas d'utilisation analytiques uniques. Ces modèles d'architecture permettent un traitement efficace des données et réduisent l'utilisation des ressources.

Étapes d'implémentation

- Comprendre les caractéristiques des données : analysez les caractéristiques de vos données et les modèles d'accès afin d'identifier la bonne configuration pour vos ressources cloud. Les caractéristiques clés à prendre en considération sont les suivantes :
  - Type de données : structurées, semi-structurées, non structurées
  - Croissance des données : limitée, illimitée
  - Durabilité des données : persistantes, éphémères, temporaires
  - Modèles d'accès : lectures ou écritures, fréquence de mise à jour, irrégulière ou cohérente.
- Utiliser les modèles d'architecture optimaux : utilisez les modèles d'architecture qui prennent le mieux en charge les modèles d'accès et de stockage des données.
  - Modèles permettant la persistance des données
  - Let's Architect! Architectures de données modernes
  - Bases de données AWS : l'outil adapté à la tâche
- Utiliser des services spécialement conçus : utilisez des technologies adaptées à vos besoins.

  - Utilisez des technologies qui peuvent fonctionner en natif avec les données compressées.
    - Formats de fichiers pour prendre en charge la compression Athena
    - Options de formatage pour les entrées et sorties ETL dans AWS Glue
    - Chargement de fichiers de données comprimés d'Amazon S3 avec Amazon Redshift
  - Utilisez des services d'analytique spécialement conçus pour le traitement des données dans votre architecture. Pour en savoir plus sur les services analytiques de AWS, reportez-vous à AWS re:Invent 2022 - Building modern data architectures on AWS.
  - Utilisez le moteur de base de données qui prend le mieux en charge votre modèle de requête dominant. Gérez vos index de base de données pour une bonne efficacité des requêtes. Pour en savoir plus, reportez-vous à Bases de données AWS et AWS re:Invent 2022 - Modernize apps with purpose-built databases.
- Minimiser le transfert des données : sélectionnez des protocoles réseaux qui réduisent la quantité de capacité réseau consommée dans votre architecture.

Ressources

Documents connexes :

- COPIE de formats de données en colonnes avec Amazon Redshift
- Conversion de votre format d'enregistrement d'entrée dans Firehose
- Améliorer la performance des requêtes sur Amazon Athena grâce à une conversion en formats de colonnes
- Surveillance de la charge de base de données avec Performance Insights sur Amazon Aurora
- Surveillance de la charge de la base de données avec Performance Insights sur Amazon RDS
- Classe de stockage Amazon S3 Intelligent-Tiering
- Créez un magasin d'événements CQRS avec Amazon DynamoDB

Vidéos connexes :

- AWS re:Invent 2022 - Building data mesh architectures on AWS
- AWS re:Invent 2023 - Deep dive into Amazon Aurora and its innovations
- AWS re:Invent 2023 - Improve Amazon EBS efficiency and be more cost-efficient
- AWS re:Invent 2023 - Optimizing storage price and performance with Amazon S3

- AWSre:Invent 2023 : création et optimisation d'un lac de données sur Amazon S3
- AWS re:Invent 2023 - Advanced event-driven patterns with Amazon EventBridge

Exemples connexes :

- AWS Atelier sur les bases de données sur mesure
- Journée d'immersion dans l'architecture moderne des données AWS
- Créez un maillage de données sur AWS

Gestion des données

Mettez en œuvre des pratiques de gestion des données afin de réduire le stockage alloué nécessaire pour assurer votre charge de travail et les ressources nécessaires à son utilisation. Veillez à bien connaître vos données et utilisez des technologies et des configurations de stockage qui soutiennent plus efficacement la valeur métier des données et leur utilisation. Adoptez un cycle de vie des données offrant un stockage plus efficace et moins performant quand les exigences baissent et supprimez les données qui ne sont plus nécessaires.

Bonnes pratiques

- SUS04-BP01 Mettre en œuvre une politique de classification des données
- SUS04-BP02 Utiliser des technologies qui prennent en charge les modèles d'accès et de stockage des données
- SUS04-BP03 Utiliser des politiques pour gérer le cycle de vie de vos ensembles de données
- SUS04-BP04 Utiliser l'élasticité et l'automatisation pour étendre le stockage par blocs ou le système de fichiers
- SUS04-BP05 Supprimer les données inutiles ou redondantes
- SUS04-BP06 Utiliser des systèmes de fichiers partagés ou le stockage pour accéder aux données courantes
- SUS04-BP07 Réduire le mouvement des données entre les réseaux
- SUS04-BP08 Sauvegarder des données uniquement lorsqu'elles sont difficiles à recréer

SUS04-BP01 Mettre en œuvre une politique de classification des données

Classifiez les données pour identifier leur criticité vis-à-vis des résultats économiques, et choisissez le niveau de stockage économe en énergie approprié pour stocker les données.

Anti-modèles courants :

- Vous n'identifiez pas les ressources de données actuellement traitées ou stockées ayant des caractéristiques similaires (comme la sensibilité, la criticité métier ou les exigences réglementaires).
- Vous n'avez pas implémenté de catalogue de données pour inventorier vos ressources de données.

Avantages liés au respect de cette bonne pratique : la mise en œuvre d'une politique de classification des données vous permet de déterminer le niveau de stockage le plus économe en énergie pour les données.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

La classification des données implique d'identifier les types de données actuellement traitées ou stockées dans un système d'information détenu ou exploité par une organisation. Elle implique également de déterminer la criticité des données et l'impact possible d'une compromission, d'une perte ou d'une mauvaise utilisation de ces données.

Mettez en œuvre la politique de classification des données en partant de l'utilisation contextuelle des données et en créant un schéma de catégorisation qui prend en compte le niveau de criticité d'un jeu de données déterminé vis-à-vis des opérations d'une organisation.

Étapes d'implémentation

- Réaliser un inventaire des données :procédez à l'inventaire des différents types de données qui existent pour votre charge de travail.
- Données du groupe :déterminez la criticité, la confidentialité, l'intégrité et la disponibilité des données en fonction du risque vis-à-vis de l'organisation. Prenez en compte ces exigences pour regrouper les données dans l'un des niveaux de classification des données que vous adoptez. À titre d'exemple, consultez Quatre étapes simples pour classer vos données et sécuriser votre start-up.

- Définissez les niveaux et les politiques de classification des données : pour chaque groupe de données, définissez le niveau de classification des données (par exemple, public ou confidentiel) et les politiques de gestion. Balisez les données en conséquence. Pour en savoir plus sur les catégories de classification des données, consultez le livre blanc Classification des données.
- Révision périodique : passez régulièrement en revue et auditez votre environnement pour détecter les données non étiquetées et non classifiées. Utilisez l'automatisation pour identifier ces données, puis classez et balisez les données de manière appropriée. À titre d'exemple, consultez le catalogue de données et les robots dans AWS Glue.
- Établissez un catalogue de données : établissez un catalogue de données qui fournit des fonctionnalités d'audit et de gouvernance.
- Documentation : Documentez les politiques de classification des données et les procédures de traitement pour chaque classe de données.

Ressources

Documents connexes :

- Utilisation de l'effet de levier AWS Cloud tpour soutenir la classification des données
- Marquer les politiques de AWS Organizations

Vidéos connexes :

- AWS re:Invent 2022 - Favoriser l'agilité grâce à la gouvernance des données activée AWS
- AWS re:Invent 2023 - Protection des données et résilience grâce au stockage AWS

SUS04-BP02 Utiliser des technologies qui prennent en charge les modèles d'accès et de stockage des données

Utilisez les technologies de stockage qui prennent le mieux en charge l'accès à vos données et leur stockage pour limiter le provisionnement de ressources tout en soutenant votre charge de travail.

Anti-modèles courants :

- Vous partez du principe que toutes les charges de travail ont des modèles de stockage de données et d'accès similaires.

- Vous n’utilisez qu’un seul niveau de stockage, partant du principe que toutes les charges de travail s’intègrent dans ce niveau.
- Vous partez du principe que les modèles d’accès aux données n’évolueront pas dans le temps.

Avantages liés au respect de cette bonne pratique : en choisissant et en optimisant vos technologies de stockage en fonction des modèles d’accès aux données et de stockage, vos besoins métier demanderont moins de ressources cloud et vous améliorerez l’efficace globale de votre charge de travail cloud.

Niveau d’exposition au risque si cette bonne pratique n’est pas respectée : faible

Directives d’implémentation

Choisissez la solution de stockage la mieux adaptée à vos modèles d’accès ou envisagez de changer vos modèles d’accès en fonction de la solution de stockage pour optimiser les performances.

Étapes d’implémentation

- Évaluer les caractéristiques des données et de l’accès : évaluez les caractéristiques de vos données et le modèle d’accès afin de recueillir les caractéristiques clés de vos besoins en matière de stockage. Les caractéristiques clés à prendre en considération sont les suivantes :
  - Type de données : structurées, semi-structurées, non structurées
  - Croissance des données : limitée, illimitée
  - Durabilité des données : persistantes, éphémères, temporaires
  - Modèles d’accès : lectures ou écritures, fréquence de mise à jour, irrégulière ou cohérente.
- Choisir la bonne technologie de stockage : migrer les données vers la technologie de stockage appropriée qui prend en charge les caractéristiques des données et le modèle d’accès. Voici quelques exemples de technologies de AWS stockage et leurs principales caractéristiques :

Type	Technologie	Principales caractéristiques
Stockage d’objets	Amazon S3	Service de stockage d’objets offrant une capacité de mise à l’échelle illimitée, une haute disponibilité et plusieurs options d’accessi bilité. Le transfert et l’accès

Type	Technologie	Principales caractéristiques
		à des objets à l'intérieur et à l'extérieur d'Amazon S3 peuvent utiliser un service, tel que Transfer Acceleration ou Access Points, pour répondre à votre localisation, à vos besoins en matière de sécurité et à vos modèles d'accès.
Archivage et stockage	Amazon S3 Glacier	Classe de stockage d'Amazon S3 conçue pour l'archivage de données.
Système de fichiers partagé	Amazon Elastic File System (AmazonEFS)	Système de fichiers montable auquel plusieurs types de solutions informatiques peuvent accéder. Amazon augmente et réduit EFS automatiquement le stockage et optimise les performances pour garantir de faibles latences constantes.

