Infrastructure AWS correctement architecturée

Novembre 2016

Mentions légales

Ce document est fourni à titre informatif uniquement. Il présente l'offre de produits et les pratiques actuelles d'AWS à la date de publication de ce document, des informations qui sont susceptibles d'être modifiées sans préavis. Il incombe aux clients de procéder à leur propre évaluation indépendante des informations contenues dans ce document et chaque client est responsable de son utilisation des produits ou services AWS, chacun étant fourni « en l'état », sans garantie d'aucune sorte, qu'elle soit explicite ou implicite. Ce document n'offre pas de garantie, représentation, engagement contractuel, condition ou assurance de la part d'AWS, de ses sociétés apparentées, fournisseurs ou concédants de licence. Les responsabilités et obligations d'AWS vis-à-vis de ses clients sont régies par les contrats AWS. Le présent document ne fait partie d'aucun contrat entre AWS et ses clients et n'en modifie aucun.

Table des matières

Résumé

Ce livre blanc décrit l'**infrastructure AWS correctement architecturée**, qui permet aux clients de vérifier et d'améliorer leur architecture cloud, ainsi que de mieux comprendre l'impact métier de leurs décisions de conception. Nous abordons les principes généraux de conception aussi bien que les bonnes pratiques spécifiques, et nous fournissons des conseils dans les cinq domaines conceptuels définis comme les *piliers* de l'infrastructure correctement architecturée.

Introduction

Chez Amazon Web Services (AWS), nous sommes conscients de la valeur que représente la formation de nos clients aux bonnes pratiques architecturales, car ils pourront ainsi concevoir et piloter au sein du cloud des systèmes fiables, sécurisés, efficaces et économiques. Dans le cadre de cette démarche, nous avons développé l'infrastructure AWS correctement architecturée, qui vous permet de comprendre les avantages et les inconvénients des décisions que vous prenez lors du développement de systèmes sur AWS. Nous considérons que des systèmes à l'architecture bien conçue accroissent grandement la probabilité de la réussite commerciale.

Les architectes des solutions AWS possèdent des années d'expérience en matière d'architecture de solutions sur une très grande diversité de verticaux (solutions verticales) métier et de cas d'utilisation. En outre, nous avons contribué à la conception et à la révision de milliers d'architectures de clients sur AWS. A partir de cette expérience, nous avons identifié les bonnes pratiques et les principales stratégies d'architecture de systèmes dans le cloud.

L'infrastructure AWS correctement architecturée documente un ensemble de questions de base qui vous permettent de comprendre si une architecture spécifique respecte les bonnes pratiques du cloud. L'infrastructure offre une approche cohérente pour évaluer les systèmes par rapport aux qualités que vous escomptez de systèmes modernes basés sur le cloud, ainsi que les corrections requises pour atteindre ces qualités. A mesure qu'AWS évolue et que notre collaboration avec les clients représente une source d'enseignements de plus en plus riche, nous continuons à affiner la définition d'une architecture correctement conçue.

Ce livre blanc s'adresse à celles et à ceux qui sont dépositaires de rôles technologiques, comme les directeurs techniques, les architectes, les développeurs et les membres de l'équipe d'exploitation. Après avoir lu ce document, vous connaîtrez les stratégies et les bonnes pratiques AWS à utiliser lors de la conception et de l'exploitation d'une architecture cloud. Le présent document ne propose ni détail de mise en œuvre ni modèle architectural. Cependant, il contient des références aux ressources appropriées relatives à ces informations.

Définition de l'infrastructure AWS correctement architecturée

Tous les jours, les experts AWS aident les clients à concevoir l'architecture de leurs systèmes afin de tirer parti des meilleures pratiques dans le cloud. Nous collaborons avec vous pour parvenir à des compromis architecturaux tandis que vos conceptions évoluent. Lorsque vous déployez ces systèmes dans des environnements réels, vous découvrez les performances effectives de ces systèmes, ainsi que les conséquences de ces compromis.

Grâce aux enseignements acquis, nous avons créé l'infrastructure AWS correctement architecturée, qui constitue un ensemble de questions que vous pouvez reprendre pour évaluer le degré de conformité d'une architecture aux bonnes pratiques AWS.

L'infrastructure AWS correctement architecturée repose sur cinq piliers : la sécurité, la fiabilité, l'efficacité des performances, l'optimisation des coûts et l'excellence opérationnelle.

Nom du pilier	Description
Sécurité	Capacité à protéger les informations, les systèmes et les ressources lors de l'offre d'une valeur métier, via l'évaluation des risques et les stratégies d'atténuation.
Fiabilité	Capacité d'un système à récupérer à partir de défaillances de l'infrastructure ou d'un service, à acquérir dynamiquement les ressources de calcul nécessaires à la satisfaction de la demande et à atténuer les perturbations telles que les erreurs de configuration ou les problèmes réseau temporaires.
Efficacité des performances	Capacité à utiliser efficacement les ressources informatiques pour satisfaire aux exigences système et à maintenir cette efficacité au fur et à mesure que la demande change et que les technologies évoluent.
Optimisation des coûts	Capacité à éviter ou à supprimer les coûts superflus et les ressources sous-optimales.
Excellence opérationnelle	Capacité à exécuter et à superviser les systèmes afin d'offrir une valeur métier, et à perfectionner en permanence la prise en charge des processus et des procédures.

Lorsque vous concevez l'architecture de solutions, vous établissez des compromis entre les piliers en fonction de votre contexte métier et ces décisions professionnelles peuvent orienter vos priorités en matière d'ingénierie. Vous souhaiterez peut-être procéder à une optimisation afin de réduire les coûts aux dépens de la fiabilité dans les environnements de développement ou, dans le cas de solutions critiques, vous préfèrerez optimiser la fiabilité grâce à des coûts accrus. Dans les solutions de commerce électronique, les performances peuvent affecter le chiffre d'affaires et la propension du client à acheter. La sécurité et l'excellence opérationnelle ne donnent généralement pas lieu à des compromis avec les autres piliers.

Principes généraux de conception

L'infrastructure AWS correctement architecturée identifie un ensemble de principes généraux de conception destinés à favoriser une bonne conception dans le cloud :

- **Cesser** de **présumer les besoins en capacité :** ne devinez plus les besoins en capacité de votre infrastructure. Avant de déployer un système, lorsque vous prenez une décision en matière de capacité, il se peut que vous vous retrouviez face à des ressources inutilisées onéreuses ou à traiter les implications, en termes de performances, d'une capacité limitée. Grâce au cloud computing, ces problèmes disparaissent. Vous pouvez utiliser autant de capacité que vous le souhaitez en fonction de vos besoins, et l'agrandir ou la réduire automatiquement.
- **Tester les systèmes à l'échelle de la production :** dans le cloud, vous pouvez créer à la demande un environnement de test à l'échelle de la production, exécuter les tests, puis désactiver les ressources. Parce que vous ne payez l'environnement de test que lorsqu'il s'exécute, vous pouvez simuler votre environnement réel pour une partie du coût que représenteraient les tests sur site.
- **Recourir à l'automatisation pour faciliter l'expérimentation architecturale :** l'automatisation vous permet de créer et de répliquer vos systèmes à moindre coût, ainsi que d'économiser les frais d'un effort manuel. Vous pouvez suivre les modifications apportées à l'automatisation, auditer l'impact et rétablir les paramètres antérieurs si nécessaire.

- **Autoriser les architectures évolutives :** dans un environnement traditionnel, les décisions architecturales sont souvent implémentées comme événements uniques et statiques, avec quelques versions majeures d'un système pendant sa durée de vie. Tandis que l'activité et son contexte continuent à évoluer, ces décisions initiales peuvent entraver la capacité du système à satisfaire des exigences métier changeantes. Dans le cloud, la capacité d'automatiser et de tester à la demande réduit le risque d'impact des modifications de conception. Les systèmes peuvent ainsi évoluer au fil du temps, de telle sorte que les entreprises parviennent à tirer profit des innovations en tant que pratique standard.
- **Architectures orientées données :** dans le cloud, vous pouvez recueillir les données relatives à la façon dont vos choix architecturaux affectent le comportement de votre charge de travail. Il vous est ainsi possible de prendre des décisions factuelles relatives à l'amélioration de votre charge de travail. Comme votre infrastructure cloud est le code, vous pouvez utiliser ces données pour documenter vos choix architecturaux et vos améliorations au fil du temps.
- **Procéder à des améliorations via les « game days » :** testez l'exécution de votre architecture et de vos processus en planifiant à intervalles réguliers des « game days » afin de simuler les événements en production. Cela vous aidera à comprendre à quels endroits il est possible d'apporter des améliorations, ainsi qu'à développer une expérience organisationnelle dans la gestion de ces événements.

Les cinq piliers de l'infrastructure AWS correctement architecturée

Créer un système logiciel s'apparente à la construction d'un immeuble. Si les fondations ne sont pas solides, de réels problèmes structurels peuvent saper l'intégrité et la fonction de l'immeuble. Lors de la conception architecturale de solutions technologiques, si vous négligez les cinq piliers de la sécurité, de la fiabilité, de l'efficacité des performances, de l'optimisation des coûts et de l'expérience opérationnelle, la création d'un système qui répond à vos attentes et à vos exigences peut se révéler difficile. Lorsque vous intégrez ces piliers à votre architecture, vous contribuez à créer des systèmes stables et efficaces. Vous pouvez ainsi vous concentrer sur d'autres aspects de la conception, telles que les exigences fonctionnelles.

Cette section décrit chacun des cinq piliers et inclut les définitions, bonnes pratiques, questions, considérations et services clés AWS appropriés.

Pilier « Sécurité »

Le pilier **Sécurité** inclut la capacité à protéger les informations, les systèmes et les ressources tout en offrant une valeur métier, via l'évaluation des risques et les stratégies d'atténuation.

Principes de conception

Dans le cloud, il existe un certain nombre de principes qui peuvent vous aider à renforcer la sécurité de votre système.

- **Appliquer la sécurité à toutes les couches :** au lieu de n'exécuter les dispositifs de sécurité (les pare-feux, par exemple) qu'à la périphérie de votre infrastructure, utilisez les pare-feux et autres contrôles de sécurité sur l'ensemble de vos ressources (chaque serveur virtuel, chaque équilibreur de charge et chaque sous-réseau, par exemple).
- **Activer la traçabilité :** enregistrez et auditez toutes les actions et toutes les modifications apportées à votre environnement.
- **Mettre en œuvre le principe du privilège minimum :** assurez-vous que l'autorisation est adaptée à chaque interaction avec vos ressources AWS et implémentez directement des contrôles d'accès logiques puissants sur les ressources.

- **Se concentrer sur la sécurisation de votre système :** avec le modèle Responsabilité partagée AWS, vous pouvez vous concentrer sur la sécurisation de vos applications, données et systèmes d'exploitation, pendant qu'AWS fournit une infrastructure et des services sécurisés.
- **Automatiser les bonnes pratiques de sécurité :** les mécanismes de sécurité basés sur les logiciels améliorent votre capacité à évoluer plus rapidement et plus économiquement, et ce en toute sécurité. Créez et enregistrez une image corrigée et renforcée d'un serveur virtuel, puis utilisez automatiquement cette image sur chaque nouveau serveur que vous lancez. Créez une architecture de zone de confiance complète, définie et gérée dans un modèle via le contrôle de révision. Automatisez la réponse aux événements habituels et aux événements de sécurité anormaux.

Définition

Les cinq zones de bonnes pratiques en matière de sécurité dans le cloud sont les suivantes :

1. Gestion des identités et des accès
2. Contrôles de détection
3. Protection de l'infrastructure
4. Protection des données
5. Réponse aux incidents

Avant de concevoir l'architecture d'un système, vous devez mettre en place les pratiques qui influent sur la sécurité. Vous voudrez contrôler qui peut faire quoi. En outre, il est important que vous puissiez identifier les incidents de sécurité, protéger vos systèmes et vos services, et assurer la confidentialité et l'intégrité des données via leur protection. Pour répondre aux incidents de sécurité, vous devez disposer d'un processus bien défini et expérimenté. Ces outils et techniques sont importants, parce qu'ils prennent en charge des objectifs tels que la prévention des pertes financières ou la conformité aux obligations réglementaires.

Le modèle Responsabilité partagée AWS permet aux organisations qui adoptent le cloud d'atteindre leurs objectifs de sécurité et de conformité. Comme AWS sécurise physiquement l'infrastructure qui prend en charge nos services cloud, les clients AWS peuvent se concentrer sur l'utilisation de services pour concrétiser leurs objectifs. Le cloud AWS offre aussi un plus grand accès aux données de sécurité, ainsi qu'une approche automatisée pour répondre aux événements de sécurité.

Bonnes pratiques

IAM (Identity and Access Management)

La gestion des identités et des accès constitue un aspect essentiel du programme de sécurité des informations ; elle garantit que seuls les utilisateurs autorisés et authentifiés peuvent accéder à vos ressources, et uniquement de la manière prévue. Par exemple, vous définirez des mandataires (utilisateurs, groupes, services et rôles qui interviennent dans votre compte), déploierez des stratégies alignées sur ces mandataires et implémenterez une gestion robuste des informations d'identification. Ces éléments de gestion des privilèges composent les concepts centraux de l'authentification et de l'autorisation.

Dans AWS, la gestion des privilèges est principalement prise en charge par le service AWS Identity and Access Management (IAM), qui permet aux clients de contrôler l'accès aux ressources et services AWS pour les utilisateurs. Vous pouvez appliquer des politiques détaillées qui attribuent des autorisations à un utilisateur, un groupe, un rôle ou une ressource. Vous avez aussi la possibilité d'exiger des pratiques de mot de passe fort, comme le niveau de complexité, en évitant la réutilisation et en employant Multi-Factor Authentication (MFA). Vous pouvez utiliser la fédération avec votre service d'annuaire existant. Pour les charges de travail qui nécessitent que les systèmes aient accès à AWS, IAM permet un accès sécurisé via les profils d'instance, la fédération d'identité et les informations d'identification temporaires.

Les questions suivantes sont axées sur des considérations relatives à la gestion des privilèges (pour obtenir la liste des questions, réponses et bonnes pratiques relatives à la sécurité, consultez l'annexe).

SEC 1. Comment protégez-vous l'accès aux informations d'identification du compte racine (root) AWS et leur utilisation ?

SEC 2. Comment définissez-vous les rôles et les responsabilités des utilisateurs système pour contrôler l'accès humain à AWS Management Console et aux API ?

SEC 3. Comment limitez-vous l'accès automatique aux ressources AWS ? (par exemple, applications, scripts et/ou outils ou services tiers)

Il est essentiel d'assurer la protection des informations d'identification du compte racine et, à cette fin, AWS recommande d'attacher l'authentification multifacteur (MFA) au compte racine et de verrouiller les informations d'identification avec l'authentification MFA dans un emplacement sécurisé physiquement. Le service IAM vous permet de créer et de gérer d'autres permissions utilisateur (non-racine), ainsi que d'établir des niveaux d'accès aux ressources.

Contrôles de détection

Vous pouvez utiliser les contrôles de détection pour identifier un incident potentiel de sécurité. Ces contrôles constituent une partie essentielle des infrastructures de gouvernance et peuvent être utilisés pour prendre en charge un processus de qualité, une obligation de conformité ou une obligation légale, ainsi que pour l'identification des menaces et les tentatives de réponse. Il existe différents types de contrôles de détection. Par exemple, la conduite d'un inventaire des ressources et de leurs attributs détaillés favorise une prise de décision plus efficace (et les contrôles du cycle de vie) pour contribuer à établir des lignes de base opérationnelles. Ou vous pouvez utiliser un audit interne (examen des contrôles associés aux systèmes d'informations) pour garantir que les pratiques satisfont aux politiques et aux exigences, et que vous avez défini les notifications correctes d'alerte automatique en fonction des conditions définies. Ces contrôles sont des facteurs réactifs importants qui aident les organisations à identifier et à comprendre l'étendue des activités anormales.

