Infrastructure AWS correctement architecturée

Octobre 2015

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 avis préalable. 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 ne crée pas de garanties, représentations, engagements contractuels, conditions ou assurances à l'encontre d'AWS, de ses affiliés, fournisseurs ou donneurs 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 et ne modifie aucun contrat entre AWS et ses clients.

Table des matières

Résumé

Ce livre blanc décrit l'**infrastructure AWS correctement architecturée**, qui permet aux clients d'évaluer et d'améliorer leurs architectures cloud et 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 quatre 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 dans le 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 la là, 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. Au fur et à mesure que la plateforme AWS évoluera et que notre collaboration avec les clients sera de plus en plus riche d'enseignements, nous continuerons 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 saurez quelles sont les stratégies et les bonnes pratiques AWS à utiliser lors de la conception d'une architecture cloud. Il ne contient pas de modèles d'architecture ou de détails sur l'implémentation ; cependant, il propose des références aux ressources appropriées où figurent 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 quatre piliers : la sécurité, la fiabilité, l'efficacité des performances et l'optimisation des coûts, que nous définissons comme suit :

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 d'un incident dans une infrastructure ou un service, à acquérir dynamiquement des ressources informatiques pour répondre à la demande, et à réduire les perturbations telles qu'erreurs de configuration ou 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.

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 :

- **Cessez de présumer vos 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.
- **Testez les systèmes à l'échelle de la production :** dans un environnement traditionnel qui ne s'appuie pas sur le cloud, la création en double d'un environnement aux seules fins de test entraîne généralement un coût prohibitif. Par conséquent, la plupart des environnements de test ne sont pas testés aux niveaux réels de la demande de production. Dans le cloud, vous pouvez créer un environnement dupliqué à la demande, 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.
- **Réduisez le risque d'une modification d'architecture :** comme vous pouvez automatiser la création d'environnements de test qui émulent vos configurations de production, les tests peuvent être exécutés sans peine. Vous pouvez aussi supprimer la sérialisation des tests qui intervient dans les environnements locaux, où les équipes doivent passer par une file d'attente pour utiliser les ressources des tests.
- **Recourez à l'automatisation pour faciliter l'expérimentation architecturale :** l'automatisation vous permet de créer et de répliquer vos systèmes à moindre coût (pas d'effort manuel). Vous pouvez suivre les modifications apportées à l'automatisation, auditer l'impact et rétablir les paramètres antérieurs si nécessaire.
- **Autorisez les architectures évolutives :** dans un environnement traditionnel, les décisions architecturales sont souvent implémentées comme un événement unique et statique, 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 peuvent tirer profit des nouvelles innovations en tant que pratique standard.

Les quatre piliers de l'infrastructure AWS correctement architecturée

Créer un système logiciel s'apparente à la construction d'un immeuble. Si la fondation n'est pas solide, des problèmes structurels risquent d'apparaître qui sapent l'intégrité et la fonction de l'immeuble. Lors de la conception architecturale de solutions technologiques, si vous négligez les quatre piliers de la sécurité, de la fiabilité, de l'efficacité des performances et de l'optimisation des coûts, la création d'un système qui satisfait 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 quatre piliers et inclut les définitions, bonnes pratiques, questions, considérations et services clés AWS appropriés.

Pilier « Sécurité »

Le pilier **Sécurité** englobe la 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.

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 simplement exécuter les dispositifs de sécurité (les pare-feux, par exemple) à la pointe de votre infrastructure, utilisez les pare-feux et autres contrôles de sécurité sur l'ensemble de vos ressources (chaque serveur virtuel, répartiteur de charge et sous-réseau).
- **Activer la traçabilité :** enregistrez et auditez toutes les actions et toutes les modifications apportées à votre environnement.
- **Automatiser les réponses aux événements de sécurité :** assurez une surveillance et déclenchez automatiquement les réponses aux alertes basées sur les événements ou sur les conditions.

- **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 de référence personnalisée d'un serveur virtuel, puis utilisez automatiquement cette image sur chaque nouveau serveur que vous lancez. Créez une infrastructure complète définie et gérée dans un modèle.

Définition

La sécurité dans le cloud se compose de quatre zones :

1. Protection des données
2. Gestion des privilèges
3. Protection de l'infrastructure
4. Contrôles de détection

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

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 ; le *privilège minimum* limite l'accès au niveau le plus bas possible tout en continuant à autoriser les fonctions normales ; et, enfin, 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.

La protection des données implique l'utilisation de contrôles et de modèles destinés à maintenir la confidentialité de vos données tout en préservant leur intégrité et en garantissant leur disponibilité lorsque vous en avez besoin.

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 sont axées sur des considérations relatives à la sécurité des données (pour obtenir la liste des questions, réponses et meilleures pratiques relatives à la sécurité, consultez l'annexe) :

SEC 1. Comment chiffrez-vous et protégez-vous vos données au repos ?