Type	Technologie	Principales caractéristiques
Système de fichiers partagé	Amazon FSx	Construit sur les dernières solutions AWS informati ques pour prendre en charge quatre systèmes de fichiers couramment utilisés : Open NetApp ONTAPZFS, Windows File Server et Lustre. FSxLa latence, le débit et le débit d'Amazon IOPS varient selon le système de fichiers et doivent être pris en compte lors de la sélection du système de fichiers adapté à vos besoins en matière de charge de travail.
Stockage en mode bloc	Boutique Amazon Elastic Block (AmazonEBS)	Service de stockage par blocs évolutif et performant conçu pour Amazon Elastic Compute Cloud (AmazonEC 2). Amazon EBS inclut le stockage SSD sauvegardé pour les charges de travail transactionnelles intensives et le stockage HDD sauvegard é pour les charges de travail IOPS gourmandes en débit.

Type	Technologie	Principales caractéristiques
Base de données relationnelle	Amazon Aurora, Amazon RDS, Amazon Redshift	Conçu pour prendre en charge ACID (atomicité, cohérence, isolation, durabilité) les transactions et maintenir l'intégrité référentielle et la forte cohérence des données. De nombreuses applications traditionnelles, systèmes de planification des ressources d'entreprise (ERP), de gestion de la relation client (CRM) et de commerce électronique utilisent des bases de données relationnelles pour stocker leurs données.
Base de données clé-valeur	Amazon DynamoDB	Optimisées pour les modèles d’accès courants, généralement pour stocker et récupérer de gros volumes de données. Les applications Web à trafic élevé, les systèmes d’e-commerce et les applications de jeu sont des cas d’utilisation typiques pour les bases de données clé-valeur.

- Automatisez l'allocation de stockage : pour les systèmes de stockage de taille fixe, tels qu'Amazon EBS ou AmazonFSx, surveillez l'espace de stockage disponible et automatisez l'allocation de stockage lorsqu'un seuil est atteint. Vous pouvez tirer parti CloudWatch d'Amazon pour collecter et analyser différentes statistiques pour Amazon EBS et Amazon FSx.
- Choisissez la bonne classe de stockage : choisissez la classe de stockage adaptée à vos données.
  - Les classes de stockage Amazon S3 peuvent être configurées au niveau de l’objet. Un compartiment unique peut contenir les objets stockés dans toutes les classes de stockage.

- Vous pouvez utiliser les stratégies de cycle de vie Amazon S3 pour faire passer automatiquement des objets d'une classe de stockage à une autre ou supprimer des données sans aucune modification au niveau de l'application. Ces mécanismes de stockage vous imposent généralement de faire un compromis entre l'efficacité des ressources, la latence d'accès et la fiabilité.

Ressources

Documents connexes :

- Types de EBS volumes Amazon
- Boutique d'EC2instances Amazon
- Amazon S3 Intelligent Tiering
- Caractéristiques d'Amazon EBS I/O
- Utilisation des classes de stockage Amazon S3
- Qu'est-ce qu'Amazon S3 Glacier ?

Vidéos connexes :

- AWS re:Invent 2023 - Améliorez EBS l'efficacité d'Amazon et soyez plus rentable
- AWS re:Invent 2023 - Optimisation du prix et des performances du stockage avec Amazon S3
- AWS re:Invent 2023 - Création et optimisation d'un lac de données sur Amazon S3
- AWS re:Invent 2022 - Création d'architectures de données modernes sur AWS
- AWS re:Invent 2022 - Modernisez les applications avec des bases de données spécialement conçues
- AWS re:Invent 2022 - Création d'architectures de maillage de données sur AWS
- AWS re:Invent 2023 - Découvrez Amazon Aurora et ses innovations
- AWS re:Invent 2023 - Modélisation avancée des données avec Amazon DynamoDB

Exemples connexes :

- Exemples Amazon S3
- AWS Atelier sur les bases de données spécialement conçues

- Bases de données pour développeurs
- AWS Journée d'immersion dans l'architecture de données moderne
- Créez un maillage de données sur AWS

SUS04-BP03 Utiliser des politiques pour gérer le cycle de vie de vos ensembles de données

Gérez le cycle de vie de toutes vos données et appliquez automatiquement la suppression pour réduire au minimum le stockage total requis pour votre charge de travail.

Anti-modèles courants :

- Vous supprimez manuellement les données.
- Vous ne supprimez aucune donnée de vos charges de travail.
- Vous ne déplacez pas les données vers des niveaux de stockage plus écoénergétiques en fonction de leurs exigences de conservation et d’accès.

Avantages liés au respect de cette bonne pratique : l’utilisation de politiques de cycle de vie des données garantit un accès et une rétention efficaces des données dans une charge de travail.

Niveau de risque encouru si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

Les exigences en matière de conservation et d’accès des jeux de données varient généralement au cours de leur cycle de vie. Par exemple, votre application peut nécessiter un accès fréquent à certains jeux de données pendant une période limitée. Après cela, ces jeux de données sont rarement consultés. Pour améliorer l’efficacité du stockage de données et du calcul au fil du temps, mettez en œuvre des politiques de cycle de vie, qui sont des règles qui définissent la manière dont les données sont traitées au fil du temps.

Avec les règles de configuration du cycle de vie, vous pouvez demander au service de stockage spécifique de transférer un jeu de données vers des niveaux de stockage plus écoénergétiques, de l’archiver ou de le supprimer. Cette pratique minimise le stockage et l’extraction actifs des données, ce qui entraîne une réduction de la consommation d’énergie. En outre, des pratiques telles que l’archivage ou la suppression de données obsolètes soutiennent la conformité réglementaire et la gouvernance des données.

Étapes d'implémentation

- Utiliser la classification des données : classez les jeux de données dans votre charge de travail.
- Définir des règles de traitement : définissez des procédures de traitement pour chaque classe de données.
- Activer l'automatisation : définissez des politiques de cycle de vie automatisées pour appliquer des règles de cycle de vie. Voici quelques exemples de la configuration des politiques de cycle de vie automatisé pour différents services de stockage AWS :

Service de stockage	Comment définir des politiques de cycle de vie automatisées
Amazon S3	Vous pouvez utiliser Amazon S3 Lifecycle afin de gérer vos objets au cours de leur cycle de vie. Si vos schémas d'accès sont inconnus, changeants ou imprévisibles, vous pouvez utiliser Amazon S3 Intelligent-Tiering, qui surveille les schémas d'accès et déplace automatiquement les objets qui n'ont pas été accédés vers des niveaux d'accès moins coûteux. Vous pouvez tirer parti des métriques Amazon S3 Storage Lens pour identifier les opportunités d'optimisation et les lacunes dans la gestion du cycle de vie.
Amazon Elastic Block Store	Vous pouvez utiliser Amazon Data Lifecycle Manager pour automatiser la création, la rétention et la suppression des instantanés EBS et des AMI basées sur EBS.
Amazon Elastic File System	La fonction de gestion du cycle de vie Amazon EFS gère automatiquement le stockage de vos systèmes de fichiers.
Amazon Elastic Container Registry	Les politiques de cycle de vie d'Amazon ECR automatisent le nettoyage des images de

Service de stockage	Comment définir des politiques de cycle de vie automatisées
	conteneur en faisant expirer des images selon l'ancienneté ou le décompte.
AWS Elemental MediaStore	Vous pouvez utiliser une politique de cycle de vie des objets qui régit la durée de stockage des objets dans le conteneur MediaStore.

- Supprimer les ressources inutilisées : supprimez les volumes, les instantanés et les données inutilisés dont la période de conservation est dépassée. Utilisez des fonctionnalités de service natives telles qu'Amazon DynamoDB Time To Live ou la conservation des journaux Amazon CloudWatch pour la suppression.
- Regrouper et compresser : regroupez et compressez les données le cas échéant en fonction des règles de cycle de vie.

Ressources

Documents connexes :

- Optimisez vos règles de cycle de vie Amazon S3 grâce à l'analyse des classes de stockage Amazon S3
- Évaluation des ressources avec AWS Config Rules

Vidéos connexes :

- AWS re:Invent 2021 - Amazon S3 Lifecycle best practices to optimize your storage spend
- AWS re:Invent 2023 - Optimizing storage price and performance with Amazon S3
- Simplifiez le cycle de vie de vos données et optimisez les coûts de stockage avec Amazon S3 Lifecycle
- Réduisez vos coûts de stockage en utilisant Amazon S3 Storage Lens

SUS04-BP04 Utiliser l'élasticité et l'automatisation pour étendre le stockage par blocs ou le système de fichiers

Utilisez l’élasticité et l’automatisation pour étendre le stockage par blocs ou le système de fichiers au fur et à mesure que le volume de données augmente afin de minimiser le stockage total provisionné.

Anti-modèles courants :

- Vous provisionnez un grand bloc de stockage ou un grand système de fichiers pour vos besoins futurs.
- Vous surprovisionnez les opérations d'entrée et de sortie par seconde (IOPS) de votre système de fichiers.
- Vous ne contrôlez pas l’utilisation de vos volumes de données.

AAvantages liés au respect de cette bonne pratique : la réduction du surprovisionnement du système de stockage réduit les ressources inactives et améliore l’efficacité globale de votre charge de travail.

Niveau de risque encouru si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

Créez des systèmes de stockage par blocs et des systèmes de fichiers avec une allocation de taille, un débit et une latence adaptés à votre charge de travail. Utilisez l’élasticité et l’automatisation pour étendre le stockage par blocs ou le système de fichiers en fonction de la croissance des données sans avoir à provisionner ces services de stockage de manière excessive.

Étapes d’implémentation

- Pour le stockage de taille fixe tel qu'Amazon EBS, vérifiez que vous surveillez la quantité de stockage utilisée par rapport à la taille de stockage globale et créez une automatisation, si possible, pour augmenter la taille de stockage lorsque vous atteignez un seuil.
- Utilisez des volumes Elastic et des services de données par bloc gérés pour automatiser l’allocation de stockage supplémentaire à mesure que vos données persistantes augmentent. Par exemple, vous pouvez utiliser Amazon EBS Elastic Volumes pour modifier la taille ou le type de volume ou ajuster les performances de vos EBS volumes Amazon.
- Choisissez la bonne classe de stockage, le bon mode de performance et le mode de débit adapté à votre système de fichiers afin de répondre aux besoins de votre entreprise, sans les dépasser.