Dans AWS, vous pouvez mettre en œuvre des contrôles de détection en traitant les journaux, les événements et les surveillances, et en assurant l'audit, l'analyse automatique et la gestion des alarmes. Les journaux AWS CloudTrail, les appels d'API AWS et Amazon CloudWatch fournissent la surveillance des métriques avec gestion des alarmes, et AWS Config offre l'historique des configurations. Les journaux de niveau service sont également disponibles : par exemple, vous pouvez utiliser Amazon Simple Storage Service (S3) pour enregistrer les demandes d'accès. Enfin, Amazon Glacier propose une fonction de verrouillage de coffre-fort pour préserver les données essentielles à la mission grâce aux contrôles de conformité conçus pour prendre en charge la rétention à long terme vérifiable.

La question suivante porte essentiellement sur les contrôles de détection en matière de sécurité :

SEC 4. Comment capturez-vous et analysez-vous les journaux ?

La gestion des journaux est essentielle dans le cadre d'une conception correctement architecturée, pour des raisons qui vont de la sécurité et de l'expertise judiciaire aux exigences réglementaires ou légales. Il est essentiel que vous analysiez les journaux et leur répondiez afin de pouvoir ainsi identifier les incidents de sécurité potentiels. AWS fournit des fonctionnalités qui simplifient l'implémentation de la gestion des journaux en offrant aux clients la possibilité de définir un cycle de vie de rétention des données, ou de spécifier à quel emplacement les données seront conservées, archivées et/ou supprimées. La gestion des données fiables et prévisibles en devient plus simple et plus économique.

Protection de l'infrastructure

La protection de l'infrastructure englobe les méthodologies de contrôle, comme la protection fiable et l'authentification multifacteur, nécessaires pour satisfaire les bonnes pratiques et les obligations industrielles ou réglementaires. L'utilisation de ces méthodologies est essentielle au succès des opérations en cours, que ce soit dans le cloud ou sur site.

Dans AWS, vous pouvez implémenter l'inspection des paquets avec état et sans état, à l'aide des technologies natives AWS ou de produits et services de partenaires disponibles via AWS Marketplace. Vous pouvez aussi utiliser Amazon Virtual Private Cloud (VPC) pour créer un environnement privé, sécurisé et évolutif, dans lequel vous pouvez définir votre topologie, y compris les passerelles, tables de routage et/ou sous-réseaux privés.

Les questions suivantes portent essentiellement sur la protection de l'infrastructure en matière de sécurité :

SEC 5. Comment appliquez-vous la protection des limites aux niveaux réseau et hôte ?

SEC 6. Comment mettez-vous à profit les fonctions de sécurité au niveau des services AWS ?

SEC 7. Comment protégez-vous l'intégrité des systèmes d'exploitation sur vos instances Amazon EC2 ?

Plusieurs couches de défense sont conseillées dans tout type d'environnement et, dans le cas de la protection de l'infrastructure, la plupart des concepts et méthodes sont valides pour les modèles cloud et locaux. L'application d'une protection des limites et la surveillance des points d'entrée et de sortie, ainsi que la journalisation, la supervision et les alertes, sont toutes essentielles à un plan de sécurité efficace des informations.

Comme évoqué dans la section *Principes de conception* ci-dessus, les clients AWS peuvent personnaliser ou renforcer la configuration d'une instance EC2, et maintenir de façon persistante cette configuration dans un Amazon Machine Image (AMI) immuable. Puis, qu'ils soient déclenchés par Auto Scaling ou lancés manuellement, tous les nouveaux serveurs virtuels (instances) lancés avec cet AMI reçoivent la configuration renforcée.

Protection des données

Avant de concevoir l'architecture d'un quelconque système, les pratiques de base qui influent sur la sécurité doivent être en place. Par exemple, la *classification des données* fournit un moyen de classifier les données organisationnelles en fonction des niveaux de sensibilité et le *chiffrement* protège les données en les rendant inintelligibles en cas d'accès non autorisé. Ces outils et techniques sont importants, parce qu'ils prennent en charge des objectifs tels que la prévention des pertes financières ou la conformité aux obligations réglementaires.

Dans AWS, les pratiques suivantes facilitent la protection des données :

- Les clients AWS conservent le contrôle intégral de leurs données.
- AWS vous permet de chiffrer vos données et de gérer vos clés plus facilement, rotation régulière des clés incluse, ce qui peut être facilement automatisé en mode natif par AWS ou assuré par un client.
- La journalisation détaillée est disponible et contient des informations importantes, telles que les accès aux fichiers et les modifications.
- AWS a conçu les systèmes de stockage pour une résilience exceptionnelle. A titre d'exemple, Amazon Simple Storage Service (S3) est conçu pour une durabilité de 99,999999999 % (« onze-neuf »). (Par exemple, si vous stockez 10 000 objets avec Amazon S3, vous pouvez en moyenne vous attendre à la perte d'un objet tous les 10 000 000 ans.)
- Le versioning, qui peut faire partie d'un processus de gestion du cycle de vie des données plus étendu, assure une protection contre les remplacements ou suppressions accidentels, et dommages similaires.
- AWS n'initie jamais de mouvement de données entre régions. Le contenu affecté à une région demeure dans celle-ci jusqu'à ce que le client active explicitement une fonction ou exploite un service qui fournit une telle fonctionnalité.

Les questions suivantes portent essentiellement sur la sécurité des données :

SEC 8. Comment classez-vous les données ?

SEC 9. Comment chiffrez-vous et protégez-vous vos données au repos ?

SEC 10. Comment gérez-vous les clés ?

SEC 11. Comment chiffrez-vous et protégez-vous vos données en transit ?

AWS fournit également plusieurs options pour chiffrer les données au repos et en transit. Nous intégrons à nos produits et services des fonctionnalités qui facilitent le chiffrement de vos données. Par exemple, nous avons implémenté le chiffrement côté serveur pour Amazon S3 afin de vous permettre de stocker plus facilement vos données d'une manière chiffrée. Vous pouvez aussi prendre les dispositions nécessaires pour que la totalité du processus de chiffrement et déchiffrement HTTPS (généralement appelé terminaison SSL) soit gérée par Elastic Load Balancing.

Réponse aux incidents

Même avec des contrôles de prévention et de détection extrêmement mûrs, les organisations doivent continuer à mettre en place des processus pour répondre aux incidents ayant un impact potentiel sur la sécurité et les atténuer. L'architecture de votre charge de travail affecte fortement la possibilité de vos équipes à fonctionner efficacement pendant un incident afin d'isoler ou de contenir les systèmes, et de rétablir les opérations à un état correct connu. La mise en place des outils et des accès préalablement à tout incident de sécurité, puis la pratique régulière de réponse aux incidents garantira que l'architecture est mise à jour pour accueillir un examen et une récupération opportuns.

Dans AWS, les pratiques suivantes facilitent la réponse efficace aux incidents :

- La journalisation détaillée est disponible et contient des informations importantes, telles que les accès aux fichiers et les modifications.
- Les événements peuvent être automatiquement traités et déclencher les scripts qui automatisent les runbooks au travers de l'utilisation des API AWS.
- Vous pouvez pré-provisionner les outils et un « espace propre » avec AWS CloudFormation. Cela vous permet de procéder à une expertise dans un environnement sécurisé et isolé.

Les questions suivantes portent essentiellement sur la réponse aux incidents :

SEC 12. Comment vous assurez-vous que vous disposez de la réponse aux incidents appropriée ?

Assurez-vous que vous avez un moyen d'accorder rapidement l'accès à votre équipe InfoSec, et automatisez l'isolation des instances aussi bien que la capture des données et des états pour l'analyse experte.

Services AWS clés

Le service AWS essentiel à la sécurité est AWS Identity and Access Management (IAM), qui vous permet de contrôler de façon sécurisée l'accès aux ressources et services AWS pour vos utilisateurs. Les services et fonctions suivants prennent en charge les quatre zones de sécurité :

Identity and Access Management : IAM vous permet de contrôler en toute sécurité l'accès aux services et ressources AWS. L'authentification multifacteur (MFA) fournit un niveau de sécurité supplémentaire par-dessus votre nom d'utilisateur et votre mot de passe.

Contrôles de détection : AWS CloudTrail enregistre les appels d'API AWS, AWS Config fournit un inventaire détaillé de votre configuration et de vos ressources AWS, et Amazon CloudWatch est un service de supervision des ressources AWS.

Protection de l'infrastructure : Amazon Virtual Private Cloud (VPC) vous permet de mettre en service une section isolée et privée du cloud AWS où vous pouvez lancer les ressources AWS dans un réseau virtuel.

Protection des données : les services tels qu'Elastic Load Balancing, Amazon Elastic Block Store (EBS), Amazon Simple Storage Service (S3) et Amazon Relational Database Service (RDS) incluent les capacités de chiffrement pour protéger vos données en transit et au repos. AWS Key Management Service (KMS) permet aux clients de créer et de contrôler plus facilement les clés utilisées pour le chiffrement.

Réponse aux incidents : IAM doit être utilisé pour accorder les autorisations appropriées aux équipes de réponse aux incidents. Amazon CloudFormation peut être utilisé pour créer un environnement fiable afin de conduire les investigations.

Ressources

Consultez les ressources suivantes pour en savoir plus sur nos bonnes pratiques de sécurité.

Documentation et blogs

- Centre de sécurité AWS
- Conformité AWS
- Blog sur la sécurité AWS

Livres blancs

- Présentation de la sécurité AWS
- Bonnes pratiques de sécurité AWS
- Risque et conformité AWS

Vidéos

- Security of the AWS Cloud (Sécurité du cloud AWS)
- Shared Responsibility Overview (Présentation de la responsabilité partagée)

Pilier « Fiabilité »

Le pilier **Fiabilité** englobe la possibilité d'un système de récupérer à partir de perturbations de l'infrastructure ou d'un service, d'acquérir dynamiquement les ressources de calcul pour satisfaire à la demande et d'atténuer les perturbations telles que les erreurs de configuration ou les problèmes réseau temporaires.

Principes de conception

Dans le cloud, il existe un certain nombre de principes qui peuvent vous aider à accroître la fiabilité :

- **Tester les procédures de récupération :** dans un environnement local, les tests sont souvent conduits pour prouver que le système fonctionne dans un scénario particulier ; les tests ne sont généralement pas utilisés pour valider les stratégies de récupération. Dans le cloud, vous pouvez tester de quelle façon votre système échoue et valider vos procédures de récupération. Vous pouvez utiliser l'automatisation pour simuler différentes défaillances ou recréer les scénarios qui y ont conduit précédemment. Cela expose les chemins de défaillance que vous pouvez tester et rectifier *avant* un scénario de défaillance réelle, en réduisant le risque d'échec de composants qui n'ont pas été testés avant.
- **Récupération automatique après incident :** en surveillant un système pour les indicateurs de performance clé, vous pouvez déclencher l'automatisation en cas de violation d'un seuil. Cela permet une notification automatique et un suivi des défaillances, et les processus de récupération automatique qui contournent ou réparent la défaillance. Avec une automatisation sophistiquée, il est possible d'anticiper et de corriger les défaillances avant qu'elles ne se produisent.
- **Mise à l'échelle horizontale pour augmenter la disponibilité cumulée du système :** remplacez une ressource volumineuse par plusieurs petites ressources pour réduire l'impact d'une défaillance unique sur le système global. Répartissez les demandes entre plusieurs ressources plus petites pour garantir qu'elles ne partagent pas un point de défaillance commun.

- **Arrêt de la capacité d'estimation :** une cause courante de défaillance des systèmes locaux est celle de la saturation des ressources, quand les demandes placées sur un système dépassent la capacité de ce système (tel est souvent l'objectif des attaques par déni de service). Dans le cloud, vous pouvez surveiller la demande et l'utilisation du système, et automatiser l'ajout ou la suppression de ressources afin de maintenir le niveau optimal de satisfaction de la demande sans sur-allocation ou sous-allocation.
- **Gérer les modifications par l'automatisation** : les modifications de votre infrastructure doivent s'effectuer via l'automatisation. Les changements qui doivent être gérés sont ceux apportés à l'automatisation.

Définition

Les trois zones de bonnes pratiques en matière de fiabilité dans le cloud sont les suivantes :

1. Fondations
2. Gestion des modifications
3. Gestion des défaillances

Pour parvenir à la fiabilité, un système doit avoir en place une fondation et une supervision correctement planifiées, avec les mécanismes pour gérer les modifications en matière de demande ou d'exigence. Le système doit être conçu pour détecter les défaillances et se réparer automatiquement.

Bonnes pratiques

Fondations

Avant de concevoir l'architecture d'un quelconque système, les exigences qui influent sur la fiabilité doivent être en place. Par exemple, vous devez avoir une bande passante réseau suffisante pour votre centre de données. Ces exigences sont parfois négligées (parce qu'elles sont au-delà de la simple portée d'un projet). Cette négligence peut avoir un impact significatif sur la capacité à proposer un système fiable. Dans un environnement local, ces exigences peuvent entraîner de longs délais d'attente en raison des dépendances et, par conséquent, doivent être intégrées lors de la planification initiale.

Avec AWS, le plupart des exigences en matière de fondation sont déjà intégrées ou peuvent être satisfaites en fonction des besoins. Le cloud étant conçu pour être fondamentalement illimité, il est de la responsabilité d'AWS de satisfaire l'exigence de capacités suffisantes de réseau et de calcul, tandis que vous êtes libre de modifier la taille et l'allocation des ressources, comme la taille des dispositifs de stockage, à la demande.

Les questions suivantes portent sur les considérations essentielles relatives à la fiabilité :

FIA 1. Comment gérez-vous les Service Limits AWS pour vos comptes ?

FIA 2. Comment planifiez-vous la topologie de votre réseau sur AWS ?

AWS définit des limites de service (limite supérieure de la quantité que votre équipe peut demander pour chaque ressource) pour vous protéger d'une sur-allocation accidentelle des ressources. Vous devez avoir la gouvernance et les processus en place pour surveiller et modifier ces limites en fonction de vos besoins métier. Lorsque vous choisissez le cloud, il se peut que vous ayez besoin de planifier l'intégration aux ressources locales existantes (approche hybride). Un modèle hybride permet la transition progressive vers une approche cloud tout en un au fil du temps ; par conséquent, il est important que vous disposiez d'une conception de la façon dont AWS et vos ressources locales interagiront en tant que topologie réseau.

Gestion des modifications

Le fait d'être conscient de la façon dont le changement affecte un système vous permet une planification proactive, tandis que la supervision vous permet d'identifier rapidement les tendances qui pourraient conduire à des problèmes de capacité ou à des violations de contrat de niveau de service (SLA). Dans les environnements traditionnels, les processus de contrôle des modifications sont souvent manuels et doivent être soigneusement coordonnés ave l'audit pour contrôler efficacement les personnes autorisées à effectuer des modifications et à quel moment.

Avec AWS, vous pouvez surveiller le comportement d'un système et automatiser la réponse aux indicateurs de performance clé, par exemple en ajoutant des serveurs au fur et à mesure qu'un système gagne de nouveaux utilisateurs. Vous pouvez contrôler les personnes qui ont l'autorisation d'apporter des modifications au système et d'auditer l'historique de ces modifications.

Les questions suivantes portent essentiellement sur les considérations relatives aux modifications en matière de fiabilité :

FIA 3. Comment surveillez-vous les ressources AWS ?

FIA 4. Comment votre système s'adapte-t-il aux modifications à la demande ?

FIA 5. Comment exécutez-vous les modifications ?

Lorsque vous concevez l'architecture d'un système pour ajouter ou supprimer automatiquement des ressources en réponse à des modifications à la demande, cela accroît non seulement la fiabilité, mais garantit aussi que la réussite commerciale ne devient pas un poids. Avec la supervision en place, votre équipe est automatiquement avertie quand les indicateurs de performance clé s'écartent des normes attendues. La journalisation automatique des modifications apportées à votre environnement vous permet d'auditer et d'identifier rapidement les actions susceptibles d'avoir un impact sur la fiabilité. Les contrôles de la gestion des modifications assurent que vous appliquez les règles offrant la fiabilité dont vous avez besoin.

Gestion des défaillances

Dans un système de complexité raisonnable, il est attendu que des défaillances se produisent et il est généralement intéressant de savoir comment devenir conscient de ces échecs, y répondre et empêcher qu'ils ne se renouvellent.

Dans AWS, nous mettons à profit l'automatisation pour réagir aux données de supervision. Par exemple, lorsqu'une métrique particulière franchit un seuil, vous pouvez déclencher une action automatique pour corriger le problème. De même, plutôt que de tenter de diagnostiquer et de corriger une ressource défaillante qui fait partie de votre environnement de production, vous pouvez la remplacer par une nouvelle ressource et exécuter l'analyse de cette ressource hors bande. Comme le cloud vous permet de maintenir les versions temporaires d'un système complet à bas coût, vous pouvez utiliser les tests automatiques pour vérifier les processus complets de récupération.

Les questions suivantes portent essentiellement sur la gestion des défaillances en termes de fiabilité :

FIA 6. Comment sauvegardez-vous les données ?

FIA 7. Comment votre système supporte-t-il les défaillances de composants ?

FIA 8. Comment testez-vous la résilience ?

FIA 9. Comment planifiez-vous la reprise après sinistre ?

Sauvegardez régulièrement vos données et testez vos fichiers de sauvegarde pour vous assurer de pouvoir récupérer aussi bien à partir d'erreurs logiques que d'erreurs physiques. La clé de la gestion des défaillances réside dans des tests réguliers et automatiques des systèmes par le biais d'échecs et de récupérations (idéalement selon un planning régulier et déclenché également après des modifications significatives du système). Suivez activement les indicateurs de performance clé, tels que l'objectif de délai de récupération et l'objectif de point de récupération, pour évaluer la résilience d'un système (notamment dans les scénarios de test de défaillance). Un tel suivi vous aide à identifier et à atténuer les points uniques de défaillance. L'objectif est de tester intégralement vos processus de récupération système de telle sorte que vous soyez assuré de récupérer l'ensemble de vos données et de continuer à servir vos clients, même en présence de problèmes continus. Vos processus de récupération doivent être aussi bien maîtrisés que vos processus normaux de production.

Services AWS clés

Le service AWS qui constitue la clé pour garantir la fiabilité est Amazon CloudWatch, qui contrôle les métriques en temps réel. Les autres services et fonctions qui prennent en charge les trois zones de la fiabilité sont les suivants :

Fondations : AWS Identity and Access Management (IAM) vous permet de contrôler en toute sécurité l'accès aux services et ressources AWS. Amazon VPC vous permet de mettre en service une section isolée et privée du cloud AWS, où vous pouvez lancer les ressources AWS au sein d'un réseau virtuel.

Gestion des modifications : AWS CloudTrail enregistre les appels d'API AWS pour votre compte et vous délivre les fichiers journaux à des fins d'audit. AWS Config fournit un inventaire détaillé de votre configuration et de vos ressources AWS, et enregistre continuellement les changements de configuration.

Gestion des défaillances : AWS CloudFormation fournit des modèles pour la création de ressources AWS et les alloue d'une manière ordonnée et prévisible.

Ressources

Consultez les ressources suivantes pour en savoir plus sur nos bonnes pratiques en matière de fiabilité.

Vidéo et rapport d'analyse

- Embracing Failure : Fault-Injection and Service Reliability (Prise en compte des défaillances : injection d'erreurs et fiabilité des services)
- Benchmarking Availability and Reliability in the Cloud (Test de disponibilité et de fiabilité dans le cloud)

Documentation et blogs

- Service Limits
- Blog consacré aux rapports des Service Limits

Livres blancs

- Approches de sauvegarde, d'archivage et de restauration avec AWS
- Gestion de votre infrastructure AWS à l'échelle
- Reprise après sinistre AWS
- Options de connectivité AWS Amazon VPC

AWS Support

- AWS Premium Support
- Trusted Advisor

Pilier « Efficacité des performances »

Le pilier **Efficacité des performances** se concentre sur l'utilisation efficace des ressources de calcul pour répondre aux exigences et sur le maintien de cette efficacité au fur et à mesure que la demande change et que les technologies évoluent.

Principes de conception

Dans le cloud, il existe un certain nombre de principes qui peuvent vous aider à garantir l'efficacité des performances :

- **Démocratiser les technologies avancées :** les technologies difficiles à implémenter peuvent devenir plus faciles à utiliser en intégrant cette connaissance et cette complexité dans le domaine du fournisseur du cloud. Au lieu que votre équipe informatique apprenne à héberger et exécuter une nouvelle technologie, elle peut simplement l'utiliser comme service. Par exemple, les bases de données NoSQL, le transcodage multimédia et le Machine Learning sont trois technologies requérant une expertise qui n'est pas répartie également au sein de la communauté technique. Dans le cloud, ces technologies deviennent des services que votre équipe peut utiliser tout en se concentrant sur le développement du produit plutôt que sur l'allocation et la gestion des ressources.
- **Portée mondiale en quelques minutes :** déployez aisément votre système dans plusieurs régions du monde en quelques clics à peine. Vous pourrez ainsi offrir à vos clients une latence plus faible et une meilleure expérience à un coût minimal.
- **Utilisation des architectures *sans serveur* :** dans le cloud, les architectures sans serveur suppriment la nécessité d'exécuter et de gérer des serveurs pour effectuer les activités traditionnelles de calcul. Par exemple, les services de stockage peuvent faire office de sites web statiques et supprimer la nécessité de services web, tandis que les services d'événements peuvent héberger automatiquement le code. Cela supprime non seulement le poids opérationnel lié à la gestion de ces serveurs, mais peut aussi réduire les coûts des transactions, car ces services gérés œuvrent à l'échelle du cloud.

- **Expérimentation plus fréquente :** avec les ressources virtuelles et automatisables, vous pouvez rapidement exécuter des tests comparatifs à l'aide de différents types d'instances, stockages ou configurations.
- **Sympathie mécanique :** utilisez l'approche technologique qui s'aligne le mieux sur ce que vous tentez d'obtenir. Par exemple, pensez aux modèles d'accès aux données lorsque vous sélectionnez les approches base de données ou stockage.

Définition

Les quatre zones de bonnes pratiques en matière d'efficacité des performances dans le cloud sont les suivantes :

1. Sélection (calcul, stockage, base de données, réseau)
2. Révision
3. Supervision
4. Compromis

Adoptez une approche orientée données pour sélectionner une architecture hautes performances. Recueillez les données sur tous les aspects de l'architecture, depuis la conception de haut niveau jusqu'à la sélection et la configuration des types de ressource. En vérifiant vos choix selon une base cyclique, vous vous assurez de tirer profit de la plateforme AWS en constante évolution. La supervision vous garantit d'être conscient de tout écart par rapport aux performances attendues et de prendre les mesures nécessaires. Enfin, votre architecture peut établir des compromis pour améliorer les performances, tels que la compression ou la mise en cache, ou l'assouplissement des exigences de cohérence.

Bonnes pratiques

Sélection

La solution optimale d'un système particulier varie en fonction de votre type de charge de travail, souvent avec plusieurs approches combinées. Les systèmes correctement architecturés utilisent plusieurs solutions et autorisent différentes fonctions pour améliorer les performances.

Dans AWS, les ressources sont virtualisées et disponibles dans un certain nombre de types et de configuration différents. Il est ainsi plus facile de trouver une approche qui correspond à vos besoins ; vous pouvez également rechercher des options qui ne sont pas facilement accessibles avec une infrastructure locale. Par exemple, un service géré tel qu'Amazon DynamoDB fournit une base de données NoSQL entièrement gérée, qui offre, quelle que soit l'échelle, une latence inférieure à 10 millisecondes.

Les questions suivantes portent essentiellement sur la sélection :

PERF 1. Comment sélectionner l'architecture aux meilleures performances ?

Lorsque vous sélectionnez les modèles et l'implémentation de votre architecture, utilisez une approche pilotée par les données pour bénéficier de la solution la plus optimale. AWS Solutions Architects, AWS Reference Architectures et AWS Partners peuvent vous aider à sélectionner une architecture basée sur ce que vous avez appris, mais les données obtenues via les comparaisons ou le test de charge seront requises pour optimiser votre architecture.

Votre architecture combinera probablement un certain nombre de différentes approches architecturales (pilotage par les événement, ETL ou pipeline, par exemple). L'implémentation de votre architecture utilisera les services AWS qui sont spécifiques à l'optimisation des performances de votre architecture. Dans les sections suivantes, nous examinerons les quatre principaux types de ressources que vous devez prendre en compte (calcul, stockage, base de données et réseau).

Calcul

La solution de calcul optimale d'un système particulier peut varier selon la conception de l'application, les modèles d'utilisation et les paramètres de configuration. Les architectures peuvent utiliser différentes solutions de calcul pour divers composants et activer différentes fonctions pour améliorer les performances. La sélection d'une solution de calcul incorrecte pour une architecture peut conduire à une efficacité moindre des performances.

Dans AWS, le calcul est disponible sous trois formes : instances, conteneurs et fonctions :

- Les **instances** sont des serveurs virtualisés et, par conséquent, vous pouvez modifier les capacités à l'aide d'un simple clic sur un bouton ou d'un appel d'API. Comme dans le cloud les décisions relatives aux ressources ne sont plus fixes, vous pouvez expérimenter différents types de serveur. Dans AWS, ces *instances* de serveur virtuel se présentent selon différentes tailles et familles, et offrent ainsi une large variété de capacités, telles que disques SSD et unités GPU.
- Les **conteneurs** sont une méthode de virtualisation de système d'exploitation qui vous permet d'exécuter une application et ses dépendances dans des processus isolés des ressources.
- Les **fonctions** extraient l'environnement d'exécution du code que vous voulez exécuter. Par exemple, AWS Lambda vous permet d'exécuter du code sans exécuter une instance.

Les exemples de questions suivantes portent essentiellement sur le calcul :

PERF 2. Comment sélectionnez-vous la solution de calcul ?

Lors de la conception de l'architecture de votre solution de calcul, vous tirez parti des mécanismes d'élasticité disponibles pour garantir que vous avez la capacité suffisante pour soutenir les performances alors que la demande évolue.

Stockage

La solution de stockage optimale pour un système particulier varie en fonction du type de méthode d'accès (bloc, fichier ou objet), des modèles d'accès (aléatoire ou séquentiel), du débit requis, de la fréquence des accès (en ligne, hors connexion, archivage), de la fréquence des mises à jour (WORM, dynamiques) et des contraintes de disponibilité et de durabilité. Les systèmes correctement architecturés utilisent plusieurs solutions de stockage et autorisent différentes fonctions pour améliorer les performances.

Dans AWS, le stockage est virtualisé et disponible dans un certain nombre de types différents. Il est ainsi plus facile de faire correspondre plus étroitement vos méthodes de stockage à vos besoins et de proposer des options de stockage qui ne sont pas facilement accessibles avec une infrastructure locale. Par exemple, Amazon Simple Storage Service (S3) est conçu pour une durabilité de 99,999999999 % (lisez « onze-neuf »). Vous pouvez aussi passer des disques HDD aux disques SSD et déplacer sans peine les disques virtuels d'une instance à l'autre en quelques secondes.

Les questions suivantes se concentrent essentiellement sur le stockage en termes d'efficacité des performances :

PERF 3. Comment sélectionnez-vous la solution de stockage ?

Lorsque vous sélectionnez une solution de stockage, s'assurer qu'elle s'aligne sur vos modèles d'accès sera essentiel pour parvenir aux performances que vous souhaitez.

Base de données

La solution de base de données optimale pour un système particulier peut varier en fonction des exigences de disponibilité, de cohérence, de tolérance des partitions, de latence, de durabilité, d'évolutivité et de capacité des requêtes. De nombreux systèmes utilisent différentes solutions de base de données pour divers sous-systèmes et activent différentes fonctions pour améliorer les performances. La sélection d'une solution de base de données et de fonctionnalités incorrectes pour un système peut conduire à une efficacité moindre des performances.

Dans AWS, Amazon Relational Database Service (RDS) fournit une base de données relationnelle entièrement gérée. Avec Amazon RDS, vous pouvez dimensionner les ressources de calcul et de stockage de votre base de données, souvent sans aucune interruption. Amazon DynamoDB est une base de données NoSQL entièrement gérée, qui offre, quelle que soit l'échelle, une latence inférieure à 10 millisecondes. Amazon Redshift est un entrepôt de données géré d'une capacité de plusieurs Po (pétaoctets), qui permet de changer le nombre ou le type de nœuds au fur et à mesure que vos performances ou besoins en capacité évoluent.

Les questions suivantes se concentrent essentiellement sur les bases de données en termes d'efficacité des performances :

PERF 4. Comment sélectionnez-vous votre solution de base de données ?

Même si l'approche de base de données (RDBMS, NoSQL, etc.) d'une charge de travail a un impact réel sur l'efficacité des performances, il s'agit souvent d'un domaine choisi conformément aux valeurs par défaut de l'organisation plutôt qu'au travers d'une approche pilotée par les données. De même que pour le stockage, il est essentiel de prendre en compte les modèles d'accès de votre charge de travail, ainsi que le fait que d'autres solutions de bases de données puissent résoudre le problème plus efficacement (comme l'utilisation d'un moteur de recherche ou d'un entrepôt de données).

Réseau

La solution de réseau optimale pour un système particulier varie en fonction des exigences de latence et de débit, entre autres. Les contraintes physiques, telles que les ressources utilisateur ou les ressources locales, pilotent les options d'emplacement, qui peuvent être décalées à l'aide de techniques périphériques ou d'emplacement des ressources.

Dans AWS, le réseau est virtualisé et disponible dans un certain nombre de types et de configuration différents. Cela facilite une correspondance plus étroite entre vos besoins et vos méthodes de mise en réseau. AWS propose des fonctionnalités produit (types d'instance réseau très élevés, instances optimisées Amazon EBS, Amazon S3 Transfer Acceleration, Amazon CloudFront dynamique, par exemple) pour optimiser le trafic réseau. AWS propose également des fonctionnalités réseau (routage de latence Amazon Route53, points de terminaison Amazon VPC et AWS Direct Connect, par exemple) afin de réduire la distance ou l'instabilité réseau.

Les questions suivantes se concentrent essentiellement sur le stockage en termes d'efficacité des performances :

PERF 5. Comment sélectionnez-vous votre solution réseau ?

Lorsque vous sélectionnez votre solution réseau, vous devez prendre en compte l'emplacement. Avec AWS, vous pouvez choisir de placer les ressources près de l'emplacement où elles seront utilisées afin de réduire la distance. En tirant parti des régions, groupes de placement et emplacements périphériques, vous pouvez améliorer les performances de façon significative.