  - EFSPerformances d'Amazon
  - Performances des EBS volumes Amazon sur les instances Linux
- Définissez des niveaux cibles d’utilisation des volumes de données et redimensionnez les volumes en dehors des plages attendues.
- Dimensionnez correctement les volumes en lecture seule en fonction des données.
- Migrez les données vers des magasins d’objets pour éviter d’allouer la capacité excédentaire des tailles de volume fixes vers le stockage par bloc.
- Examinez régulièrement les volumes Elastic et les systèmes de fichiers pour mettre fin aux volumes inutilisés et réduire les ressources surprovisionnées pour les adapter à la taille actuelle des données.

Ressources

Documents connexes :

- Étendre le système de fichiers après le redimensionnement d'un volume EBS
- Modifier un volume à l'aide d'Amazon EBS Elastic Volumes
- Documentation Amazon FSx
- Qu’est ce qu’Amazon Elastic File System ?

Vidéos connexes :

- Présentation approfondie d'Amazon EBS Elastic Volumes
- Stratégies d'optimisation d'Amazon EBS et de Snapshot pour de meilleures performances et des économies
- Optimisation d'Amazon EFS en termes de coûts et de performances, en utilisant les meilleures pratiques

SUS04-BP05 Supprimer les données inutiles ou redondantes

Supprimez les données inutiles ou redondantes pour minimiser les ressources de stockage requises pour stocker vos jeux de données.

Anti-modèles courants :

- Vous dupliquez des données qui peuvent être facilement obtenues ou recréées.
- Vous sauvegardez toutes les données sans tenir compte de leur criticité.
- Vous ne supprimez les données que de façon irrégulière, sur les événements opérationnels ou pas du tout.
- Vous stockez les données de manière redondante, quelle que soit la durabilité du service de stockage.
- Vous activez la gestion des versions sans aucune justification commerciale.

Avantages liés au respect de cette bonne pratique : la suppression des données inutiles réduit la taille de stockage requise pour votre charge de travail et son impact environnemental.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

La suppression des jeux de données inutiles et redondants permet de réduire les coûts de stockage et l'empreinte environnementale. Cette pratique peut également rendre l'informatique plus efficace, car les ressources de calcul traitent uniquement des données importantes au lieu de données inutiles. Automatisez la suppression des données inutiles. Utilisez des technologies qui dédupliquent les données au niveau du fichier et du bloc. Utilisez les fonctionnalités des services pour la réplication et la redondance des données natives.

Étapes d'implémentation

- Évaluer les jeux de données publics : déterminez si vous pouvez éviter de stocker des données en utilisant des jeux de données existants publiquement accessibles dans AWS Data Exchange et les données ouvertes sur AWS.
- Dédupliquer les données : utilisez des mécanismes qui peuvent dédupliquer les données au niveau du bloc et de l'objet. Voici quelques exemples de déduplication des données sur AWS :

Service de stockage	Mécanismes de déduplication
Amazon S3	Utilisez AWS Lake Formation FindMatches pour rechercher les enregistrements correspondants dans un jeu de données (y compris ceux sans identifiant) à l'aide de la nouvelle transformation FindMatches ML.

Service de stockage	Mécanismes de déduplication
Amazon FSx	Utilisez la déduplication des données sur Amazon FSx for Windows
Instantanés volumes Amazon Elastic Block Store	Les instantanés sont des sauvegardes incrémentielles, ce qui signifie que seuls les blocs de l'appareil qui ont changé depuis l'instantané le plus récent sont enregistrés.

- Utiliser des politiques de cycle de vie : utilisez des politiques de cycle de vie pour automatiser la suppression des données inutiles. Tirez parti de fonctionnalités de service natives telles qu'Amazon DynamoDB Time To Live, Amazon S3 Lifecycle ou la rétention des journaux Amazon CloudWatch pour la suppression.
- Utiliser la virtualisation des données : utilisez les capacités de virtualisation des données sur AWS afin de maintenir les données à leur source et d'éviter leur duplication.
  - Virtualisation des données natives dand le cloud sur AWS
  - Optimiser le modèle de données à l'aide du partage de données Amazon Redshift
- Utiliser la sauvegarde incrémentielle : utilisez une technologie de sauvegarde capable d'effectuer des sauvegardes incrémentielles.
- Utiliser la durabilité native : tirez parti de la durabilité d'Amazon S3 et de la réplication d'Amazon EBS pour atteindre vos objectifs de durabilité au lieu de recourir à des technologies autogérées (telles qu'un réseau redondant de disques indépendants (RAID)).
- Utiliser une journalisation efficace : centralisez les données de journaux et de suivi, dédupliquez les entrées de journaux identiques et établissez des mécanismes pour ajuster le niveau de détail, si nécessaire.
- Utiliser une mise en cache efficace : préremplissez les caches uniquement lorsque cela est justifié.
- Établissez la surveillance et l'automatisation des caches pour redimensionner correctement les caches.
- Supprimer les ressources de version antérieure : supprimez les déploiements et les ressources obsolètes des magasins d'objets et des caches périphériques lors de la transmission des nouvelles versions de votre charge de travail.

Ressources

Documents connexes :

- Modification de la conservation des données de journaux dans CloudWatch Logs
- Déduplication des données sur Amazon FSx for Windows File Server
- Les fonctions d'Amazon FSx for ONTAP incluent la déduplication des données
- Invalider des fichiers sur Amazon CloudFront
- Utilisation d'AWS Backup pour la sauvegarde et la restauration des systèmes de fichiers Amazon EFS
- Qu'est-ce qu'Amazon CloudWatch Logs ?
- Utilisation de sauvegardes sur Amazon RDS
- Intégrez et dédupliquez des ensembles de données à l'aide de AWS Lake Formation

Vidéos connexes :

- Cas d'utilisation du partage de données pour Amazon Redshift

Exemples connexes :

- Comment analyser les journaux d'accès au serveur Amazon S3 à l'aide d'Amazon Athena ?

SUS04-BP06 Utiliser des systèmes de fichiers partagés ou le stockage pour accéder aux données courantes

Adoptez des systèmes de fichiers ou de stockage partagés pour éviter la duplication des données et permettre une infrastructure plus efficace pour votre charge de travail.

Anti-modèles courants :

- Vous mettez en service le stockage pour chaque client individuel.
- Vous ne détachez pas le volume de données des clients inactifs.
- Vous ne fournissez pas d'accès au stockage pour les plateformes et les systèmes.

Avantages liés au respect de cette bonne pratique : ’utilisation de systèmes de fichiers ou de stockage partagés permet de partager des données avec un ou plusieurs consommateurs sans avoir à les copier. Cela permet de réduire les ressources de stockage nécessaires à la charge de travail.

Niveau de risque encouru si cette bonne pratique n’est pas respectée : moyen

Directives d’implémentation

Si plusieurs utilisateurs ou applications accèdent aux mêmes jeux de données, l’utilisation de la technologie de stockage partagé est cruciale pour avoir une infrastructure efficace pour votre charge de travail. La technologie de stockage partagé fournit un emplacement central pour stocker et gérer les jeux de données et éviter la duplication des données. Elle assure également la cohérence des données entre les différents systèmes. En outre, la technologie de stockage partagé permet d’utiliser plus efficacement la puissance de calcul, car plusieurs ressources informatiques peuvent accéder aux données et les traiter simultanément en parallèle.

Ne récupérez les données de ces services de stockage partagé qu’en fonction des besoins et détachez les volumes inutilisés pour libérer des ressources.

Étapes d’implémentation

- Utiliser un stockage partagé : migrez les données vers le stockage partagé lorsque les données ont plusieurs consommateurs. Voici quelques exemples de technologie de stockage partagé sur AWS :

Option de stockage	Utilisation
Amazon EBS Multi-Attach	Amazon EBS Multi-Attach vous permet d’attacher un volume SSD IOPS provisionnés (io1 ou io2) à plusieurs instances basées sur Nitro situées dans la même zone de disponibi lité.
Amazon EFS	Reportez-vous à Quand choisir Amazon EFS.
Amazon FSx	Reportez-vous à Choisir un système de fichiers Amazon FSx.
Amazon S3	Les applications qui ne nécessitent pas de structure de système de fichiers et qui sont conçues pour fonctionner avec le stockage

Option de stockage	Utilisation
	d'objet peuvent utiliser Amazon S3 comme une solution de stockage d'objet massivement évolutive, durable et peu coûteuse.

- Récupérer les données requises : copiez des données vers ou récupérez des données depuis des systèmes de fichiers partagés uniquement si nécessaire. Par exemple, vous pouvez créer un système de fichiers Amazon FSx pour Lustre soutenu par Amazon S3 et charger uniquement le sous-ensemble de données requis pour le traitement des tâches sur Amazon FSx.
- Supprimer les données inutiles : supprimez les données selon vos modèles d'utilisation comme indiqué dans SUS04-BP03 Utiliser des politiques pour gérer le cycle de vie de vos ensembles de données.
- Détacher les clients inactifs : détachez les volumes des clients qui ne les utilisent pas activement.

Ressources

Documents connexes :

- Liaison de votre système de fichiers à un compartiment Amazon S3
- Utilisation d'Amazon EFS pour AWS Lambda dans vos applications sans serveur
- Amazon EFS Intelligent-Tiering optimise les coûts liés aux charges de travail en fonction de l'évolution des modèles d'accès
- Utilisation d'Amazon FSx avec votre référentiel de données sur site

Vidéos connexes :

- Optimisation des coûts de stockage avec Amazon EFS
- AWS re:Invent 2023 - What's new with AWS file storage
- AWS re:Invent 2023 - File storage for builders and data scientists on Amazon Elastic File System

SUS04-BP07 Réduire le mouvement des données entre les réseaux

Utilisez des systèmes de fichiers partagés ou un stockage objet pour accéder aux données communes et minimiser les ressources réseau totales requises pour prendre en charge le déplacement des données de votre charge de travail.