Révision

Lors de la conception architecturale des solutions, vous avez la possibilité de choisir au sein d'un ensemble fini d'options. Cependant, au fil du temps de nouvelles technologies et approches deviennent disponibles qui peuvent améliorer les performances de votre architecture.

Avec AWS, vous pouvez tirer parti de votre innovation continuelle, qui est pilotée par le besoin client. Nous proposons régulièrement de nouveaux emplacements périphériques, régions, services et fonctions. N'importe lequel d'entre eux peut améliorer positivement l'efficacité des performances de votre architecture.

Les questions suivantes se concentrent essentiellement sur la vérification de l'efficacité des performances :

PERF 6. Comment savez-vous que vous continuez à avoir le type de ressource le plus approprié tandis que de nouveaux types de ressource et de nouvelles fonctionnalités sont introduits ?

Comprendre où votre architecture est contrainte par les performances vous permet d'examiner les futures versions à même d'alléger cette contrainte.

Supervision

Une fois que vous avez implémenté votre architecture, vous devez surveiller ses performances de telle sorte que vous puissiez corriger les problèmes éventuels avant que vos clients n'en soient conscients. La surveillance des métriques doit être utilisée pour déclencher des alarmes en cas de dépassement des seuils. L'alarme peut déclencher une action automatique pour contourner les composants aux performances médiocres.

Avec AWS, Amazon CloudWatch offre la possibilité de superviser et d'envoyer des alarmes de notification ; vous pouvez aussi utiliser l'automatisation pour contourner les problèmes de performance et déclencher les actions nécessaires via Amazon Kinesis, Amazon Simple Queue Service (SQS) et AWS Lambda.

Les questions suivantes se concentrent essentiellement sur la surveillance de l'efficacité des performances :

PERF 7. Comment surveillez-vous les ressources après leur lancement pour vous assurer qu'elles se comportent comme prévu ?

S'assurer que vous ne voyez pas un trop grand nombre de faux positifs, ou que vous ne soyez pas surchargé de données, est la clé pour obtenir une solution de supervision efficace. Les déclencheurs automatiques évitent l'erreur humaine et peuvent réduire le temps de correction des problèmes. Planifiez des « game days » où les simulations sont conduites dans l'environnement de production, pour tester votre solution de gestion des alarmes et garantir qu'elle identifie correctement les problèmes.

Compromis

Lorsque vous concevez l'architecture de vos solutions, pensez à procéder à des compromis afin que vous puissiez sélectionner une approche optimale. En fonction de votre situation, vous pouvez négocier entre la cohérence, la durabilité et l'espace d'un côté, et le temps ou la latence de l'autre, pour offrir des performances plus élevées.

Avec AWS, vous pouvez atteindre une portée mondiale en quelques minutes et déployer les ressources dans plusieurs emplacements à travers le monde pour être plus proches de vos utilisateurs finaux. Vous pouvez aussi ajouter dynamiquement des réplicas en lecture seule aux banques d'informations telles que les bases de données afin de réduire la charge sur la base de données principale. AWS propose aussi des solutions de mise en cache, telles qu'Amazon ElastiCache, qui fournit un magasin de données en mémoire ou cache, et Amazon CloudFront, qui met en cache les copies de votre contenu statique et les rapproche des utilisateurs finaux.

Les questions suivantes portent essentiellement sur les compromis espace-temps en termes d'efficacité des performances :

PERF 8. Comment puis-je utiliser les compromis pour améliorer les performances ?

Les compromis peuvent améliorer la complexité de votre architecture et nécessiter un test de charge pour garantir qu'un avantage mesurable est obtenu.

Services AWS clés

Le principal service AWS pour l'efficacité des performances est Amazon CloudWatch, qui surveille vos ressources et systèmes, en offrant une visibilité de vos performances globales et de votre état de fonctionnement. Les services suivants sont importants dans les domaines d'efficacité des performances suivants :

Sélection :

Calcul : Auto Scaling est essentiel pour vous garantir que vous avez assez d'instances pour satisfaire la demande et préserver la réactivité.

Stockage : Amazon EBS offre un large éventail d'options de stockage (telles que SSD et PIOPS) qui vous permettent d'optimiser votre cas d'utilisation. Amazon S3 offre une livraison de contenu sans serveur et Amazon S3 Transfer Acceleration permet un transfert rapide, facile et sécurisé de fichiers sur des longues distances.

Base de données : Amazon RDS offre un large éventail de fonctions de base de données (telles que les IOPS provisionnées et les réplicas en lecture) qui vous permettent d'optimiser votre cas d'utilisation. Amazon DynamoDB offre, quelle que soit l'échelle, une latence inférieure à 10 millisecondes.

Réseau : Amazon Route 53 fournit un routage basé sur la latence. Les points de terminaison Amazon VPC et Direct peuvent réduire la distance ou l'instabilité réseau.

Révision : le blog AWS et la section Nouveautés du site web AWS constituent des ressources d'information relatives aux fonctionnalités et services récemment lancés.

Surveillance : Amazon CloudWatch fournit les métriques, alarmes et notifications que vous pouvez intégrer à votre solution de supervision existante, et que vous pouvez utiliser avec AWS Lambda pour déclencher les actions.

Compromis : Amazon ElastiCache, Amazon CloudFront et AWS Snowball sont des services qui vous permettent d'améliorer les performances. Les réplicas en lecture dans Amazon RDS peuvent vous permettre de dimensionner les charges de travail à lecture intensive.

Ressources

Consultez les ressources suivantes pour en savoir plus sur nos bonnes pratiques en matière d'efficacité des performances.

Vidéos

- Performance Channel (Canal de performances)
- Performance Benchmarking on AWS (Comparaison des performances sur AWS)

Documentation

- Optimisation des performances Amazon S3
- Performances des volume Amazon EBS

Pilier « Optimisation des coûts »

Le pilier **Optimisation des coûts** englobe le processus continuel d'affinage et d'amélioration d'un système pendant la totalité de son cycle de vie. Depuis la conception initiale de votre toute première preuve de concept (Proof of Concept) jusqu'au fonctionnement continu des charges de travail en production, l'adoption des pratiques du présent livre blanc vous permet de créer et de faire fonctionner des systèmes rentables qui atteignent les résultats commerciaux escomptés et réduisent les coûts, permettant ainsi à votre activité d'optimiser son retour sur investissement.

Principes de conception

Dans le cloud, il existe un certain nombre de principes qui peuvent vous aider à atteindre l'optimisation des coûts :

- **Adopter un modèle de consommation :** ne payez que les ressources de calcul que vous consommez, et augmentez ou diminuez l'utilisation en fonction de vos exigences métier, et non en utilisant des prévisions élaborées. Par exemple, les environnements de développement et de test sont généralement utilisés uniquement huit heures par jour, pendant les semaines ouvrées. Vous pouvez interrompre ces ressources lorsqu'elles sont inutilisées afin de réaliser des économies pouvant atteindre jusqu'à 75 % (40 heures au lieu de 168).
- **Bénéficier d'économies d'échelle :** grâce au cloud computing, vous pouvez parvenir à un coût variable moindre que par vous-même, car AWS permet des économies d'échelle supérieures. Des centaines de milliers de clients sont regroupés dans le cloud AWS, ce qui se traduit par des prix de facturation à l'utilisation inférieurs.
- **Cesser de dépenser de l'argent sur les opérations des centres de données :** comme AWS a la lourde charge de monter les serveurs en rack, de les empiler et de les alimenter, vous pouvez vous concentrer sur vos clients et sur les projets métier plutôt que sur l'infrastructure informatique.
- **Analyser et attribuer les dépenses :** le cloud facilite l'identification précise de l'utilisation et du coût des systèmes, ce qui permet ensuite l'attribution transparente des coûts informatiques à leurs détenteurs métier. La mesure du retour sur investissement en est simplifiée et, par conséquent, offre aux détenteurs de systèmes l'opportunité d'optimiser leurs ressources et de réduire leurs coûts.
- **Utilisation des services gérés pour réduire le coût de possession :** dans le cloud, les services gérés suppriment la charge opérationnelle de maintenance de serveurs pour des tâches telles que l'envoi de courriers électroniques ou la gestion de bases de données. Et, comme les services gérés interviennent à l'échelle du cloud, ils peuvent offrir un coût moindre par transaction ou service.

Définition

Les quatre zones de bonnes pratiques en matière d'optimisation des coûts dans le cloud sont les suivantes :

1. Ressources économiques
2. Correspondance de l'offre et de la demande
3. Sensibilisation aux dépenses
4. Optimisation au fil du temps

Comme pour les autres piliers, il existe d'autres compromis à prendre en compte. Par exemple, voulez-vous optimiser pour accélérer la mise sur le marché ou pour des raisons de coût ? Dans certains cas, il est préférable d'optimiser la vitesse, avec une mise sur le marché rapide, la livraison de nouvelles fonctions ou le simple respect d'une échéance, plutôt que d'investir dans une optimisation des coûts initiaux. Les décisions de conception sont parfois guidées par la précipitation, par opposition aux données empiriques, car la tentation est toujours présente de surcompenser « juste au cas où », plutôt que de consacrer du temps à des essais comparatifs pour déterminer le déploiement le plus optimal en termes de coût. Cela conduit souvent à des déploiements incroyablement sur-provisionnés et sous-optimisés. Les sections suivantes fournissent des techniques et des conseils stratégiques en matière d'optimisation initiale et continue des coûts de votre déploiement.

Bonnes pratiques

Ressources économiques

L'utilisation des instances et ressources appropriées de votre système constitue la clé des économies de coût. Par exemple, un processus de reporting peut nécessiter jusqu'à cinq heures pour s'exécuter sur un serveur, mais un plus grand serveur deux fois plus cher n'aura besoin que d'une heure. Le résultat sera le même dans les deux cas, mais le plus petit serveur entraînera un coût plus élevé au fil du temps.

Un système correctement architecturé utilise les ressources les plus rentables, ce qui peut avoir un impact économique positif et significatif. Vous avez aussi l'opportunité d'utiliser les services gérés pour réduire les coûts. Par exemple, plutôt que de maintenir des serveurs pour remettre les courriers électroniques, vous pouvez utiliser un service qui facture par message.

AWS propose une grande variété d'options de tarification flexibles et économiques pour acquérir les instances Amazon EC2 de la façon qui correspond le mieux à vos besoins. Les *instances à la demande* vous permettent de payer la capacité de calcul à l'heure, sans aucun engagement minimum requis. Les *instances réservées* vous permettent de réserver des capacités et offrent des économies pouvant atteindre 75 % de la tarification à la demande. Avec les *instances ponctuelles*, vous pouvez faire une offre sur la capacité non utilisée d'Amazon EC2 avec des remises significatives. Les instances ponctuelles conviennent quand le système peut tolérer l'utilisation d'une flotte de serveurs où les serveurs individuels peuvent aller et venir dynamiquement, comme lors de l'utilisation de HPC et du Big Data.

Les questions suivantes portent essentiellement sur la sélection de ressources économiques à des fins d'optimisation des coûts :

COUT 1. Lorsque vous sélectionnez les services AWS pour votre solution, prenez-vous le coût en compte ?

COUT 2. Avez-vous dimensionné les ressources pour satisfaire vos cibles de coût ?

COUT 3. Avez-vous sélectionné le modèle de tarification approprié pour satisfaire vos cibles de coût ?

A l'aide d'outils tels que AWS Trusted Advisor qui permettent de contrôler régulièrement votre utilisation d'AWS, vous pouvez surveiller activement cette dernière et ajuster vos déploiements en conséquence.

Correspondance de l'offre et de la demande

L'adéquation optimale de l'offre et de la demande fournit les coûts les plus bas pour un système, mais il doit aussi exister une offre supplémentaire suffisante pour permettre l'allocation de temps et les défaillances de ressources individuelles. La demande peut être fixe ou variable, et nécessiter des métriques et des automatisations afin de s'assurer que la gestion ne devient pas un coût important.

Dans AWS, vous pouvez allouer automatiquement les ressources pour répondre à la demande. Auto Scaling et les approches fondées sur la demande, la mémoire tampon ou le temps vous permettent d'ajouter ou de supprimer des ressources selon vos besoins. Si vous pouvez anticiper des modifications de la demande, vous pouvez économiser plus d'argent et garantir que vos ressources correspondent aux besoins de votre système.

Les questions suivantes portent essentiellement sur les correspondances de l'offre et de la demande à des fins d'optimisation des coûts :

COUT 4. Comment êtes-vous sûr que la capacité correspond à ce dont vous avez besoin, sans le dépasser de façon substantielle ?

Lors de la conception architecturale destinée à faire correspondre l'offre et la demande, vous voulez réfléchir activement aux modèles d'utilisation et au temps nécessaire pour approvisionner de nouvelles ressources.

Sensibilisation aux dépenses

La flexibilité et l'agilité accrues que permet le cloud favorise l'innovation, ainsi que le développement et le déploiement à un rythme soutenu. Le cloud élimine les processus manuels et le temps associé à l'allocation de l'infrastructure locale, y compris l'identification des spécifications matérielles, la négociation des devis, la gestion des bons de commande, la planification des livraisons et le déploiement des ressources. Cependant, cette facilité d'utilisation et cette capacité à la demande pratiquement illimitée peuvent nécessiter une nouvelle façon d'envisager les dépenses.

De nombreuses entreprises sont composées de plusieurs systèmes dirigés par diverses équipes. La capacité d'attribuer des coûts de ressource aux responsables individuels d'activité ou de produit oriente un comportement d'utilisation efficace et contribue à réduire les gaspillages. L'attribution précise des coûts vous permet aussi de comprendre quels produits sont réellement rentables et vous permet de prendre des décisions mieux fondées quant aux emplacements d'affectation du budget.

Les questions suivantes portent essentiellement sur la sensibilisation aux dépenses à des fins d'optimisation des coûts :

COUT 5. Avez-vous considéré les charges de transfert des données lors de la conception de votre architecture ?

COUT 6. Comment surveillez-vous l'utilisation et les dépenses ?

COUT 7. Mettez-vous hors service les ressources dont vous n'avez plus besoin ou arrêtez-vous celles qui ne sont pas nécessaires temporairement ?

COUT 8. Quels contrôles d'accès et procédures avez-vous en place pour régir l'utilisation d'AWS ?

Vous pouvez utiliser les balises de répartition des coûts pour classer vos coûts AWS par catégorie et en effectuer le suivi. Lorsque vous appliquez des balises à vos ressources AWS (telles que les instances Amazon EC2 ou les compartiments Amazon S3), AWS génère un rapport de répartition des coûts faisant apparaître votre consommation et les coûts regroupés par balise. Vous pouvez appliquer des balises qui représentent les catégories professionnelles (telles que les centres de coût, les noms de système ou les propriétaires) pour organiser vos coûts sur plusieurs services.

En associant les ressources balisées à la totalité du suivi du cycle de vie (employés, projets), il devient possible d'identifier les ressources orphelines ou les projets qui ne génèrent plus de valeur pour l'activité et qui doivent être mis hors service. Vous pouvez configurer des alertes de facturation pour être informé des dépassements de budget prévus, et le Calculateur de coûts mensuels AWS vous permet de calculer vos coûts de transfert des données.

Optimisation au fil du temps

Tandis qu'AWS propose de nouveaux services et de nouvelles fonctionnalités, une bonne pratique consiste à vérifier vos décisions architecturales existantes afin de garantir qu'elles continuent à être les plus économiques. Lorsque vos exigences évoluent, n'hésitez pas à désactiver des ressources et des services entiers, ou les systèmes qui ne vous sont plus nécessaires.

Comme les services gérés d'AWS peuvent souvent optimiser une solution de façon significative, il est judicieux d'être conscient des nouveaux services gérés tandis qu'ils deviennent disponibles. Par exemple, l'exécution d'une base de données Amazon RDS peut être moins onéreuse que celle de votre propre base de données sur Amazon EC2.

Les questions suivantes portent essentiellement sur les réévaluations de coût à des fins d'optimisation des coûts :

COUT 9. Comment gérez-vous et/ou envisagez-vous l'adoption de nouveaux services ?

En vérifiant régulièrement vos déploiements, il est souvent possible d'utiliser les nouveaux services AWS pour diminuer vos coûts. De même, évaluer en quoi les nouveaux services peuvent aider permet d'économiser de l'argent. Par exemple, AWS RDS for Aurora peut vous aider à réduire les coûts des bases de données relationnelles.

Services AWS clés

La fonction AWS clé qui prend en charge l'optimisation des coûts est celle des balises d'allocation, qui vous aident à maîtriser les coûts d'un système. Les services et fonctions suivants sont importants dans les quatre domaines d'optimisation des coûts :

Ressources rentables : vous pouvez utiliser les instances réservées et les capacités prépayées pour réduire votre coût. AWS Trusted Advisor permet d'inspecter votre environnement AWS et de rechercher des opportunités pour économiser de l'argent.

Correspondance de l'offre et de la demande : Auto Scaling vous permet d'ajouter ou de supprimer des ressources pour correspondre à la demande sans dépassement budgétaire

Sensibilisation aux dépenses : les alarmes Amazon CloudWatch et les notifications Amazon Simple Notification Service (SNS) vous préviennent si vous vous apprêtez à dépasser le montant budgété, ou en cas de pronostic d'un tel dépassement.

Optimisation au fil du temps : le blog AWS et la section *Nouveautés* du site web AWS constituent des ressources d'information relatives aux fonctionnalités et services récemment lancés. AWS Trusted Advisor inspecte votre environnement AWS et recherche des possibilités d'économies en éliminant les ressources inutilisées ou inactives, ou en choisissant la capacité des instances réservées.

Ressources

Consultez les ressources suivantes pour en savoir plus sur les bonnes pratiques AWS en matière d'optimisation des coûts.

Vidéo

- Optimisation des coûts sur AWS

Documentation

- Centre d'optimisation des coûts du cloud AWS

Outils

- Calculateur du coût total de possession (TCO) AWS
- Rapports de facturation détaillés AWS
- Calculateur de coûts mensuels AWS
- Explorateur de coûts AWS

Pilier « Excellence opérationnelle »

Le pilier **Excellence opérationnelle** inclut les pratiques et procédures opérationnelles utilisées pour gérer les charges de travail en production. Cela inclut la façon dont les modifications planifiées sont exécutées, ainsi que les réponses aux événements opérationnels inattendus. L'exécution des modifications et les réponses doivent être automatiques. Tous les processus et procédures de l'excellence opérationnelle doivent être documentés, testés et revus régulièrement.

Principes de conception

Dans le cloud, il existe un certain nombre de principes qui pilotent l'excellence opérationnelle :

- **Exécuter les opérations avec le code :** quand il existe des procédures ou processus répétitifs communs, utilisez l'automatisation. Par exemple, envisagez l'automatisation de la gestion de la configuration, des modifications et des réponses aux événements.
- **Aligner les processus opérationnels sur les objectifs métier :** collectez les métriques qui indiquent l'excellence opérationnelle dans la satisfaction des objectifs métier. Comme l'objectif doit être de réduire le ratio du signal sonore dans les métriques, la supervision opérationnelle et les réponses sont ciblées pour prendre en charge les besoins essentiels à l'activité. La collecte de métriques superflues empêche des réponses efficaces aux événements opérationnels inattendus en compliquant la supervision et les réponses.
- **Procéder à des modifications régulières, petites et incrémentielles :** les charges de travail doivent être conçues pour permettre que les composants soient mis à jour régulièrement. Les modifications doivent être faites par petits incréments, non par lots volumineux, et doivent pouvoir être annulées sans que les opérations n'en soient affectées. Mettez les procédures opérationnelles en place pour permettre l'implémentation de ces modifications sans interruption pour la maintenance ou le remplacement de composants de services dépendants.

- **Tester les réponses aux événements inattendus :** les charges de travail doivent être testées en cas de défaillances des composants ou autres événements opérationnels inattendus. Il est important de tester et de comprendre les procédures pour répondre aux événements opérationnels, de telle sorte qu'elles soient respectées lorsque des événements opérationnels se produisent. Créez des « game days » afin que vous puissiez tester les réponses à des événements opérationnels simulés et aux injections d'échecs.
- **Tirer les leçons des échecs et des événements opérationnels :** les processus doivent être en place de telle sorte que tous les types d'événements opérationnels et d'échecs soient capturés, vérifiés et utilisés en vue d'améliorations. Les révisions transversales régulières des opérations fonctionnelles doivent se traduire par des améliorations de processus qui pilotent l'excellence opérationnelle.
- **Maintenir à jour les procédures opérationnelles :** les guides de processus et de procédure doivent être adaptés au fur et à mesure de l'évolution des environnements et des opérations. Cela inclut la mise à jour des runbooks d'opérations régulières (procédures opérationnelles standard), ainsi que des playbooks (plans de réponse aux événements opérationnels inattendus ou aux échecs de production). Les directives et les apprentissages relatifs aux opérations doivent être partagés entre les équipes pour empêcher les fautes répétées. Pensez à utiliser un wiki ou une base de connaissances interne pour ces informations. Les informations qui doivent être évaluées inclut les métriques des opérations, les anomalies inattendues, les déploiements ayant échoué, les échecs système et les réponses inefficaces ou inappropriées aux défaillances. La documentation sur le système et sur l'architecture doit aussi être capturée et mise à jour à l'aide de l'automatisation au fur et à mesure que les environnements et les opérations évoluent.

Définition

Les trois zones de bonnes pratiques en matière d'excellence opérationnelle dans le cloud sont les suivantes :

1. Préparation
2. Transactions
3. Réponses

Pour piloter l'excellence opérationnelle, la préparation est essentielle. Nombre de problèmes opérationnels peuvent être évités en respectant les meilleures pratiques lors de la conception de la charge de travail, et les correctifs sont moins coûteux s'ils sont implémentés dans les phases de conception plutôt que dans les phases de production. Les procédures et processus opérationnels doivent être intégralement planifiés, testés et revus. Les charges de travail doivent évoluer et être modifiées de manière automatique et gérable. Les modifications doivent être réduites, fréquentes et incrémentielles, tout en n'ayant pas d'impact sur les opérations continues. Les équipes opérationnelles doivent être prêtes à répondre aux échecs et événements opérationnels, et disposer de processus en place pour en tirer des enseignements.

Bonnes pratiques

Préparation

Une préparation efficace est requise pour piloter l'excellence opérationnelle. Les listes de contrôle des opérations garantissent que les charges de travail sont prêtes pour l'opération de production et empêchent une promotion en production non intentionnelle sans préparation efficace. Les charges de travail doivent avoir des conseils opérationnels auxquels les équipes d'exploitation peuvent se reporter lors de l'exécution des tâches quotidiennes normales (runbooks), ainsi que des instructions pour répondre aux événements opérationnels inattendus (playbooks). Les playbooks doivent inclure les plans de réponse, ainsi que les chemins de réaffectation et les notifications aux parties prenantes. Il doit aussi être procédé aux révisions régulières des événements du cycle de vie professionnel qui peuvent piloter les modifications dans les opérations (événements marketing, ventes flash, etc.). Tous les runbooks et playbooks doivent être testés de telle sorte que les écarts ou les problèmes puissent être identifiés, et que les risques potentiels soient atténués. Les mécanismes de suivi des échecs et d'apprentissage doivent être en place. Les environnements, l'architecture et les paramètres de configuration des ressources en leur sein doivent être documentés d'une façon qui permet d'identifier aisément les composants à des fins de suivi et de dépannage. Les modifications apportées à la configuration doivent également être suivies et automatisées.

Dans AWS, il existe plusieurs méthodes, services et fonctions qui peuvent être utilisés pour prendre en charge la réactivité opérationnelle et la possibilité de se préparer aux opérations quotidiennes normales aussi bien qu'aux événements opérationnels inattendus. La réactivité opérationnelle peut continuer à inclure les révisions par des pairs ou transversales manuelles afin de garantir une vision. Les services AWS tels qu'AWS CloudFormation peuvent être utilisés pour s'assurer que les environnements contiennent toutes les ressources requises lorsqu'elles sont déployées en production, et que la configuration de l'environnement repose sur de bonnes pratiques testées, ce qui réduit le risque d'erreur humaine. L'implémentation d'Auto Scaling, ou autres mécanismes de dimensionnement automatique, permet aux charges de travail de répondre automatiquement lorsque des événements liés à l'activité affectent les besoins opérationnels. Des services comme AWS Config, avec les règles AWS Config associées, créent des mécanismes pour suivre automatiquement les modifications et y répondre dans vos environnements et charges de travail AWS. Il importe aussi d'utiliser des fonctionnalités comme le balisage pour s'assurer que toutes les ressources d'une charge de travail peuvent être facilement identifiées lorsque c'est nécessaire durant les opérations et les réponses.

Les questions suivantes portent essentiellement sur la préparation de l'excellence opérationnelle :

OPE 1. Quelles bonnes pratiques utilisez-vous pour les opérations du cloud ?

OPE 2. Comment assurez-vous la gestion de la configuration de votre charge de travail ?

Assurez-vous que la documentation ne devienne pas obsolète au fur et à mesure que les procédures évoluent. Vérifiez aussi qu'elle soit exhaustive. Sans les conceptions d'application, les configurations d'environnement, les configurations de ressource, les plans de réponse et les plans d'atténuation, la documentation n'est pas complète. Si la documentation n'est pas mise à jour et testée régulièrement, elle ne sera pas utile lorsque des événements opérationnels inattendus se produiront. Si les charges de travail ne sont pas revues avant la production, les opérations seront affectées lorsque des problèmes non détectés se produiront. Si les ressources ne sont pas documentées, lorsque des événements opérationnels se produisent, déterminer comment répondre sera plus difficile tandis que les ressources correctes sont identifiées.

Transactions

Les opérations doivent être normalisées et gérables sur une base régulière. L'accent doit être mis sur l'automatisation, les modifications réduites et fréquentes, les tests réguliers d'assurance qualité et les mécanismes définis, pour suivre, auditer, annuler ou vérifier les changements. Ceux-ci ne doivent être ni volumineux ni rares ; ils ne doivent pas nécessiter une interruption planifiée, et ne requièrent pas d'exécution manuelle. Un vaste ensemble de journaux et de métriques basées sur les indicateurs opérationnels clés d'une charge de travail doit être recueilli et contrôlé afin d'assurer des opérations continues.

Dans AWS, vous pouvez configurer un pipeline d'intégration continue / de déploiement continu (référentiel de code source, systèmes de génération, automatisation du déploiement et des tests). Les processus de gestion des versions, qu'ils soient manuels ou automatiques, doivent être testés et reposer sur des modifications incrémentielles réduites et des versions suivies. Vous devez pouvoir rétablir les modifications qui introduisent des problèmes opérationnels sans entraîner d'impact opérationnel. L'assurance qualité des modifications doit inclure les stratégies d'atténuation des risques, telles que Blue/Green, Canary et tests A/B. Les listes de contrôle des opérations doivent être utilisées pour évaluer la réactivité d'une charge de travail en vue de la production. Regroupez les journaux pour une supervision et des alertes centralisées. Assurez-vous que les alertes déclenchent des réponses automatiques, notifications et réaffectations incluses. Concevez aussi des surveillances pour les anomalies, pas seulement pour les échecs.

Les questions suivantes portent essentiellement sur le fonctionnement de la charge de travail en vue de l'excellence opérationnelle :

OPE 3. Comment faites-vous évoluer votre charge de travail tout en réduisant l'impact des modifications ?

OPS 4. Comment surveillez-vous votre charge de travail pour vous assurer qu'elle se comporte comme prévu ?

Les opérations régulières, ainsi que les réponses aux événements non planifiés, doivent être automatisées. Evitez les processus manuels pour les déploiements, la gestion des versions, les modifications et les annulations. Les versions ne doivent pas être des lots volumineux exécutés de façon irrégulière. Les restaurations sont plus difficiles en cas de modifications volumineuses, et l'impossibilité d'avoir un plan de restauration ou d'atténuer les impacts des échecs, empêche la continuité des opérations. Alignez la supervision sur les besoins métier de telle sorte que les réponses gèrent la continuité de l'activité de manière efficace. Une supervision ponctuelle et non centralisée, avec des réponses manuelles, aura un plus grand impact sur les opérations en cas d'événements inattendus.

Réponses

Les réponses aux événements opérationnels non prévus doivent être automatisées. Cela ne concerne pas seulement les alertes, mais aussi l'atténuation, la correction, la restauration et la récupération. Les alertes doivent être opportunes et appeler des réaffectations lorsque les réponses ne sont pas adaptées à la réduction de l'impact des événements opérationnels. Les mécanismes d'assurance qualité doivent être en place pour annuler automatiquement les déploiements ayant échoué. Les réponses doivent suivre un playbook prédéfini qui inclut les parties prenantes, le processus de réaffectation et les procédures. Les chemins de réaffectation doivent être définis et inclure les capacités de réaffectation hiérarchique et fonctionnelle. La réaffectation hiérarchique doit être automatique et la priorité réaffectée doit se traduire dans les notifications des parties prenantes.

Dans AWS, il existe plusieurs mécanismes pour garantir une gestion des alertes et des notifications appropriée en réponse aux événements opérationnels non prévus, ainsi que des réponses automatiques. Des outils doivent aussi être en place pour superviser de façon centralisée les charges de travail, et créer des alertes et des notifications efficaces basées sur tous les journaux et métriques disponibles qui se rapportent aux indicateurs opérationnels clés. Cela inclut les alertes et les notifications en cas d'anomalies hors limites, et pas seulement en cas de défaillance d'un composant ou d'un service. Les réponses doivent être activées pour les services et environnements AWS dépendants, ainsi que pour l'intégrité applicative de la charge de travail. L'analyse des causes racine doit être exécutée après les événements opérationnels et utilisée pour améliorer à la fois l'architecture et les plans de réponse.

Les questions suivantes portent essentiellement sur la réponse aux événements dans le cadre de l'excellence opérationnelle :

OPE 5. Comment répondez-vous aux événements opérationnels non prévus ?

OPE 6. Comment est gérée la réaffectation lorsque vous répondez à des événements opérationnels non prévus ?

Si les plans de réponse sont faits de manière ponctuelle et non définie, les résultats sont imprévisibles et exacerbent souvent l'impact d'un événement. Si les réponses sont basées sur des playbooks obsolètes, ou si un playbook n'est pas accessible en cas de problèmes, les résultats sont aussi imprévisibles. Si un processus n'est pas en place pour vérifier les événements, les futurs événements opérationnels seront plus durs à prévenir et auront aussi le même impact.

Services AWS clés

Il existe deux services principaux qui peuvent être utilisés pour piloter l'excellence opérationnelle. AWS CloudFormation permet de créer des modèles basés sur les bonnes pratiques et d'allouer les ressources d'une manière ordonnée et prévisible. Amazon CloudWatch permet de surveiller les métriques, de recueillir les journaux, de générer les alertes et de déclencher les réponses. Les autres services et fonctions qui prennent en charge les trois zones de l'excellence opérationnelle sont les suivants :

Préparation : AWS Config fournit un inventaire détaillé de votre configuration et de vos ressources AWS, et enregistre continuellement les changements de configuration. AWS Service Catalog permet de créer un ensemble normalisé d'offres de services alignées sur les bonnes pratiques. La conception de charges de travail qui utilisent l'automatisation avec des services comme Auto Scaling et Amazon SQS constitue une excellente méthode pour garantir les opérations continues en cas d'événements opérationnels inattendus.