Anti-modèles courants :

- Vous stockez toutes les données dans la même Région AWS, indépendamment de l'endroit où se trouvent les utilisateurs des données.
- Vous n'optimisez ni la taille ni le format des données avant de les déplacer sur le réseau.

Avantages liés au respect de cette bonne pratique : l'optimisation du déplacement des données sur le réseau réduit les ressources réseau totales nécessaires à la charge de travail et diminue son impact environnemental.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Le déplacement des données dans votre entreprise nécessite des ressources de calcul, de réseau et de stockage. Utilisez des techniques pour minimiser les déplacements de données et améliorer l'efficacité globale de votre charge de travail.

Étapes d'implémentation

- Utiliser la proximité : tenez compte de la proximité des données ou des utilisateurs comme facteur de décision lors de la sélection d'une région pour votre charge de travail.
- Partitionner les services : partitionnez les services consommés par région afin que les données spécifiques à une région soient stockées dans la région où elles sont consommées.
- Utiliser des formats de fichiers efficaces : utilisez des formats de fichiers efficaces (tels que Parquet ou ORC) et compressez les données avant de les déplacer sur le réseau.
- Minimiser le mouvement des données : ne déplacez pas les données inutilisées. Voici quelques exemples qui peuvent vous aider à éviter de déplacer des données inutilisées :
  - Réduisez les réponses de l'API aux seules données pertinentes.
  - Agrégez les données lorsqu'elles sont détaillées (les informations au niveau de l'enregistrement ne sont pas requises).
  - Reportez-vous à Atelier Well-Architected : optimiser le modèle de données à l'aide du partage de données Amazon Redshift.
  - Envisagez le partage de données entre comptes dans AWS Lake Formation.
- Utilisez des services de périphérie : utilisez des services qui peuvent vous aider à exécuter du code au plus près des utilisateurs de votre charge de travail.

Service	Utilisation
Lambda@Edge	Utilisez ce service pour les opérations exigeantes en puissance de calcul qui sont exécutées lorsque des objets ne sont pas dans le cache.
Fonctions CloudFront	Utilisez ce système pour des cas d'utilisa tion simples tels que les manipulations de requêtes/réponses HTTP(s) qui peuvent être lancées par des fonctions de courte durée.
AWS IoT Greengrass	Exécutez du calcul local, une messagerie et une mise en cache de données pour les appareils connectés.

Ressources

Documents connexes :

- Optimisation de votre infrastructure AWS pour la durabilité, partie 3 : mise en réseau
- Infrastructure mondiale AWS
- Fonctions clés d'Amazon CloudFront, y compris le réseau périphérique mondial CloudFront
- Compression des requêtes HTTP dans Amazon OpenSearch Service
- Compression intermédiaire de données avec Amazon EMR
- Chargement de fichiers de données comprimés d'Amazon S3 vers Amazon Redshift
- Diffusion de fichiers compressés avec Amazon CloudFront

Vidéos connexes :

- Demystifying data transfer on AWS

SUS04-BP08 Sauvegarder des données uniquement lorsqu'elles sont difficiles à recréer

Évitez de sauvegarder les données qui n'ont aucune valeur commerciale afin de minimiser les besoins en ressources de stockage pour votre charge de travail.

Anti-modèles courants :

- Vous n'avez aucune stratégie de sauvegarde en place pour vos données.
- Vous sauvegardez des données qui peuvent être facilement recréées.

Avantages liés au respect de cette bonne pratique : éviter de sauvegarder des données non critiques réduit les ressources de stockage requises pour la charge de travail et réduit son impact environnemental.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Le fait d'éviter la sauvegarde de données inutiles peut contribuer à réduire les coûts et les ressources de stockage utilisées par la charge de travail. Sauvegardez uniquement les données ayant une valeur opérationnelle ou nécessaires pour répondre aux exigences en matière de conformité. Examinez les politiques de sauvegarde et excluez tout magasin éphémère n'apportant aucune valeur dans un scénario de récupération.

Étapes d'implémentation

- Classer les données : mettre en œuvre la politique de classification des données telle que décrite dans SUS04-BP01 Mettre en œuvre une politique de classification des données.
- Concevoir une stratégie de sauvegarde : tirez parti de l'importance de votre classification des données et concevez une stratégie de sauvegarde en fonction de vos objectif de délai de reprise (RTO) et objectif de point de reprise (RPO). Évitez de sauvegarder les données non critiques.
  - Excluez les données qui peuvent être facilement recréées.
  - Excluez les données éphémères de vos sauvegardes.
  - Excluez les copies locales des données, sauf si le temps nécessaire pour restaurer ces données à partir d'un emplacement commun dépasse vos contrats de niveau de service (SLA).
- Utiliser une sauvegarde automatisée : utilisez une solution automatisée ou un service géré pour sauvegarder les données essentielles à l'entreprise.

- AWS Backup est un service entièrement géré qui vous permet de facilement centraliser et automatiser la protection des données sur les services AWS dans le cloud et sur site. Pour obtenir des conseils pratiques sur la façon de créer des sauvegardes automatisées à l'aide de AWS Backup, consultez la section Test de la sauvegarde et de la restauration de données.
- Automating backups and optimizing backup costs for Amazon EFS à l'aide de AWS Backup.

Ressources

Bonnes pratiques associées :

- REL09-BP01 Identifier et sauvegarder toutes les données qui doivent être sauvegardées, ou reproduire les données à partir de sources
- REL09-BP03 Effectuer automatiquement la sauvegarde des données
- REL13-BP02 Utiliser des stratégies de reprise définies pour répondre aux objectifs de reprise

Documents connexes :

- Utilisation d'AWS Backup pour la sauvegarde et la restauration des systèmes de fichiers Amazon EFS
- Instantanés Amazon EBS
- Utilisation de sauvegardes sur Amazon Relational Database Service
- Partenaire APN : partenaires pouvant faciliter la sauvegarde
- AWS Marketplace : produits pouvant être utilisés pour la sauvegarde
- Sauvegarde d'Amazon EFS
- S'auvegarde d'Amazon FSx for Windows File Server
- Backup et restauration d'Amazon ElastiCache (Redis OSS)

Vidéos connexes :

- AWS re:Invent 2023 - Backup and disaster recovery strategies for increased resilience
- AWS re:Invent 2023 - What's new with AWS Backup
- AWS re:Invent 2021 - Backup, disaster recovery, and ransomware protection with AWS

Matériel et services

Recherchez des possibilités de réduire les impacts en matière de durabilité de la charge de travail en modifiant vos pratiques de gestion du matériel. Réduisez la quantité de matériel nécessaire à allouer et à déployer, et sélectionnez le matériel et les services les plus efficaces pour votre charge de travail individuelle.

Bonnes pratiques

- SUS05-BP01 Utilisez le minimum de matériel pour répondre à vos besoins
- SUS05-BP02 Utiliser des types d'instance ayant le moins d'impact
- SUS05-BP03 Utiliser des services gérés
- SUS05-BP04 Optimiser votre utilisation des accélérateurs de calcul matériels

SUS05-BP01 Utilisez le minimum de matériel pour répondre à vos besoins

Utilisez la quantité minimale de matériel pour votre charge de travail afin de répondre efficacement aux besoins de votre entreprise.

Anti-modèles courants :

- Vous ne surveillez pas l'utilisation des ressources.
- Vous disposez de ressources avec un faible niveau d'utilisation dans votre architecture.
- Vous n'examinez pas l'utilisation du matériel statique pour déterminer s'il doit être redimensionné.
- Vous ne définissez pas d'objectifs d'utilisation du matériel pour votre infrastructure informatique en fonction de votre activitéKPIs.

Avantages liés au respect de cette bonne pratique : la rationalisation de vos ressources cloud permet de réduire l'impact environnemental d'une charge de travail, d'économiser de l'argent et de maintenir les repères de performance.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Sélectionnez de manière optimale le nombre total de composants matériels requis pour votre charge de travail afin d'améliorer son efficacité globale. AWS Cloud Il offre la flexibilité nécessaire pour étendre ou réduire le nombre de ressources de manière dynamique grâce à divers mécanismes,

tels que AWS Auto Scaling, et pour répondre à l'évolution de la demande. Il fournit également APIset SDKs permet de modifier les ressources avec un minimum d'effort. Utilisez ces capacités pour apporter des modifications fréquentes à vos mises en œuvre de charges de travail. En outre, utilisez les directives de redimensionnement issues des AWS outils pour exploiter efficacement vos ressources cloud et répondre aux besoins de votre entreprise.

Étapes d'implémentation

- Choisissez le type d'instance : choisissez le type d'instance qui répond le mieux à vos besoins. Pour déterminer comment choisir des instances Amazon Elastic Compute Cloud et utiliser des mécanismes tels que la sélection d'instances basée sur les attributs, consultez les rubriques suivantes :
  - Comment choisir le type d'EC2instance Amazon adapté à ma charge de travail ?
  - Sélection du type d'instance basée sur les attributs pour Amazon EC2 Fleet.
  - Créer un groupe Auto Scaling en utilisant la sélection du type d'instance basée sur des attributs.
- Mettre à l'échelle : diminuez les charges de travail variables par petits paliers.
- Utilisez plusieurs options d'achat de calcul : équilibrez la flexibilité, la capacité de mise à l'échelle et les économies de coûts des instances grâce à plusieurs options d'achat de calcul.
  - Les instances EC2 Amazon On-Demand sont parfaitement adaptées aux nouvelles charges de travail dynamiques et complexes qui ne peuvent pas être liées au type d'instance, à l'emplacement ou à la durée.
  - Les instances Amazon EC2 Spot constituent un excellent moyen de compléter les autres options pour les applications tolérantes aux pannes et flexibles.
  - Tirez parti des Compute Savings Plans pour des charges de travail stables qui offrent de la flexibilité si vos besoins (tels que l'AZ, la région, les familles d'instances ou les types d'instances) changent.
- Tirez parti de la diversité des instances et des zones de disponibilité : optimisez la disponibilité des applications et tirez parti de la capacité excédentaire en diversifiant vos instances et vos zones de disponibilité.
- Instances à la bonne taille : utilisez les recommandations de redimensionnement des AWS outils pour ajuster votre charge de travail. Pour en savoir plus, consultez Optimiser vos coûts avec les recommandations de dimensionnement et Dimensionnement adéquat : Dimensionnement des instances en fonction des charges de travail
  - Utilisez des recommandations de redimensionnement dans AWS Cost Explorer ou AWS Compute Optimizerpour identifier les opportunités de redimensionnement.