Opérations : AWS CodeCommit, AWS CodeDeploy et AWS CodePipeline permettent de gérer et d'automatiser les modifications du code sur les charges de travail AWS. Utilisez les kits SDK AWS ou les bibliothèques tierces pour automatiser les modifications opérationnelles. Utilisez AWS CloudTrail pour auditer et suivre les modifications apportées aux environnements AWS.

Réponses : tirez parti de toutes les fonctionnalités du service Amazon CloudWatch afin d'obtenir des réponses efficaces et automatiques. Les alarmes Amazon CloudWatch permettent de définir des seuils pour les alertes et les notifications, et les événements Amazon CloudWatch peuvent déclencher les notifications et les réponses automatiques.

Ressources

Reportez-vous aux ressources suivantes pour en savoir plus sur nos bonnes pratiques en matière d'excellence opérationnelle.

Vidéos

- AWS re :Invent 2015 - Opérations de développement chez Amazon
- AWS re :Invent 2015 - Dimensionnement des opérations d'infrastructure
- AWS Summit 2016 - Opérations de développement, intégration continue et déploiement sur AWS

Documentation et blogs

- Opérations de développement et AWS
- Présentation de l'intégration continue
- Présentation de la livraison continue
- Blog Opérations de développement AWS

Livres blancs

- AWS Cloud Adoption Framework - Perspective des opérations
- Présentation de DevOps sur AWS
- Liste de contrôle opérationnelle AWS

AWS Support

- Contrôle des opérations du cloud AWS
- AWS Premium Support
- AWS Trusted Advisor

Conclusion

L'infrastructure AWS correctement architecturée fournit les bonnes pratiques à travers cinq piliers permettant de concevoir et de gérer dans le cloud des systèmes fiables, sécurisés et économiques. L'infrastructure AWS documente un ensemble de questions qui vous permettent de contrôler une architecture existante ou suggérée, et de définir aussi un ensemble de bonnes pratiques AWS pour chaque pilier. L'utilisation de l'infrastructure dans votre architecture vous aidera à produire des systèmes stables et efficaces, qui vous permettent de vous concentrer sur vos exigences fonctionnelles.

Collaborateurs

Les personnes et organisations suivantes ont participé à l'élaboration de ce document :

- Philip Fitzsimons, responsable architecture des solutions, Amazon Web Services
- Erin Rifkin, responsable produit senior, Amazon Web Services
- Max Ramsay, architecte principal de solutions de sécurité, Amazon Web Services
- Scott Paddock, architecte de solutions de sécurité, Amazon Web Services
- Jon Steele, responsable technique senior, Amazon Web Services
- Callum Hughes, architecte de solutions, Amazon Web Services

Historique du document

20 novembre 2015. Mise à jour de l'annexe avec les informations des journaux Amazon CloudWatch.

20 novembre 2016. Mise à jour de l'infrastructure pour inclure le pilier Excellence opérationnelle, et réviser et mettre à jour les autres piliers de façon à réduire la duplication et à intégrer les leçons tirées de l'exécution de révisions auprès de milliers de clients.

Annexe : Questions, réponses et bonnes pratiques relatives à l'infrastructure correctement architecturée

Cette annexe contient la liste complète des questions et réponses, bonnes pratiques incluses, relatives à l'infrastructure correctement architecturée, organisées par pilier :

Pilier « Sécurité »

IAM (Identity and Access Management)

SEC 1. comment protégez-vous l'accès aux informations d'identification du compte racine (root) AWS et leur utilisation ?

Les informations d'identification du compte root (racine) AWS sont similaires à celles de l'administrateur local ou racine des autres systèmes d'exploitation et doivent être utilisées avec parcimonie. La bonne pratique actuelle consiste à créer les utilisateurs AWS Identity and Access Management (IAM), à les associer à un groupe administrateur et à utiliser le compte IAM pour gérer le compte. Le compte racine AWS ne doit pas avoir de clés d'API, doit avoir un mot de passe fort et doit être associé à un périphérique Multi-Factor Authentication (MFA) matériel. Il s'ensuit que l'utilisation de l'identité racine n'est possible qu'au travers d'AWS Management Console et que le compte racine ne peut pas être utilisé pour les appels d'API. Notez que certains revendeurs ou certaines régions ne distribuent pas ou ne prennent pas en charge les informations d'identification des comptes racine AWS.

Bonnes pratiques :

- **MFA et utilisation minimale de la racine** Les informations d'identification du compte racine AWS ne sont utilisées que pour les activités minimales requises.
- **Non utilisation de la racine**

SEC 2. Comment définissez-vous les rôles et les responsabilités des utilisateurs système pour contrôler l'accès humain à AWS Management Console et aux API ?

La bonne pratique actuelle consiste pour les clients à séparer les rôles et les responsabilités définis des utilisateurs système en créant des groupes d'utilisateurs. Les groupes d'utilisateurs peuvent être définis à l'aide de plusieurs technologies : les groupes IAM (Identity and Access Management), les rôles IAM pour l'accès entre comptes, les identités web, via l'intégration SAML (Security Assertion Markup Language) (par exemple, définition des rôles dans Active Directory) ou à l'aide d'une solution tierce (par exemple, Okta, Ping Identity ou autre technique personnalisée) qui s'intègre généralement via SAML ou AWS Security Token Service (STS). L'utilisation d'un compte partagé est fortement déconseillée.

Bonnes pratiques :

- **Cycle de vie géré des employés** Les stratégies de cycle de vie des employés sont définies et appliquées.
- **Privilège minimum** Les utilisateurs, les groupes et les rôles sont clairement définis, et ne reçoivent que les privilèges minimaux nécessaires à l'exécution des exigences métier.

SEC 3. Comment limitez-vous l'accès automatique aux ressources AWS ? (par exemple, applications, scripts et/ou outils ou services tiers)

L'accès systématique doit être défini de manière similaire, car les groupes d'utilisateurs sont créés pour les personnes. Pour les instances Amazon EC2, ces groupes sont appelés rôles IAM pour EC2. La bonne pratique actuelle consiste à utiliser les rôles IAM pour EC2 et un SDK ou une interface de ligne de commande AWS, qui possède une prise en charge intégrée pour extraire les rôles IAM pour les informations d'identification EC2. Généralement, les informations d'identification utilisateur sont injectées dans les instances EC2, mais le codage en dur des informations d'identification dans les scripts et le code source est fortement déconseillé.

Bonnes pratiques :

- **Informations d'identification statiques utilisées pour l'accès automatique** Stockez ces informations en toute sécurité.
- **Authentification dynamique pour l'accès automatique** Gérez l'authentification à l'aide des profils d'instance ou d'Amazon STS.

Contrôles de détection

SEC 4. Comment capturez-vous et analysez-vous les journaux ?

Les journaux de capture sont essentiels pour tout examiner, des performances aux incidents de sécurité. La bonne pratique actuelle consiste à ce que les journaux soient régulièrement déplacés de la source directement dans un système de traitement des journaux (par exemple, CloudWatch Logs, Splunk, Papertrail, etc.) ou stockés dans un compartiment Amazon S3 en vue d'un traitement ultérieur basé sur les besoins professionnels. Les sources communes des journaux sont les API AWS et les journaux liés aux utilisateurs AWS (par exemple, AWS CloudTrail), les journaux spécifiques aux services AWS (par exemple, Amazon S3, Amazon CloudFront, etc.), les journaux générés par les systèmes d'exploitation et les journaux propres aux applications tiers. Vous pouvez utiliser les journaux Amazon CloudWatch pour surveiller, stocker et atteindre vos fichiers journaux à partir des instances Amazon EC2, d'AWS CloudTrail ou d'autres sources.

Bonnes pratiques :

- **Surveillance d'activité appropriée : journaux Amazon CloudWatch, événements, journaux de flux VPC, journaux ELB, journaux de compartiment S3, etc.**
- **AWS Cloud Trail activé**
- **Journaux du système d'exploitation ou des applications surveillés.**

Protection de l'infrastructure

SEC 5. Comment appliquez-vous la protection des limites aux niveaux réseau et hôte ?

Dans les centres de données locaux, une approche DMZ sépare les systèmes en zones fiables et non fiables à l'aide de pare-feux. Sur AWS, les pare-feux avec état et sans état sont utilisés. Les pare-feux avec état sont appelés groupes de sécurité, et les pare-feux sans état sont appelés listes de contrôle d'accès (ACL) qui protègent les sous-réseaux dans un Amazon Virtual Private Cloud (VPC). La bonne pratique actuelle consiste à exécuter un système dans un VPC, et à définir la sécurité basée sur les rôles dans des groupes de sécurité (par exemple, couche web, couche applications, etc.) et la sécurité basée sur les emplacements dans des listes de contrôle d'accès (ACL) réseau (par exemple, couche Elastic Load Balancing dans un sous-réseau par zone de disponibilité, couche web dans un autre sous-réseau par zone de disponibilité, etc.).

Bonnes pratiques :

- **Trafic réseau contrôlé dans un VPC** Par exemple, utilisez les pare-feux, groupes de sécurité, les ACL réseau, un hôte bastion, etc.
- **Trafic réseau contrôlé à la limite** Par exemple, utilisez AWS WAF, les pare-feux basés sur les hôtes, les groupes de sécurité, les ACL réseau, etc.

SEC 6. Comment mettez-vous à profit les fonctions de sécurité au niveau des services AWS ?

Les services AWS offrent des fonctionnalités de sécurité supplémentaires (par exemple, stratégies de compartiment Amazon S3, Amazon SQS, Amazon DynamoDB, stratégies clés KMS, etc.).

Bonnes pratiques :

- **Utilisation de fonctionnalités supplémentaires si nécessaire**

SEC 7. Comment protégez-vous l'intégrité du système d'exploitation sur vos instances Amazon EC2 ?

Un autre contrôle traditionnel consiste à protéger l'intégrité du système d'exploitation. Cela s'effectue facilement dans Amazon EC2 à l'aide des techniques traditionnelles basées sur les hôtes (par exemple, OSSEC, Tripwire, Trend Micro Deep Security, etc.).

Bonnes pratiques :

- **Intégrité des fichiers** Les contrôles d'intégrité des fichiers sont utilisés pour les instances EC2.
- **Détection d'intrusion sur EC2** Les contrôles de détection d'intrusion basés sur les hôtes sont utilisés pour les instances EC2.
- **AWS Marketplace ou solution partenaire** Solution d'AWS Marketplace ou d'un partenaire APN.
- **Outil de gestion de configuration** Utilisation d'un AMI personnalisé ou d'outils de configuration de gestion (c'est-à-dire, Puppet ou Chef), sécurisés par défaut.

Protection des données

SEC8. Comment classez-vous les données ?

La classification des données offre un moyen de classer par catégorie les données organisationnelles en fonction des niveaux de sensibilité. Cela inclut les types de données disponibles, l'emplacement des données, les niveaux d'accès et la protection des données (via le chiffrement ou le contrôle d'accès, par exemple).

Bonnes pratiques :

- **Utilisation du schéma de classification des données**
- **Toutes les données sont traitées comme sensibles**

SEC 9. Comment chiffrez-vous et protégez-vous vos données au repos ?

Un contrôle traditionnel de sécurité consiste à chiffrer les données au repos. AWS prend en charge cette fonction côté client (par exemple, support SDK, support système d'exploitation, Windows Bitlocker, dm-crypt, Trend Micro SafeNet, etc.) et côté serveur (par exemple, Amazon S3). Vous pouvez aussi utiliser le chiffrement côté serveur et les volumes chiffrés Amazon Elastic Block Store.

Bonnes pratiques :

- **Non obligatoire** Le chiffrement des données au repos n'est pas obligatoire
- **Chiffrement au repos**

SEC 10. Comment gérez-vous les clés ?

Les clés sont des secrets qui doivent être protégés, tandis qu'une stratégie appropriée de rotation doit être définie et utilisée. La bonne pratique consiste à ne pas coder en dur ces secrets dans des scripts et des applications, mais cela se produit souvent.

Bonnes pratiques :

- **AWS CloudHSM** Utilisez AWS CloudHSM.
- **Utilisation des contrôles des services AWS** Les données au repos sont chiffrées à l'aide de contrôles spécifiques aux services AWS (par exemple, chiffrement côté serveur Amazon S3, volumes chiffrés Amazon EBS, Amazon Relational Database Service (RDS), Transparent Data Encryption (TDE), etc.).
- **Utilisation côté client** Les données au repos sont chiffrées à l'aide de techniques côté client.
- **AWS Marketplace ou solution partenaire** Solution d'AWS Marketplace ou d'un partenaire APN. (Par exemple, SafeNet, TrendMicro, etc.).

SEC 11. Comment chiffrez-vous et protégez-vous vos données en transit ?

Une bonne pratique consiste à protéger les données en transit à l'aide du chiffrement. AWS prend en charge l'utilisation des points de terminaison chiffrés pour les API de service. En outre, les clients peuvent utiliser différentes techniques au sein de leurs instances Amazon EC2.

Bonnes pratiques :

- **Non obligatoire** Chiffrement non obligatoire sur les données en transit.
- **Communications chiffrées**TLS ou équivalent est utilisé pour les communications si nécessaire.

Réponse aux incidents

SEC 12. Comment vous assurez-vous que vous disposez de la réponse aux incidents appropriée ?

La mise en place des outils et des accès préalablement à tout incident de sécurité, puis la pratique régulière de réponse aux incidents garantira que l'architecture est mise à jour pour accueillir un examen et une récupération opportuns.

Bonnes pratiques :

- **Accès pré-provisionné** Infosec possède l'accès correct ou le moyen d'obtenir rapidement un accès. Ceci doit être préalablement alloué de telle sorte qu'une réponse appropriée puisse être apportée à un incident.
- **Outils prédéployés** Infosec a les outils appropriés prédéployés dans AWS de telle sorte qu'une réponse appropriée puisse être apportée à un incident
- **Game Days dans un environnement autre que de production** Les simulations de réponse aux incidents sont conduites régulièrement dans l'environnement autre que de production et les leçons tirées sont intégrées à l'architecture et aux opérations.
- **Game Days dans un environnement de production** Les simulations de réponse aux incidents sont conduites régulièrement dans l'environnement de production et les leçons tirées sont intégrées à l'architecture et aux opérations.

Pilier « Fiabilité »

Fondations

FIA 1. Comment gérez-vous les Service Limits AWS pour vos comptes ?

Les comptes AWS sont attribués avec les limites de service par défaut pour empêcher de nouveaux utilisateurs d'allouer involontairement plus de ressources que nécessaire. Les clients AWS doivent évaluer leurs besoins en services AWS et demander les modifications appropriées de leurs limites pour chaque région utilisée.

Bonnes pratiques :

- **Surveiller et gérer les limites :** évaluez votre utilisation potentielle sur AWS, augmentez vos limites régionales en conséquence et autorisez la croissance planifiée de l'utilisation.
- **Configurer la supervision automatique :** implémentez les outils, tels que les SDK, pour être informé lorsque les seuils sont atteints.
- **Etre conscient des Service Limits fixes :** connaissez les Service Limits non modifiables et de leur architecture.
- **Assurez-vous qu'il existe un écart suffisant entre vos Service Limits et votre utilisation maximale pour pouvoir gérer le basculement**
- **Les Service Limits sont prises en compte à travers tous les comptes et toutes les régions appropriés**

FIA 2. Comment planifiez-vous la topologie de votre réseau sur AWS ?

Les applications peuvent exister dans un ou plusieurs environnements : EC2 Classic, VPC ou VPC par défaut. Les considérations réseau telles que la connexion système, la gestion des adresses IP élastiques/publiques, la gestion du VPC/des adresses privées, et la résolution de noms sont essentielles pour exploiter les ressources du cloud. Les déploiements correctement planifiés et documentés sont essentiels pour réduire le risque de chevauchement et de conflit.

Bonnes pratiques :

- **Connexion retour au centre de données non nécessaire**
- **Connexion hautement disponible entre AWS et l'environnement local (si applicable) :** plusieurs circuits DX, plusieurs tunnels VPN, appliances AWS Marketplace si applicable.
- **Connexion réseau hautement disponible pour les utilisateurs de la charge de travail :** répartition de charge et/ou proxy hautement disponible, solution DNS, appliances AWS Marketplace, etc.
- **Plages d'adresses IP sans chevauchement :** l'utilisation de vos plages d'adresses IP et sous-réseaux de votre cloud privé virtuel ne doit pas se chevaucher ni chevaucher d'autres environnements cloud ou vos environnements locaux.
- **Allocation de sous-réseau IP :** les plages d'adresses IP Amazon VPC doivent être assez grandes pour répondre aux exigences d'une application, y compris la prise en compte d'une future extension ou allocation d'adresses IP aux sous-réseaux via les zones de disponibilité.

Gestion des modifications

FIA 4. Comment votre système s'adapte-t-il aux modifications à la demande ?

Un système évolutif peut offrir une élasticité pour ajouter ou supprimer des ressources automatiquement de telle sorte qu'elles correspondent étroitement à la demande en cours à un instant donné, quel qu'il soit.

Bonnes pratiques :

- **Dimensionnement automatique :** utilisez les services pouvant être dimensionnés automatiquement, tels qu'Amazon S3, Amazon CloudFront, Auto Scaling, Amazon DynamoDB, AWS Elastic Beanstalk, etc.
- **Test de charge :** adoptez une méthodologie de test de charge pour déterminer si l'activité de dimensionnement satisfait aux exigences de l'application.

FIA 3. Comment surveillez-vous les ressources AWS ?

Les journaux et les métriques constituent un outil puissant pour obtenir un aperçu de l'état de vos applications. Vous pouvez configurer votre système pour surveiller les journaux et les métriques, et envoyer des notifications lorsque les seuils sont franchis ou que des événements significatifs se produisent.

Idéalement, quand les seuils de performance basse sont franchis ou que des défaillances se produisent, le système doit avoir été architecturé pour se réparer automatiquement ou se dimensionner en conséquence.

Bonnes pratiques :

- **Supervision :** surveillez vos applications avec Amazon CloudWatch ou des outils tiers.
- **Notification :** prévoyez de recevoir des notifications lorsque des événements significatifs se produisent.
- **Réponse automatique :** utilisez l'automatisation pour prendre des mesures en cas de détection d'une défaillance, par exemple le remplacement de composants défectueux.

FIA 5. Comment exécutez-vous les modifications ?

Les modifications non contrôlées de votre environnement rendent difficile la prévision de l'effet d'un changement. La modification contrôlée des applications et des ressources AWS allouées est nécessaire pour garantir que les applications et l'environnement d'exploitation exécutent des logiciels connus, qui peuvent être corrigés ou remplacés de manière prévisible.

Bonnes pratiques :

- **Automatisation** : automatisez les déploiements et les correctifs.

Gestion des défaillances

FIA 6. Comment sauvegardez-vous les données ?

Sauvegardez les données, les applications et les environnements d'exploitation (définis comme systèmes d'exploitation configurés avec les applications) pour satisfaire aux exigences du délai moyen de récupération (MTTR, Mean Time To Recovery) et des objectifs du point de récupération (RPO, Recovery Point Objectives).

Bonnes pratiques :

- **Sauvegardes automatiques :** utilisez les fonctions AWS, les solutions AWS Marketplace ou les logiciels tiers pour automatiser les sauvegardes.
- **Test régulier de la récupération** : validez la satisfaction de l'implémentation du processus de sauvegarde à l'objectif de délai de récupération et aux objectifs du point de récupération via un test de récupération.

FIA 7. Comment votre système supporte-t-il les défaillances de composants ?

Vos applications obéissent-elles à une exigence, implicite ou explicite, de haute disponibilité et de délai moyen de récupération bas ? Si tel est le cas, concevez l'architecture de vos applications par rapport à la résilience et répartissez-les de façon à supporter les pannes. Pour atteindre de plus hauts niveaux de disponibilité, cette distribution doit être répartie sur plusieurs emplacements physiques. Concevez l'architecture des couches individuelles (par exemple, serveur web, base de données) à des fins de résilience, ce qui inclut la supervision, la réparation automatique et la notification des défaillances ou perturbations significatives.

Bonnes pratiques :

- **Région Multi-AZ** : distribuez la charge des applications entre plusieurs zones de disponibilité /régions (DNS, ELB, Application Load Balancer, API Gateway, par exemple)

- **Dépendances à couplage lâche** : par exemple, utilisez les systèmes à file d'attente, les systèmes de streaming, les flux de travail, les répartiteurs de charge, etc.
- **Dégradation appropriée** : lorsque les dépendances d'un composant sont non saines, le composant lui-même n'est pas rapporté comme non sain. Il peut continuer à traiter les demandes d'une manière dégradée.
- **Réparation automatique** : utilisez les fonctions automatiques pour détecter les défaillances et exécuter une action de correction. Surveillez en permanence l'état de votre système et planifiez de recevoir les notifications de tout événement important.

FIA 8. Comment testez-vous votre résilience ?

Lors du test de résilience, il se peut que vous trouviez des bogues latents qui ne peuvent apparaître qu'en production. La pratique régulière de vos procédures au travers de « game days » aidera votre organisation à exécuter vos procédures de façon homogène.

Bonnes pratiques :

- **Playbook** : ayez un playbook pour les scénarios d'échec.
- **Injection d'échec** : testez régulièrement les échecs (avec Chaos Monkey, par exemple), en assurant la couverture des chemins d'échec.
- **Planifier les game days**
- **Analyse des causes premières** : exécutez les vérifications des échecs système en fonction des événements significatifs pour évaluer l'architecture.

FIA 9. Comment planifiez-vous la reprise après sinistre ?

La récupération des données étant essentielle, la restauration doit être obligatoire à partir de méthodes de sauvegarde. Votre définition et exécution des objectifs, ressources, emplacements et fonctions de ces données doivent être conformes aux objectifs RTO et RPO.

Bonnes pratiques :

- **Objectifs définis :** définissez les objectifs RTO et RPO.
- **Reprise après sinistre :** établissez une stratégie de reprise après sinistre.
- **Dérive de configuration :** assurez-vous que les Amazon Machine Images (AMI) et l'état de configuration du système sont à jour sur le site/la région de reprise après sinistre.
- **Reprise après sinistre testée et validée :** testez régulièrement le basculement vers la reprise après sinistre pour vous assurer que les objectifs RTO et RPO sont satisfaits.
- **Implémentation de la récupération automatique :** utilisez AWS et/ou des outils tiers pour automatiser la récupération système.

Pilier « Performances »

Sélection

PERF 1. Comment sélectionner l'architecture aux meilleures performances ?

La solution optimale d'un système particulier varie en fonction du type de charge de travail, souvent avec plusieurs approches combinées. Les systèmes correctement architecturés utilisent plusieurs solutions et autorisent différentes fonctions pour améliorer les performances.

Bonnes pratiques :

- **Comparaison** : testez une charge de travail connue sur AWS et utilisez-la pour estimer la meilleure sélection.
- **Test de charge** : déployez la version la plus récente de votre système sur AWS à l'aide de différents types et tailles de ressource, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 2. Comment avez-vous sélectionné la solution de calcul ?

La solution de calcul optimale d'un système particulier peut varier selon la conception de l'application, les modèles d'utilisation et les paramètres de configuration. Les architectures peuvent utiliser différentes solutions de calcul pour divers composants et activer différentes fonctions pour améliorer les performances. La sélection d'une solution de calcul incorrecte pour une architecture peut conduire à une efficacité moindre des performances.

Bonnes pratiques :

- **Prendre en compte les options** : envisagez les différentes options d'utilisation des instances, conteneurs et fonctions pour obtenir les meilleures performances.
- **Options de configuration d'instance** : si vous utilisez les instances, prenez en compte les options de configuration, telles que la famille, les tailles d'instance et les fonctionnalités (GPU, E/S, extension de capacité).
- **Options de configuration de conteneur** : si vous utilisez les conteneurs, prenez en compte les options de configuration, telles que la mémoire, l'UC et la configuration de location du conteneur.
- **Options de configuration de fonction** : si vous utilisez les fonctions, prenez en compte les options de configuration telles que la mémoire, le runtime et l'état.
- **Elasticité** : utilisez l'élasticité (Auto Scaling, Amazon EC2 Container Service (ECS), AWS Lambda, par exemple) pour satisfaire les modifications à la demande.

PERF 3. Comment sélectionnez-vous la solution de stockage ?

La solution de stockage optimale pour un système particulier varie en fonction du type de méthode d'accès (bloc, fichier ou objet), des modèles d'accès (aléatoire ou séquentiel), du débit requis, de la fréquence des accès (en ligne, hors connexion, archivage), de la fréquence des mises à jour (WORM, dynamiques) et des contraintes de disponibilité et de durabilité. Les systèmes correctement architecturés utilisent plusieurs solutions de stockage et autorisent différentes fonctions pour améliorer les performances.

Bonnes pratiques :

- **Prendre en compte les caractéristiques** : prenez en compte les différentes caractéristiques (partage, taille de fichier, taille de cache, modèles d'accès, latence, débit, persistance des données, par exemple) obligatoires pour sélectionner les services que vous devez utiliser (Amazon S3, Amazon EBS, Amazon Elastic File System (EFS), stockage d'instance EC2)
- **Prendre en compte les options de configuration** : envisagez les options de configuration telles que PIOPS, disques SSD, disques magnétiques et Amazon S3 Transfer Acceleration.
- **Prendre en compte les modèles d'accès** : optimisez votre utilisation des systèmes de stockage en fonction des modèles d'accès (agrégation, distribution des clés, partitionnement, par exemple).

PERF 4. Comment sélectionnez-vous votre solution de base de données ?

La solution de base de données optimale pour un système particulier peut varier en fonction des exigences de disponibilité, de cohérence, de tolérance des partitions, de latence, de durabilité, d'évolutivité et de capacité des requêtes. De nombreux systèmes utilisent différentes solutions de base de données pour divers sous-systèmes et activent différentes fonctions pour améliorer les performances. La sélection d'une solution de base de données et de fonctionnalités incorrectes pour un système peut conduire à une efficacité moindre des performances.

Bonnes pratiques :

- **Prendre en compte les caractéristiques** : prenez en compte les différentes caractéristiques (disponibilité, cohérence, tolérance des partitions, latence, durabilité, évolutivité, capacité des requêtes, par exemple) de telle sorte que vous puissiez sélectionner l'approche de base de données la plus performante à utiliser (relationnelle, No-SQL, entrepôt, en mémoire).
- **Prendre en compte les options de configuration** : envisagez les options de configuration telles que l'optimisation du stockage, les paramètres du niveau base de données, la mémoire et le cache.
- **Prendre en compte les modèles d'accès** : optimisez votre utilisation des systèmes de base de données en fonction de votre modèle d'accès (index, distribution des clés, partition, dimensionnement horizontal).
- **Prendre en compte d'autres approches** : envisagez d'autres approches pour fournir des données interrogeables telles que les index de recherche, les entrepôts de données et le big data.

PERF 5. Comment configurez-vous votre solution réseau ?

La solution de réseau optimale pour un système particulier varie en fonction des exigences de latence et de débit, entre autres. Les contraintes physiques, telles que les ressources utilisateur ou les ressources locales, pilotent les options d'emplacement, qui peuvent être décalées à l'aide de techniques périphériques ou d'emplacement des ressources.

Bonnes pratiques :

- **Prendre en compte l'emplacement** : envisagez vos options d'emplacement (région, zone de disponibilité, groupes de placement, périphérie) pour réduire la latence réseau.
- **Prendre en compte les fonctionnalités produit** : prenez en compte les fonctionnalités produit (capacité réseau des instances EC2, types d'instance réseau très élevés, instances optimisées Amazon EBS, Amazon S3 Transfer Acceleration, Dynamic Amazon CloudFront) pour optimiser le trafic réseau.

- **Prendre en compte les fonctionnalités réseau** : envisagez les fonctionnalités réseau (routage de latence Amazon Route 53, points de terminaison Amazon VPC et AWS Direct Connect, par exemple) afin de réduire la distance ou l'instabilité réseau.
- **ACL réseau appropriées** : utilisez l'ensemble minimal d'ACL réseau pour gérer le débit réseau.
- **Prendre en compte le déchargement du chiffrement** : envisagez l'utilisation de l'équilibrage de charge pour décharger la terminaison du chiffrement (TLS).
- **Prendre en compte les protocoles** : envisagez les protocoles dont vous avez besoin pour optimiser les performances réseau.

Révision

PERF 6. Comment savez-vous que vous continuez à avoir le type de ressource le plus approprié tandis que de nouveaux types de ressource et de nouvelles fonctionnalités sont introduits ?

Lors de la conception architecturale des solutions, vous avez la possibilité de choisir au sein d'un ensemble fini d'options. Cependant, au fil du temps de nouvelles technologies et approches deviennent disponibles qui peuvent améliorer les performances de votre architecture.

Bonnes pratiques :

- **Vérifier** : ayez un processus pour vérifier les nouveaux types de ressource et les tailles. Exécutez à nouveau les tests de performance pour évaluer les améliorations de l'efficacité des performances.

Supervision

PERF 7. Comment surveillez-vous les ressources après leur lancement pour vous assurer qu'elles se comportent comme prévu ?

Les performances du système peuvent se dégrader au fil du temps en raison de facteurs internes et/ou externes. La supervision des performances des systèmes vous permet d'identifier cette dégradation et de corriger les facteurs internes ou externes (tels que le système d'exploitation ou la charge de l'application).

Bonnes pratiques :

- **Supervision** : utilisez Amazon CloudWatch, des outils tiers ou des outils de supervision personnalisés pour surveiller les performances.
- **Notifications basées sur les alarmes** : recevez une alerte automatique de votre système de supervision si les métriques excèdent les limites sécurisées.
- **Actions basées sur les déclencheurs** : définissez les alarmes qui déclenchent des actions automatiques pour corriger un problème ou le faire remonter.

Compromis

PERF 8. Comment puis-je utiliser les compromis pour améliorer les performances ?

Lorsque vous concevez l'architecture de vos solutions, pensez activement à procéder à des compromis afin de sélectionner une approche optimale. Souvent, vous pouvez négocier entre la cohérence, la durabilité et l'espace d'un côté, et le temps ou la latence de l'autre, pour offrir des performances plus élevées.

Bonnes pratiques :

- **Prendre en compte les services** : utilisez les services qui améliorent les performances, comme Amazon ElastiCache, Amazon CloudFront et AWS Snowball.
- **Prendre en compte les modèles** : utilisez des modèles pour améliorer les performances, comme la mise en cache, les réplicas en lecture, les partitions, la compression et la mémoire tampon.

Pilier « Optimisation des coûts »

Ressources économiques

COUT 1. Lorsque vous sélectionnez les services AWS pour votre solution, prenez-vous le coût en compte ?