- Négocier des accords de niveau de service (SLAs) : négociez pour SLAs permettre de réduire temporairement la capacité pendant que l'automatisation déploie des ressources de remplacement.

Ressources

Documents connexes :

- Optimisation de votre AWS infrastructure pour la durabilité, partie I : calcul
- Sélection du type d'instance basée sur les attributs pour Auto Scaling for Amazon Fleet EC2
- AWS Compute Optimizer Documentation
- Utilisation de Lambda : optimisation de la performance
- Documentation sur la scalabilité automatique

Vidéos connexes :

- AWS re:Invent 2023 - Nouveautés d'Amazon EC2
- AWS re:Invent 2023 - Économies intelligentes : stratégies d'optimisation des coûts d'Amazon Elastic Compute Cloud
- AWS re:Invent 2022 - Optimisation d'Amazon Elastic Kubernetes Service en termes de performances et de coûts AWS
- AWS re:Invent 2023 - Informatique durable : réduction des coûts et des émissions de carbone grâce à AWS

SUS05-BP02 Utiliser des types d’instance ayant le moins d’impact

Contrôlez et utilisez en permanence de nouveaux types d’instances pour tirer parti des améliorations de l’efficacité énergétique.

Anti-modèles courants :

- Vous n’utilisez qu’une seule famille d’instances.
- Vous n’utilisez que des instances x86.
- Vous spécifiez un type d’instance dans votre configuration Amazon EC2 Auto Scaling.

- Vous utilisez des instances AWS de manière non conforme à leur utilisation prévue (par exemple, vous utilisez des instances optimisées pour le calcul pour une charge de travail exigeante en mémoire).
- Vous n'évaluez pas régulièrement de nouveaux types d'instance.
- Vous ne vérifiez pas les recommandations des outils de redimensionnement AWS tels que AWS Compute Optimizer.

Avantages liés au respect de cette bonne pratique : en utilisant des instances économes en énergie et dimensionnées, vous pouvez grandement réduire l'impact sur l'environnement et le coût de votre charge de travail.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

L'utilisation d'instances efficaces dans les charges de travail du cloud est cruciale pour réduire l'utilisation des ressources et pour une meilleure rentabilité. Contrôlez de façon continue le lancement de nouveaux types d'instances et profitez d'améliorations de l'efficacité énergétique, dont ces types d'instances conçus pour soutenir des charges de travail spécifiques comme l'entraînement et l'inférence du machine learning et le transcodage vidéo.

Étapes d'implémentation

- Découvrez et explorez les types d'instances : découvrez les types d'instances qui peuvent réduire l'impact environnemental de votre charge de travail.
  - Abonnez-vous à What's New with AWS pour rester au fait des dernières technologies et instances AWS.
  - Découvrez les différents types d'instance AWS.
  - Découvrez les instances AWS Graviton qui offrent les meilleures performances par watt d'énergie consommée dans Amazon EC2 en regardant re:Invent 2020 - Deep dive on AWS Graviton2 processor-powered Amazon EC2 instances et Deep dive into AWS Graviton3 and Amazon EC2 C7g instances.
- Utiliser les types d'instances ayant le moins d'impact : planifiez et migrez votre charge de travail vers les types d'instance avec le moins d'impact.
  - Définissez un processus pour évaluer les nouvelles fonctionnalités ou instances pour votre charge de travail. Profitez de l'agilité du cloud pour tester rapidement en quoi les nouveaux types

d'instance peuvent améliorer la durabilité environnementale de votre charge de travail. Utilisez des métriques de proxy pour mesurer le nombre de ressources nécessaires pour mener à bien une unité de travail.

- Si possible, modifiez votre charge de travail pour qu'elle fonctionne avec différents nombres de processeurs et différentes quantités de mémoire afin de maximiser votre choix de type d'instance.
- Envisagez de migrer votre charge de travail vers des instances basées sur Graviton pour améliorer l'efficacité des performances de votre charge de travail. Pour en savoir plus sur le transfert de charges de travail vers AWS Graviton, consultez AWS Graviton Fast Start et considérations relatives à la transition de charges de travail vers des instances Amazon Elastic Compute AWS Cloud basées sur Graviton.
- Envisagez de sélectionner l'option AWS Graviton lorsque vous utilisez des services gérés AWS
- Migrez votre charge de travail vers des régions qui offrent des instances ayant le moindre impact en matière de durabilité et qui répondent à vos exigences métier.
- Pour les charges de travail de machine learning, tirez parti d'un matériel conçu spécialement pour votre charge de travail, par exemple AWS, Trainum, AWS Inferentia et Amazon EC2 DL1. AWS Les instances Inferentia telles que les instances Inf2 offrent des performances par watt jusqu'à 50 % supérieures à celles des instances Amazon EC2 comparables.
- Utilisez Amazon SageMaker AI Inference Recommender pour dimensionner correctement le point de terminaison d'inférence ML.
- Pour les charges de travail en dents de scie (charges de travail dont les besoins en capacité supplémentaire sont peu fréquents), il convient d'utiliser des instances de performance en rafale.
- Pour les charges de travail sans état et tolérantes aux pannes, utilisez les instances spot Amazon EC2 pour augmenter l'utilisation globale du nuage et réduire l'impact des ressources inutilisées sur le développement durable.

- Exploiter et optimiser : exploitez et optimisez votre instance de charge de travail.
  - Pour les charges de travail éphémères, évaluez les métriques d'instance Amazon CloudWatch telles que `CPUUtilization` pour identifier si l'instance est inactive ou sous-utilisée.
  - Pour les charges de travail stables, vérifiez les outils de redimensionnement AWS tels qu'AWS Compute Optimizer à intervalles réguliers pour identifier les opportunités d'optimisation et de redimensionnement des ressources de calcul. Pour d'autres exemples et recommandations, consultez les ateliers suivants :
    - Atelier Well-Architected : recommandations de redimensionnement
    - Atelier Well-Architected - Rightsizing avec Compute Optimizer

- Atelier Well-Architected : optimiser les modèles matériels et observer les indicateurs de performance clés de durabilité

Ressources

Documents connexes :

- Optimisation de votre infrastructure AWS pour la durabilité, partie 1 : calcul
- AWS Graviton
- Amazon EC2 DL1
- Flotte de réserve de capacité Amazon EC2
- Parc d'instances Spot Amazon EC2
- Fonctions : configuration des fonctions Lambda
- Sélection de type d'instance basée sur des attributs pour la flotte d'Amazon EC2
- Création d'applications durables, efficaces et optimisées en matière de coûts sur AWS
- Comment le tableau de bord de durabilité de Contino aide les clients à optimiser leur empreinte carbone

Vidéos connexes :

- AWS re:Invent 2023 - AWS Graviton: The best price performance for your AWS workloads
- AWS re:Invent 2023 - New Amazon Elastic Compute Cloud generative AI capabilities in AWS Management Console
- AWS re:Invent 2023 = What's new with Amazon Elastic Compute Cloud
- AWS re:Invent 2023 - Smart savings: Amazon Elastic Compute Cloud cost-optimization strategies
- AWS re:Invent 2021 - Deep dive into AWS Graviton3 and Amazon EC2 C7g instances
- AWS re:Invent 2022 - Build a cost-, energy-, and resource-efficient compute environment

Exemples connexes :

- Solution : conseils pour l'optimisation des charges de travail de deep learning pour atteindre la durabilité sur AWS

SUS05-BP03 Utiliser des services gérés

Utilisez les services gérés pour fonctionner plus efficacement dans le cloud.

Anti-modèles courants :

- Vous utilisez des EC2 instances Amazon à faible taux d'utilisation pour exécuter vos applications.
- Votre équipe interne ne fait que gérer la charge de travail, sans avoir le temps de se concentrer sur l'innovation ou les simplifications.
- Vous déployez et maintenez des technologies pour des tâches qui peuvent être exécutées plus efficacement sur des services gérés.

Avantages liés au respect de cette bonne pratique :

- L'utilisation de services gérés transfère la responsabilité à AWS, qui dispose d'informations sur des millions de clients qui peuvent contribuer à de nouvelles innovations et à des gains d'efficacité.
- Le service géré répartit l'impact environnemental du service entre de nombreux utilisateurs grâce aux plans de contrôle multi-réseaux.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Les services gérés transfèrent la responsabilité du maintien d'un taux d'utilisation élevé et de l'optimisation de la durabilité du matériel déployé. AWS Les services gérés suppriment également la charge opérationnelle et administrative liée à la maintenance d'un service, ce qui permet à votre équipe de disposer de plus de temps et de se concentrer sur l'innovation.

Passez en revue votre charge de travail pour identifier les composants qui peuvent être remplacés par des services AWS gérés. Par exemple, Amazon RDS, Amazon Redshift et Amazon ElastiCache fournissent un service de base de données géré. Amazon AthenaEMR, Amazon et Amazon OpenSearch Service fournissent un service d'analyse géré.

Étapes d'implémentation

1. Faites l'inventaire de votre charge de travail : dressez l'inventaire de votre charge de travail pour les services et les composants.

2. Identifier les candidats : évaluez et identifiez les composants qui peuvent être remplacés par des services gérés. Voici quelques exemples de situations dans lesquelles vous pourriez envisager de recourir à un service géré :

Tâche	Ce qu'il faut utiliser sur AWS
Hébergement d'une base de données	Utilisez des instances Amazon Relational Database Service (RDSAmazon) gérées au lieu de gérer vos propres instances Amazon RDS sur Amazon Elastic Compute Cloud (EC2Amazon).
Héberger une charge de travail en conteneur	Utilisez AWS Fargate, au lieu de mettre en œuvre votre propre infrastructure de conteneurs.
Hébergement d'applications Web	Utilisez l'hébergement AWS Amplify en tant que service CI/CD et d'hébergement entièrement géré pour les sites Web statiques et les applications Web rendues côté serveur.

3. Créez un plan de migration : identifiez les dépendances et créez un plan de migration. Mettez à jour les runbooks et les playbooks en conséquence.
    - AWS Application Discovery Service rassemble et présente automatiquement les informations sur les dépendances et l'utilisation des applications pour vous aider à prendre des décisions en connaissance de cause pour votre programme de migration.
4. Effectuer des tests : testez le service avant de migrer vers le service géré.
5. Remplacez les services autohébergés : utilisez votre plan de migration pour remplacer les services autohébergés par des services gérés.
6. Contrôler et ajuster : surveillez continuellement le service une fois la migration terminée afin d'apporter les modifications nécessaires et d'optimiser le service.

Ressources

Documents connexes :

- AWS Cloud Produits

- AWS Calculateur du coût total de possession (TCO)
- Amazon DocumentDB
- Amazon Elastic Kubernetes Service () EKS
- Amazon Managed Streaming pour Apache Kafka (Amazon) MSK

Vidéos connexes :

- AWS re:Invent 2021 - Des opérations cloud à grande échelle avec AWS Managed Services
- AWS re:Invent 2023 - Meilleures pratiques pour opérer sur AWS

SUS05-BP04 Optimiser votre utilisation des accélérateurs de calcul matériels

Optimisez votre utilisation des instances de calcul accéléré pour réduire les exigences d'infrastructure physique de votre charge de travail.

Anti-modèles courants :

- Vous ne surveillez pas l'utilisation du GPU.
- Vous utilisez une instance à usage général pour la charge de travail alors qu'une instance spécialement conçue peut fournir des performances supérieures, des coûts plus faibles et de meilleures performances par watt.
- Vous utilisez des accélérateurs de calcul matériels pour les tâches où ils sont plus efficaces en utilisant des alternatives basées sur l'UC.

Avantages liés au respect de cette bonne pratique : en optimisant l'utilisation des accélérateurs matériels, vous pouvez réduire les exigences de votre charge de travail en matière d'infrastructure physique.

Niveau de risque encouru si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Si vous avez besoin d'une capacité de traitement élevée, vous pouvez bénéficier de l'utilisation d'instances de calcul accéléré, qui vous donnent accès à des accélérateurs de calcul matériels tels que des unités de traitement graphique (GPU) et des matrices de portes programmables sur site

(FPGA). Ces accélérateurs matériels exécutent certaines fonctions comme le traitement graphique ou la correspondance de modèles de données plus efficacement que les alternatives basées sur l'UC. De nombreuses charges de travail accélérées, telles que le rendu, le transcodage et le machine learning, sont très variables en matière d'utilisation des ressources. Exécutez ce matériel uniquement pendant le temps nécessaire et mettez-le hors service grâce à l'automatisation lorsque vous n'en avez plus besoin afin de limiter les ressources consommées.

Étapes d'implémentation

- Explorer les accélérateurs de calcul : identifiez les instances de calcul accéléré qui peuvent répondre à vos besoins.
- Utiliser du matériel conçu spécialement : pour les charges de travail de machine learning, tirez parti d'un matériel conçu spécialement pour votre charge de travail, tel qu'AWS Trainum, AWS Inferentia et Amazon EC2 DL1. AWS Les instances Inferentia telles que les instances Inf2 offrent des performances par watt jusqu'à 50 % supérieures à celles des instances comparables.
- Surveiller les métriques d'utilisation : collectez des métriques d'utilisation pour vos instances de calcul accéléré. Par exemple, vous pouvez utiliser l'agent CloudWatch pour collecter des métriques telles que `utilization_gpu` et `utilization_memory` pour vos GPU, comme indiqué dans Collecter les métriques des GPU NVIDIA avec Amazon CloudWatch.
- Redimensionner : optimisez le code, le fonctionnement du réseau et les paramètres des accélérateurs matériels pour veiller à ce que le matériel sous-jacent soit pleinement utilisé.
  - Optimiser les paramètres GPU
  - Surveillance et optimisation des GPU dans l'AMI Deep Learning
  - Optimisation des E/S pour le réglage des performances de GPU pour l'entraînement du deep learning dans l'IA Amazon SageMaker
- Maintenir à jour : utilisez les derniers pilotes GPU et bibliothèques à hautes performances.
- Libérer les instances non requises : utilisez l'automatisation pour libérer les instances GPU lorsqu'elles ne sont pas utilisées.

Ressources

Documents connexes :

- Calcul accéléré
- Let's Architect! Architecture avec des puces personnalisées et des accélérateurs

- Comment choisir le type d'instance EC2 approprié pour ma charge de travail ?
- Instances Amazon EC2 VT1
- Choisissez le meilleur accélérateur d'IA et la meilleure compilation de modèles pour l'inférence de vision par ordinateur avec l'IA Amazon SageMaker

Vidéos connexes :

- AWS re:Invent 2021 - How to select Amazon EC2 GPU instances for deep learning
- AWS Online Tech Talks - Deploying Cost-Effective Deep Learning Inference
- AWS re:Invent 2023 - Cutting-edge AI with AWS and NVIDIA
- AWS re:Invent 2022 - [NEW LAUNCH!] Introducing AWS Inferentia2-based Amazon EC2 Inf2 instances
- AWS re:Invent 2022 - Accelerate deep learning and innovate faster with AWS Trainium
- AWS re:Invent 2022 - Deep learning on AWS with NVIDIA: From training to deployment

Processus et culture

Recherchez des opportunités de réduire votre impact en matière de durabilité en modifiant vos pratiques de développement, de test et de déploiement.

Bonnes pratiques

- SUS06-BP01 Communiquer et répercuter en cascade vos objectifs de durabilité
- SUS06-BP02 Adopter des méthodes qui peuvent rapidement introduire des améliorations en matière de durabilité
- SUS06-BP03 Maintenir à jour votre charge de travail
- SUS06-BP04 Augmenter l'utilisation des environnements de génération
- SUS06-BP05 Utiliser des tests Device Farms gérés

SUS06-BP01 Communiquer et répercuter en cascade vos objectifs de durabilité

La technologie est un élément clé de la durabilité. Les équipes informatiques jouent un rôle crucial dans la mise en œuvre de changements significatifs pour atteindre les objectifs de durabilité de votre

organisation. Ces équipes doivent comprendre clairement les objectifs de durabilité de l'entreprise et s'efforcer de communiquer et de répercuter ces priorités dans l'ensemble des activités.

Anti-modèles courants :

- Vous ne connaissez pas les objectifs de durabilité de votre organisation ni comment ils s'appliquent à votre équipe.
- Vous n'êtes pas suffisamment sensibilisé et formé à l'impact environnemental des charges de travail dans le cloud.
- Vous n'êtes pas sûr des domaines spécifiques à privilégier.
- Vous n'impliquez pas vos employés et vos clients dans vos initiatives de durabilité.

Avantages liés au respect de cette bonne pratique : de l'optimisation de l'infrastructure et des systèmes à l'utilisation de technologies innovantes, les équipes informatiques peuvent réduire les émissions de carbone de l'organisation et minimiser la consommation de ressources. La communication des objectifs de durabilité peut permettre aux équipes informatiques de s'améliorer et de s'adapter en permanence à l'évolution des défis en matière de durabilité. De plus, ces optimisations durables se traduisent aussi souvent par des économies de coûts, ce qui renforce l'argumentaire.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Les principaux objectifs de durabilité des équipes informatiques devraient être d'optimiser les systèmes et les solutions afin d'accroître l'efficacité des ressources et de minimiser l'empreinte carbone de l'organisation et son impact global sur l'environnement. Les services partagés et les initiatives telles que les programmes de formation et les tableaux de bord opérationnels, peuvent aider les organisations à optimiser leurs opérations informatiques et à élaborer des solutions susceptibles de contribuer à réduire de manière significative l'empreinte carbone. Le cloud offre l'opportunité non seulement de transférer les responsabilités en matière d'infrastructure physique et d'approvisionnement en énergie au fournisseur cloud, mais également d'optimiser en permanence l'efficacité des ressources des services basés sur le cloud.

Lorsque les équipes utilisent le modèle d'efficacité et de responsabilité partagée inhérent au cloud, elles peuvent réduire de manière significative l'impact de l'organisation sur l'environnement. Cela peut à son tour contribuer aux objectifs globaux de durabilité de l'organisation et démontrer la valeur de ces équipes en tant que partenaires stratégiques sur la voie d'un futur plus durable.

Étapes d'implémentation

- Définir des buts et des objectifs : établissez des objectifs bien définis pour votre programme informatique. Cela implique de recueillir l'avis des parties prenantes responsables de différents départements tels que l'informatique, la durabilité et les finances. Ces équipes doivent définir des objectifs mesurables qui s'alignent sur les objectifs de durabilité de votre organisation, notamment dans des domaines tels que la réduction des émissions carbone et l'optimisation des ressources.
- Comprendre les limites de la comptabilisation du carbone dans votre entreprise : découvrez comment les méthodes de comptabilisation du carbone telles que le protocole sur les gaz à effet de serre (GES) sont liées à vos charges de travail dans le cloud (pour plus de détails, consultez Durabilité du cloud).
- Utiliser des solutions cloud pour la comptabilisation du carbone : utilisez des solutions cloud telles que les solutions de comptabilisation du carbone sur AWS pour suivre les émissions de GES de première, deuxième et troisième catégories dans l'ensemble de vos opérations, portefeuilles et chaînes de valeur. Grâce à ces solutions, les organisations peuvent rationaliser l'acquisition de données sur les émissions de GES, simplifier les rapports et obtenir des informations pour étayer leurs stratégies climatiques.
- Surveiller l'empreinte carbone de votre portefeuille informatique : suivez et signalez les émissions de carbone de vos systèmes informatiques. Utilisez l'outil d'empreinte carbone du clientAWS pour suivre, mesurer, réviser et prédire les émissions de carbone générées par votre utilisation d'AWS.
- Communiquer à vos équipes l'utilisation des ressources via des métriques proxy : suivez et signalez votre utilisation de ressources via des métriques proxy. Dans les modèles de tarification à la demande du cloud, l'utilisation des ressources est liée au coût, qui est une métrique généralement facile à comprendre. Au minimum, utilisez le coût comme une métrique proxy pour communiquer l'utilisation des ressources et les améliorations apportées par chaque équipe.
  - Activer la granularité horaire dans Cost Explorer et créer un rapport d'utilisation et de coût (CUR) : le rapport CUR fournit une granularité d'utilisation journalière ou horaire, des tarifs, des coûts et des attributs d'utilisation pour tous les services AWS. Utilisez Cloud Intelligence Dashboards et son tableau de bord des métriques proxy de durabilité comme point de départ pour le traitement et la visualisation des données basées sur les coûts et l'utilisation. Pour plus de détails, consultez les documents suivants :
  - Mesure et suivi de l'efficacité du cloud à l'aide de métriques proxy de durabilité, partie I : Que sont les métriques proxy ?
  - Mesure et suivi de l'efficacité du cloud à l'aide de métriques proxy de durabilité, partie II : Établissement d'un pipeline de métriques