Amazon EC2, Amazon EBS, Amazon S3, etc. sont tous trois des services AWS de « blocs de construction ». Les services gérés tels qu'Amazon RDS, Amazon DynamoDB, etc. sont des services AWS « de niveau supérieur ». En sélectionnant les services gérés et blocs de construction appropriés, vous pouvez optimiser votre architecture en termes de coût. Par exemple, en utilisant les services gérés, vous pouvez réduire ou supprimer une grande partie de votre traitement administratif et opérationnel, et vous dégager ainsi du temps pour travailler les applications et les activités liées à l'entreprise.

Bonnes pratiques :

- **Sélectionner les services pour la réduction de coût** : analysez les services pour voir ceux que vous pouvez utiliser afin de réduire le coût.
- **Optimiser les coûts de licence**
- **Optimiser les approches sans serveur et basées sur les conteneurs** : utilisez AWS Lambda, les sites web Amazon S3, Amazon DynamoDB et Amazon ECS pour réduire les coûts.
- **Optimiser avec les solutions de stockage appropriées** : utilisez la solution de stockage la plus rentable à partir des modèles d'utilisation (stockage à froid Amazon EBS, Amazon S3 Standard-Infrequent Access, Amazon Glacier, etc.).
- **Optimiser avec les bases de données appropriées** : utilisez Amazon Relational Database Service (RDS) (Postgres, MySQL, SQL Server, Oracle Server) ou Amazon DynamoDB (ou autres magasins clé-valeur, alternatives NoSQL), le cas échéant.
- **Optimiser avec les autres services de niveau application** : utilisez Amazon Simple Queue Service (SQS), Amazon Simple Notification Service (SNS) ou Amazon Simple Email Service (SES), le cas échéant.

COUT 2. Avez-vous dimensionné les ressources pour satisfaire vos cibles de coût ?

Assurez-vous que vous choisissez la taille de ressource AWS adaptée à la tâche en cours. AWS encourage l'utilisation d'évaluations comparatives afin que vous vous assuriez que le type que vous choisissez est optimisé pour sa charge de travail.

Bonnes pratiques :

- **Dimensionnement des ressources pilotées par les métriques** : exploitez les métriques de performance pour sélectionner la taille/le type approprié à l'optimisation des coûts. Provisionnez de façon appropriée le débit, le dimensionnement et le stockage de services tels qu'Amazon EC2, Amazon DynamoDB, Amazon EBS (IOPS provisionnées), Amazon RDS, Amazon EMR, réseau, etc.

COUT 3. Avez-vous sélectionné le modèle de tarification approprié pour satisfaire vos cibles de coût ?

Utilisez le modèle de tarification le plus approprié à votre charge de travail pour réduire les dépenses. Le déploiement optimal peut être les instances intégralement à la demande, un mélange d'instances à la demande et d'instances réservées, ou inclure des instances ponctuelles, le cas échéant.

Bonnes pratiques :

- **Capacité réservée et validation** : analysez régulièrement l'utilisation et l'achat d'instances réservées en conséquence (Amazon EC2, Amazon DynamoDB, Amazon S3, Amazon CloudFront, par exemple, etc.).
- **Instances ponctuelles** : utilisez les instances ponctuelles (bloc ponctuel, flotte, par exemple) pour sélectionner les charges de travail (lot, EMR, par exemple, etc.).
- **Prendre en compte le coût de la région** : envisagez les coûts dans la sélection de la région.

Correspondance de l'offre et de la demande

COUT 4. Comment êtes-vous sûr que la capacité correspond à ce dont vous avez besoin, sans le dépasser de façon substantielle ?

Pour une architecture équilibrée en termes de dépense et de performances, assurez-vous que tout ce que vous payez est utilisé et évitez les instances par trop sous-utilisées. Une métrique d'utilisation faussée dans une direction ou l'autre aura un impact négatif sur votre activité, que ce soit dans les coûts d'exploitation (dégradation des performances due à une sur-utilisation) ou dans le gaspillage de dépenses AWS (en raison d'une sur-allocation).

Bonnes pratiques :

- **Approche basée sur la demande** : utilisez Auto Scaling pour répondre à la demande variable.
- **Approche basée sur la mémoire tampon** : la mémoire tampon (utilisation d'Amazon Kinesis ou d' Amazon Simple Queue Service (SQS), par exemple) permet de reporter le travail jusqu'à ce que vous ayez la capacité suffisante pour le traiter.
- **Approche basée sur le temps** : par exemple, fonctionnement 24 heures sur 24, désactiver les instances de développement/test le weekend, suivre des planifications trimestrielles ou annuelles (par exemple, le Black Friday).

Sensibilisation aux dépenses

COUT 5. Avez-vous considéré les charges de transfert des données lors de la conception de votre architecture ?

Assurez-vous que vous surveillez les charges liées au transfert de données afin de pouvoir prendre des décisions architecturales susceptibles d'alléger certains de ces coûts. Par exemple, si vous êtes un fournisseur de contenu et que vous proposez un contenu directement depuis un compartiment Amazon S3 à vos utilisateurs finaux, il se peut que vous puissiez réduire vos coûts de façon significative si vous publiez votre contenu sur le réseau de distribution de contenu Amazon CloudFront. N'oubliez pas qu'une modification architecturale petite, mais effective, peut réduire de façon spectaculaire vos coûts d'exploitation.

Bonnes pratiques :

- **Optimiser** : concevez l'architecture de façon à optimiser le transfert des données (conception de l'application, accélération WAN, Multi-AZ, sélection de région, etc.).
- **Réseau de distribution de contenu (CDN)** : utilisez un CDN chaque fois que possible.
- **AWS Direct Connect** : analysez la situation et utilisez AWS Direct Connect chaque fois que possible.

COUT 6. Comment surveillez-vous l'utilisation et les dépenses ?

Définissez des stratégies et des procédures pour surveiller, contrôler et affecter les coûts de façon appropriée. Tirez profit des outils AWS en matière de visibilité pour savoir qui utilise quoi, et à quel coût. Vous bénéficierez ainsi d'une connaissance plus approfondie de vos besoins métier et des opérations de votre équipe.

Bonnes pratiques :

- **Baliser toutes les ressources** : vous pourrez ainsi relier les modifications de facturation et les modifications d'infrastructure et d'utilisation.
- **Exploiter les outils de facturation et de gestion des coûts** : disposez d'un processus standard pour charger et interpréter les rapports de facturation détaillés ou l'Explorateur de coûts. Surveillez régulièrement l'utilisation et les dépenses à l'aide d'Amazon CloudWatch ou d'un fournisseur tiers chaque fois que possible (par exemple : Cloudability, CloudCheckr, CloudHealth).
- **Notifications** : permettez aux membres clés de votre équipe de savoir si nos dépenses excèdent des limites bien définies.
- **Méthode de refacturation orientée finances** : utilisez cette méthode pour allouer les instances et les ressources aux centres de coûts (par exemple, balisage).

COUT 7. Mettez-vous hors service les ressources dont vous n'avez plus besoin ou arrêtez-vous celles qui ne sont pas nécessaires temporairement ?

Implémentez le contrôle des modifications et la gestion des ressources depuis le début du projet jusqu'à la fin, de telle sorte que vous puissiez identifier les modifications ou améliorations de processus nécessaires, le cas échéant. Utilisez AWS Support pour les recommandations sur l'optimisation de votre projet pour votre charge de travail : par exemple, déterminer à quel moment utiliser Auto Scaling, AWS OpsWorks, AWS Data Pipeline ou les différentes approches d'allocation Amazon EC2, ou vérifiez les recommandations de Trust Advisor sur l'optimisation des coûts.

Bonnes pratiques :

- **Automatisation** : concevez votre système de façon à bien gérer la terminaison d'une ressource tandis que vous identifiez et mettez hors service les instances non critiques ou non requises, ou les ressources avec une faible utilisation.
- **Processus défini** : ayez un processus en place pour identifier et mettre hors service les ressources orphelines.

COUT 8. Quels contrôles d'accès et procédures avez-vous en place pour régir l'utilisation d'AWS ?

Définissez des stratégies et des mécanismes pour vous assurer que les coûts appropriés sont facturés lorsque les objectifs sont atteints. En adoptant une approche d'équilibre des pouvoirs via les balises et les contrôles IAM, vous pouvez innover sans dépense excessive.

Bonnes pratiques :

- **Etablir les groupes et les rôles** : (exemple : développement/ test/production) utilisez les mécanismes de gouvernance AWS tels qu'IAM pour contrôler les personnes autorisées à faire tourner les instances et les ressources dans chaque groupe. (Ceci s'applique aux services AWS ou aux solutions tierces.)
- **Suivre le cycle de vie du projet** : suivez, mesurez et auditez le cycle de vie des projets, équipes et environnements pour éviter l'utilisation et le paiement de ressources superflues.

Optimisation au fil du temps

COUT 9. Comment gérez-vous et/ou envisagez-vous l'adoption de nouveaux services ?

Tandis qu'AWS propose de nouveaux services et de nouvelles fonctionnalités, une bonne pratique consiste à vérifier vos décisions architecturales existantes afin de garantir qu'elles continuent à être les plus économiques.

Bonnes pratiques :

- **Etablir une fonction d'optimisation des coûts**
- **Vérifier** : ayez un processus pour vérifier les nouveaux services, types de ressource et tailles. Exécutez à nouveau les tests de performance pour évaluer les réductions de coût éventuelles.

Pilier « Excellence opérationnelle »

Préparation

OPE 1. Quelles bonnes pratiques utilisez-vous pour les opérations du cloud ?

Une préparation efficace est requise pour piloter l'excellence opérationnelle. Les listes de contrôle des opérations permettent de s'assurer que les charges de travail sont prêtes pour une exploitation en production. Les listes de contrôle empêchent une promotion en production non intentionnelle sans préparation efficace.

Bonnes pratiques :

- **Liste de contrôle opérationnelle** : créez une liste de contrôle opérationnelle que vous utilisez pour déterminer si vous êtes prêt à exploiter la charge de travail.
- **Plan proactif** : ayez un plan proactif pour les événements (campagnes marketing, ventes flash, par exemple) qui vous prépare aux opportunités et aux risques susceptibles d'avoir un impact matériel sur votre activité (réputation, finances, par exemple).
- **Liste de contrôle de sécurité** : créez une liste de contrôle de sécurité que vous utilisez pour déterminer si vous êtes prêt à exploiter la charge de travail en toute sécurité (jour zéro, attaque DDoS, clés compromises, par exemple).

OPE 2. Comment assurez-vous la gestion de la configuration de votre charge de travail ?

Les environnements, l'architecture et les paramètres de configuration des ressources en leur sein doivent être documentés d'une façon qui permet d'identifier aisément les composants à des fins de suivi et de dépannage. Les modifications apportées à la configuration doivent également être suivies et automatisées.

Bonnes pratiques :

- **Suivi des ressources** : planifiez les solutions d'identification de vos ressources et de leur fonction au sein de la charge de travail (métadonnées, balisage, par exemple).
- **Documentation** : documentez votre architecture (infrastructure en tant que code, base de données de gestion de la configuration, diagrammes, notes de publication).
- **Capturer les apprentissages opérationnels** : capturez les apprentissages opérationnels au fil du temps (wiki, base de connaissances, tickets, par exemple).
- **Infrastructure immuable** : définissez une infrastructure immuable de telle sorte que vous redéployez, et non corrigez.
- **Procédures de modification automatiques** : automatiser vos procédures de modification.
- **Base de données de gestion de la configuration** : suivez toutes les modifications dans une base de données de gestion de la configuration.

Transactions

OPE 3. Comment faites-vous évoluer votre charge de travail tout en réduisant l'impact des modifications ?

L'accent doit être mis sur l'automatisation, les modifications réduites et fréquentes, les tests réguliers d'assurance qualité et les mécanismes définis, pour suivre, auditer, annuler ou vérifier les changements.

Bonnes pratiques :

- **Pipeline de déploiement** : mettez en place un pipeline d'intégration continue / de déploiement continu (référentiel de code source, systèmes de génération, automatisation du déploiement et des tests).
- **Processus de gestion des versions** : établissez un processus de gestion des versions (manuel ou automatique).
- **Petites modifications incrémentielles** : assurez-vous que vous pouvez publier de petites versions incrémentielles des composants système.
- **Modifications réversibles** : soyez prêt à annuler les modifications qui introduisent des problèmes opérationnels (annulation, bascules de fonctionnalité, par exemple).
- **Stratégies d'atténuation des risques** : utilisez des stratégies d'atténuation des risques, telles que Blue/Green, Canary et tests A/B.

OPS 4. Comment surveillez-vous votre charge de travail pour vous assurer qu'elle se comporte comme prévu ?

Les performances du système peuvent se dégrader au fil du temps en raison de facteurs internes et/ou externes. En surveillant le comportement de vos systèmes, vous pouvez identifier ces facteurs de dégradation et les corriger.

Bonnes pratiques :

- **Supervision** : utilisez Amazon CloudWatch, des outils tiers ou des outils de supervision personnalisés pour surveiller les performances.
- **Agréger les journaux** : agrégez les journaux de plusieurs sources (journaux d'application, journaux propres aux services AWS, journaux de flux VPC, CloudTrail).
- **Notifications basées sur les alarmes** : recevez une alerte automatique de votre système de supervision si les métriques excèdent les limites sécurisées.
- **Actions basées sur les déclencheurs :** les alarmes déclenchent des actions automatiques pour corriger un problème ou le faire remonter.

Réponses

OPE 5. Comment répondez-vous aux événements opérationnels non prévus ?

Préparez-vous à automatiser les réponses aux événements opérationnels non prévus. Cela ne concerne pas seulement les alertes, mais aussi l'atténuation, la correction, la restauration et la récupération.

Bonnes pratiques :

- **Playbook** : ayez un playbook que vous suivez (processus sur appel, chaîne de workflow, processus de réaffectation) et mettez à jour régulièrement.
- **Processus d'analyse des causes premières** : ayez un processus d'analyse des causes premières pour vous assurer de résoudre, documenter et corriger les problèmes de telle sorte qu'ils ne se produisent pas à l'avenir.
- **Automatiser les réponses** : gérez les événements opérationnels imprévus élégamment via les réponses automatiques (Auto Scaling, API Support, par exemple).

OPE 6. Comment gérez-vous la réaffectation lorsque vous répondez à des événements opérationnels non prévus ?

Les réponses aux événements opérationnels non prévus doivent suivre un playbook prédéfini qui inclut les parties prenantes, le processus de réaffectation et les procédures. Définissez les chemins de réaffectation et incluez les capacités de réaffectation hiérarchique et fonctionnelle. La réaffectation hiérarchique doit être automatique et la priorité réaffectée doit se traduire dans les notifications des parties prenantes.

Bonnes pratiques :

- **Documenter et allouer de façon appropriée** : mettez en place les parties prenantes et systèmes nécessaires pour recevoir les alertes en cas de réaffectation.
- **Réaffectation fonctionnelle avec approche basée sur la file d'attente** : réaffectez entre les files d'attente des équipes fonctionnelles appropriées en fonction de la priorité, de l'impact et des mécanismes d'admission.
- **Réaffectation hiérarchique** : utilisez une approche basée sur la demande ou sur le temps. Tandis que l'impact, l'échelle ou le temps de résolution/récupération de l'incident augmente, la priorité est réaffectée.
- **Chemin de réaffectation externe** : incluez le support externe, le support AWS, les partenaires AWS et l'engagement d'un support tiers dans les chemins de réaffectation.
- **Automatiser la réaffectation de priorité hiérarchique** : lorsque les seuils de demande ou de temps sont franchis, la priorité est automatiquement réaffectée.