- Optimiser et évaluer en permanence : utilisez un processus d'amélioration pour optimiser en permanence vos systèmes informatiques, y compris la charge de travail dans le cloud pour des raisons d'efficacité et de durabilité. Surveillez l'empreinte carbone avant et après la mise en œuvre de la stratégie d'optimisation. Utilisez la réduction de l'empreinte carbone pour évaluer l'efficacité.
- Favoriser une culture de la durabilité : utilisez des programmes de formation (tels que AWS Skill Builder) pour sensibiliser vos employés à la durabilité. Impliquez-les dans des initiatives de durabilité. Partagez et célébrez leurs témoignages de réussite. Utilisez des incitations pour les récompenser s'ils atteignent leurs objectifs de durabilité.

Ressources

Documents connexes :

- Comprendre les estimations de vos émissions de carbone

Vidéos connexes :

- AWS re:Invent 2023 - Accelerate data-driven circular economy initiatives with AWS
- AWS re:Invent 2023 - Sustainability innovation in AWS Global Infrastructure
- AWS re:Invent 2023 - Sustainable architecture: Past, present, and future
- AWS re:Invent 2022 - Delivering sustainable, high-performing architectures
- AWSre:Invent 2022 – L'architecture de manière durable et réduisez votre AWSempreinte carbone
- AWS re:Invent 2022 - Sustainability in AWS global infrastructure

Exemples connexes :

- Well-Architected Lab – Transformer les rapports sur les coûts et l'utilisation en rapports d'efficacité

Formations associées :

- Transformation en matière de durabilité sur AWS
- SimuLearn – Rapports sur la durabilité
- Décarbonisation avec AWS

SUS06-BP02 Adopter des méthodes qui peuvent rapidement introduire des améliorations en matière de durabilité

Adoptez des méthodes et des processus pour valider les améliorations potentielles, minimiser les coûts des tests et apporter de petites améliorations.

Anti-modèles courants :

- L'examen de la durabilité de votre application est une tâche qui n'est effectuée qu'une seule fois au début d'un projet.
- Votre charge de travail est devenue obsolète, car le processus de lancement est trop lourd pour introduire des changements mineurs dans un souci d'efficacité des ressources.
- Vous ne disposez pas de mécanismes pour améliorer votre charge de travail afin d'atteindre davantage de durabilité.

Avantages liés au respect de cette bonne pratique : en établissant un processus pour introduire et suivre les améliorations en matière de durabilité, vous serez en mesure d'adopter en permanence de nouvelles fonctionnalités et capacités, de résoudre les problèmes et d'améliorer l'efficacité de la charge de travail.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : moyen

Directives d'implémentation

Testez et validez les améliorations potentielles en matière de durabilité avant de les déployer en production. Tenez compte du coût des tests lors du calcul des avantages futurs potentiels d'une amélioration. Développez des méthodes d'essai à faible coût pour apporter de petites améliorations.

Étapes d'implémentation

- Comprenez et communiquez les objectifs de durabilité de votre organisation : comprenez les objectifs de durabilité de votre organisation, tels que la réduction des émissions de carbone ou la gestion de l'eau. Traduisez ces objectifs en exigences de durabilité pour vos charges de travail cloud. Communiquez ces exigences aux principales parties prenantes.
- Ajoutez des exigences de durabilité à votre carnet de commandes : ajoutez des exigences relatives à l'amélioration de la durabilité à votre carnet de développement.
- Itérer et améliorer : utilisez un processus d'amélioration itératif pour identifier, évaluer, prioriser, tester et déployer ces améliorations.

- Test à l'aide d'un produit minimum viable (MVP) : développez et testez des améliorations potentielles en utilisant le minimum de composants représentatifs viables afin de réduire le coût et l'impact environnemental des tests.
- Rationaliser le processus : améliorez et rationalisez en permanence vos processus de développement. À titre d'exemple, automatisez votre processus de livraison de logiciels en utilisant des pipelines d'intégration et de livraison continues (CI/CD) pour tester et déployer les améliorations potentielles afin de réduire le niveau d'effort et de limiter les erreurs causées par les processus manuels.
- Formation et sensibilisation : organisez des programmes de formation pour les membres de votre équipe afin de les sensibiliser au développement durable et à l'impact de leurs activités sur les objectifs de durabilité de votre organisation.
- Évaluer et ajuster : évaluez en permanence l'impact des améliorations et procédez aux ajustements nécessaires.

Ressources

Documents connexes :

- AWS active des solutions de durabilité

Vidéos connexes :

- AWS re:Invent 2023 - Sustainable architecture: Past, present, and future
- AWS re:Invent 2022 - Delivering sustainable, high-performing architectures
- AWSre:Invent 2022 - L'architecture de manière durable et réduisez votre AWSempreinte carbone
- AWS re:Invent 2022 - Sustainability in AWS global infrastructure
- AWS re:Invent 2023 - What's new with AWS observability and operations

SUS06-BP03 Maintenir à jour votre charge de travail

Maintenez votre charge de travail à jour pour adopter des fonctionnalités efficaces, supprimer les problèmes et améliorer l'efficacité globale de votre charge de travail.

Anti-modèles courants :

- Vous supposez que votre architecture actuelle est statique et ne sera pas mise à jour au fil du temps.
- Vous ne disposez pas de systèmes ou de rythme régulier pour évaluer la compatibilité des packages et des logiciels mis à jour avec votre charge de travail.

Avantages liés au respect de cette bonne pratique : en mettant en place un processus pour garder votre charge de travail à jour, vous pourrez adopter de nouvelles fonctionnalités et capacités, résoudre les problèmes et améliorer l'efficacité de la charge de travail.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : bas

Directives d'implémentation

Des systèmes d'exploitation, des moteurs d'exécution, des logiciels intermédiaires, des bibliothèques et des applications à jour peuvent améliorer l'efficacité de la charge de travail et faciliter l'adoption de technologies plus efficaces. Les logiciels à jour peuvent également inclure des fonctions permettant de mesurer plus précisément l'impact en matière de durabilité de votre charge de travail, car les fournisseurs proposent des fonctions pour atteindre leurs propres objectifs de durabilité. Adoptez une cadence régulière pour maintenir votre charge de travail à jour avec les dernières fonctionnalités et versions.

Étapes d'implémentation

- Définir un processus : utilisez un processus et un calendrier pour évaluer les nouvelles fonctionnalités ou instances pour votre charge de travail. Profitez de l'agilité du cloud pour tester rapidement en quoi les nouvelles fonctionnalités peuvent permettre à votre charge de travail de :
  - Réduire les impacts sur la durabilité.
  - Gagner en efficacité de la performance.
  - Supprimer les obstacles à une amélioration planifiée.
  - Améliorer votre capacité à mesurer et à gérer les impacts en matière de durabilité.
- Réaliser un inventaire : établissez l'inventaire de votre logiciel de charge de travail et de l'architecture, et identifiez les composants pouvant être mis à jour.
  - Vous pouvez utiliser la fonctionnalité AWS Systems Manager Inventory pour collecter les métadonnées du système d'exploitation (OS), de l'application et de l'instance à partir de vos instances Amazon EC2 et comprendre rapidement quelles instances exécutent les logiciels et les configurations requis par votre politique logicielle et quelles instances doivent être mises à jour.

- Découvrez la procédure de mise à jour : comprenez comment mettre à jour les composants de votre charge de travail.

Composant de charge de travail	Comment mettre à jour
Images de machine	Utilisez EC2 Image Builder pour gérer les mises à jour des images de serveur Amazon Machine Images (AMI) pour Linux ou Windows.
Images de conteneur	Utilisez Amazon Elastic Container Registry (Amazon ECR) avec votre pipeline existant pour gérer des images Amazon Elastic Container Service (Amazon ECS).
AWS Lambda	AWS Lambda comprend les fonctionnalités de gestion des versions.

- Utiliser l'automatisation : utilisez l'automatisation pour le processus de mise à jour afin de réduire le niveau d'effort nécessaire au déploiement des nouvelles fonctionnalités et de limiter les erreurs causées par les processus manuels.
  - Vous pouvez utiliser CI/CD pour mettre automatiquement à jour les AMI, les images de conteneurs et d'autres artefacts liés à votre application cloud.
  - Vous pouvez utiliser des outils tels que le gestionnaire de correctifs AWS Systems Manager pour automatiser le processus de mise à jour du système et planifier l'activité à l'aide des fenêtres de maintenance AWS Systems Manager.

Ressources

Documents connexes :

- Centre d'architecture AWS
- Nouveautés avec AWS
- Outils pour développeurs AWS

Vidéos connexes :

- AWS re:Invent 2022 - Optimize your AWS workloads with best-practice guidance
- All Things Patch : AWS Systems Manager

SUS06-BP04 Augmenter l'utilisation des environnements de génération

Augmentez l'utilisation des ressources pour développer, tester et construire vos charges de travail.

Anti-modèles courants :

- Vous provisionnez ou résiliez manuellement vos environnements de construction.
- Vous faites fonctionner vos environnements de construction indépendamment des activités de test, de construction ou de lancement (par exemple, en faisant fonctionner un environnement en dehors des heures de travail des membres de votre équipe de développement).
- Vous provisionnez trop de ressources pour vos environnements de construction.

Avantages liés au respect de cette bonne pratique : en augmentant l'utilisation des environnements de création, vous pouvez améliorer l'efficacité globale de votre charge de travail dans le cloud tout en allouant les ressources aux concepteurs pour qu'ils puissent développer, tester et créer efficacement.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : faible

Directives d'implémentation

Exploitez l'automatisation et l'infrastructure en tant que code pour mettre en place des environnements de construction lorsque cela est nécessaire et les arrêter lorsqu'ils ne sont pas utilisés. Un modèle courant consiste à planifier des périodes de disponibilité qui coïncident avec les heures de travail des membres de votre équipe de développement. Vos environnements de test doivent ressembler de près à la configuration de production. Toutefois, recherchez les possibilités d'utilisation des types d'instance avec une capacité de débordement, des instances Amazon EC2 Spot, des services de base de données à mise à l'échelle automatique, des conteneurs et des technologies sans serveur pour aligner la capacité de développement et de test sur l'utilisation. Limitez le volume de données pour répondre aux exigences du test. Si vous utilisez des données de production dans les tests, étudiez les possibilités de partager les données de production et de ne pas déplacer les données à un autre emplacement.

Étapes d'implémentation

- Utiliser l'infrastructure en tant que code : utilisez l'infrastructure en tant que code pour provisionner vos environnements de construction.
- Utilisez l'automatisation : utilisez l'automatisation pour gérer le cycle de vie de vos environnements de développement et de test et maximiser l'efficacité de vos ressources de construction.
- Maximiser l'utilisation : utilisez des stratégies pour optimiser l'utilisation des environnements de développement et de test.
  - Utilisez des environnements représentatifs viables minimum pour développer et tester les améliorations potentielles.
  - Utilisez les technologies sans serveur si possible.
  - Utilisez des instances à la demande pour compléter les appareils de vos développeurs.
  - Utilisez des types d'instance à capacité de débordement, des instances Spot et d'autres technologies pour harmoniser la capacité de création et l'utilisation.
  - Adoptez des services natifs du cloud pour l'accès à un shell d'instance sécurisé plutôt que de déployer des flottes d'hôtes bastion.
  - Mettez automatiquement à l'échelle vos ressources de construction en fonction de vos tâches de construction.

Ressources

Documents connexes :

- Gestionnaire de session AWS Systems Manager
- Instances de performance à capacité extensible Amazon EC2
- Présentation de AWS CloudFormation
- Présentation de AWS CodeBuild
- Instance Scheduler sur AWS

Vidéos connexes :

- AWS re:Invent 2023 - Continuous integration and delivery for AWS

SUS06-BP05 Utiliser des tests Device Farms gérés

Utilisez les Device Farms gérés pour tester efficacement une nouvelle fonctionnalité sur un ensemble représentatif de matériel.

Anti-modèles courants :

- Vous testez et déployez manuellement votre application sur des appareils physiques individuels.
- Vous n'utilisez pas le service de test d'applications pour tester et interagir avec vos applications (par exemple, les applications Android, iOS et Web) sur des appareils physiques réels.

Avantages liés au respect de cette bonne pratique : l'utilisation de batteries d'appareils gérés pour tester des applications compatibles avec le cloud présente de nombreux avantages :

- La solution comprend des fonctionnalités plus efficaces pour tester l'application sur de nombreux appareils différents.
- Elle élimine la nécessité d'une infrastructure interne pour les essais.
- Elle permet l'utilisation de divers types d'appareils, y compris des matériels plus anciens et moins populaires, ce qui élimine le besoin de mises à niveau inutiles des appareils.

Niveau d'exposition au risque si cette bonne pratique n'est pas respectée : faible

Directives d'implémentation

L'utilisation de Device Farms gérés peut vous aider à rationaliser le processus de test des nouvelles fonctionnalités sur un ensemble représentatif de matériel. Les tests Device Farms gérés proposent divers types d'appareils, notamment du matériel plus ancien et moins courant, et permettent d'éviter que les mises à niveau inutiles d'appareils affectent la durabilité des clients.

Étapes d'implémentation

- Définir les exigences en matière de tests : définissez vos exigences et votre plan de test (comme le type de test, les systèmes d'exploitation et le calendrier des tests).
  - Vous pouvez utiliser Amazon CloudWatch RUM pour collecter et analyser les données côté client et élaborer votre plan de test.
- Sélectionnez un parc d'appareils gérés : sélectionnez un parc d'appareils gérés capable de répondre à vos exigences en matière de tests. Par exemple, vous pouvez utiliser AWS Device

Farm pour tester et comprendre l'impact de vos modifications sur un ensemble représentatif de matériel.

- Utiliser l'automatisation : utilisez l'intégration continue/le déploiement continu (CI/CD) pour programmer et exécuter vos tests.
  - Intégration d'AWS Device Farm à votre pipeline CI/CD pour exécuter des tests Selenium sur plusieurs navigateurs
  - Création et test d'applications iOS et iPadOS avec AWS DevOps et les services mobiles
- Examiner et ajuster : examinez continuellement les résultats de vos tests et apportez les améliorations nécessaires.

Ressources

Documents connexes :

- Liste d'appareils AWS Device Farm
- Affichage du tableau de bord CloudWatch RUM

Vidéos connexes :

- AWS re:Invent 2023 - Improve your mobile and web app quality using AWS Device Farm
- AWS re:Invent 2021 - Optimize applications through end user insights with Amazon CloudWatch RUM

Exemples connexes :

- Exemple d'application AWS Device Farm pour Android
- Exemple d'application AWS Device Farm pour iOS
- Tests Web Appium pour AWS Device Farm

Conclusion

Un nombre croissant d’organisations fixent des objectifs de développement durable en réponse aux modifications de la réglementation gouvernementale, au besoin de se démarquer de la concurrence et à la demande des clients, des employés et des investisseurs. CTOs, les architectes, les développeurs et les membres de l'équipe opérationnelle cherchent des moyens de contribuer directement aux objectifs de durabilité de leur organisation. En utilisant ces principes de conception et les meilleures pratiques soutenus par les AWS services, vous pouvez prendre des décisions éclairées en équilibrant la sécurité, les coûts, les performances, la fiabilité et l'excellence opérationnelle avec des résultats de durabilité pour vos AWS Cloud charges de travail. Chaque action que vous entreprenez pour réduire l’utilisation des ressources et augmenter l’efficacité de vos charges de travail contribue à réduire l’impact environnemental et participe aux objectifs de durabilité plus larges de votre organisation.

Collaborateurs

Les personnes qui ont contribué à ce document incluent :

- Sam Mokhtari, architecte senior de solutions en matière d’efficacité, Amazon Web Services
- Brendan Sisson, architecte principal de solutions de durabilité, Amazon Web Services
- Margaret O’Toole, responsable technique de la durabilité, Amazon Web Services
- Steffen Grunwald, architecte principal des solutions de durabilité, Amazon Web Services
- Ryan Eccles, ingénieur principal, service durabilité, Amazon
- Rodney Lester, architecte principal, Amazon Web Services
- Adrian Cockcroft, Vice-président de l’architecture de durabilité, Amazon Web Services
- Ian Meyers, directeur technologique, architecture de solutions, Amazon Web Services

Suggestions de lecture

Pour en savoir plus, reportez-vous à :

- AWS Well-Architected
- AWS Centre d'architecture
- Durabilité dans le cloud
- AWS active des solutions de durabilité
- The Climate Pledge
- Objectifs de développement durable des Nations Unies
- Protocole des GES

Révisions du document

Pour être informé des mises à jour de ce livre blanc, abonnez-vous au flux RSS.

Modification	Description	Date
Mises à jour des conseils sur les bonnes pratiques	Les bonnes pratiques ont été mises à jour avec de nouveaux conseils dans les domaines suivants : SUS 1, SUS 3, SUS 4, SUS 5 et SUS 6. Les directives ont été améliorées dans ces domaines de bonnes pratiques . Une nouvelle bonne pratique a été ajoutée à SUS 6, SUS06-BP01 Communiquer et répercuter en cascade vos objectifs de durabilité. Les bonnes pratiques existantes dans SUS 6 ont été renumérotées.	6 novembre 2024
Mises à jour des conseils sur les bonnes pratiques	Petits changements dans l'ensemble du pilier.	27 juin 2024
Niveaux de risque actualisés	Mises à jour mineures pour atteindre les niveaux de risque répondant aux bonnes pratiques.	3 octobre 2023
Mises à jour des conseils sur les bonnes pratiques	Les bonnes pratiques ont été mises à jour avec de nouvelles directives dans les domaines suivants : adaptation à la demande, logiciel et architect	13 juillet 2023

	ure, données, matériel et services.	
Mise à jour du nouveau cadre	Les bonnes pratiques ont été mises à jour avec des recommandations et de nouvelles bonnes pratiques.	10 avril 2023
Livre blanc mis à jour	Les bonnes pratiques ont été mises à jour avec de nouvelles recommandations en matière d'implémentation.	15 décembre 2022
Livre blanc mis à jour	Développement des bonnes pratiques et ajout de plans d'amélioration.	20 octobre 2022
Publication initiale	Pilier Durabilité - AWS Well-Architected Framework publié.	2 décembre 2021

Avis

Il incombe aux clients de procéder à une évaluation indépendante des informations contenues dans le présent document. Ce document : (a) est fourni à titre informatif uniquement, (b) représente les offres de AWS produits et les pratiques actuelles, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ni aucune assurance de la part de AWS ses filiales, fournisseurs ou concédants de licence. AWS les produits ou services sont fournis « tels quels » sans garanties, déclarations ou conditions d'aucune sorte, qu'elles soient explicites ou implicites. Les responsabilités et obligations AWS de ses clients sont régies par AWS des accords, et ce document ne fait partie d'aucun accord conclu entre AWS et ses clients et ne les modifie pas.

© 2023, Amazon Web Services, Inc. ou ses sociétés apparentées. Tous droits réservés.

AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le AWS glossaire dans la Glossaire AWS référence.