SEC 2. 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 qu'Amazon S3 vous permette de stocker plus facilement vos données sous une forme 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.

Gestion des privilèges

La gestion des privilèges est un aspect essentiel du programme de sécurité des informations ; il garantit que seuls les utilisateurs autorisés et authentifiés puissent accéder à vos ressources, et uniquement de la manière prévue. Par exemple, une liste de contrôle d'accès (ACL) est une liste d'autorisations d'accès attachées à un objet, les contrôles d'accès basés sur les rôles sont un ensemble d'autorisations conformes au rôle ou à la fonction d'un utilisateur final, et la gestion des mots de passe inclut des exigences de complexité et des intervalles de modification. Ces éléments de gestion des privilèges sont essentiels dans une architecture de sécurité des informations, car ils représentent les concepts centraux de l'authentification utilisateur 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 la complexité, la réutilisation et l'authentification multifacteur (MFA) ; vous pouvez aussi utiliser la fédération avec votre service d'annuaire existant.

Les questions suivantes portent essentiellement sur la gestion des privilèges en matière de sécurité :

SEC 3. comment protégez-vous l'accès aux informations d'identification du compte racine (root) AWS et leur utilisation ?

SEC 4. Comment définissez-vous les rôles et les responsabilités des utilisateurs système pour contrôler l'accès humain à AWS Management Console et aux API ?

SEC 5. Comment limitez-vous l'accès automatique (à partir d'applications, de scripts ou d'outils et services tiers, par exemple) aux ressources AWS ?

SEC 6. Comment gérez-vous les clés et les informations d'identification ?

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.

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

SEC 8. Comment appliquez-vous la protection au niveau des services AWS ?

SEC 9. 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.

Contrôles de détection

Vous pouvez utiliser les contrôles de détection pour détecter ou identifier une faille de sécurité. Ils constituent une partie normale des infrastructures de gouvernance et peuvent être utilisés pour prendre en charge un processus de qualité, une obligation de conformité légale et/ou une identification des menaces et les tentatives de réponse. Il existe différents types de contrôles de détection. Par exemple, l'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, les services suivants prennent en charge les contrôles de détection :

- **AWS CloudTrail :** service web qui enregistre les appels d'API, dont l'identité de l'appel, la durée de l'appel, l'adresse IP source, les paramètres et les éléments de réponse.
- **Amazon CloudWatch :** service de supervision des ressources AWS qui enregistre différents aspects tels que, entre autres, l'activité unité centrale, disque et réseau d'Amazon Elastic Compute Cloud (EC2), les instances de base de données Amazon Relational Database Service (RDS) et les volumes Amazon Elastic Block Store (EBS). CloudWatch offre la possibilité de définir des alarmes sur ces métriques et sur d'autres.
- **AWS Config :** service d'inventaire et d'historique de configuration qui fournit des informations sur les configurations et les modifications de l'infrastructure au fil du temps.

- **Amazon Simple Storage Service (S3) :** avec l'audit de l'accès aux données Amazon S3, les clients peuvent configurer les compartiments Amazon S3 pour enregistrer les détails des demandes d'accès, y compris le type, la ressource, la date et l'heure.
- **Amazon Glacier :** les clients peuvent utiliser la 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 10. Comment capturez-vous et analysez-vous les journaux AWS ?

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. 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.

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 :

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.

Gestion des privilèges : IAM vous permet de contrôler de façon sécurisée 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.

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.

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.

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 de 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.

Définition

La fiabilité dans le cloud se compose de trois zones :

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 système, les exigences en termes de fondation 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 sont axées sur des considérations relatives à la fiabilité (pour obtenir la liste complète des questions, réponses et bonnes pratiques relatives à la fiabilité, consultez l'annexe) :

REL 1. Comment gérez-vous les limites des services AWS pour votre compte ?

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

REL 3. Disposez-vous d'un chemin de réaffectation pour traiter les problèmes techniques ?

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 surveille 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. Enfin, vous voudrez vous assurer que votre équipe informatique dispose de la formation et des processus mis à jour pour prendre en charge l'utilisation d'un cloud public, ainsi que des contrats effectifs de partenariat ou d'assistance, le cas échéant.

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 :

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

REL 5. Comment surveillez-vous les ressources AWS ?

REL 6. Comment exécutez-vous la gestion des 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 :

REL 7. Comment sauvegardez-vous les données ?

REL 8. Comment votre système supporte-t-il les défaillances de composants ?

REL 9. Comment planifiez-vous la récupération ?

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 l'adéquation d'un système (notamment dans les scénarios de test de défaillance) et pour vous aider 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 permet la création d'un modèle des 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 Documentation (Documentation sur les limites des services)
- Service Limit Reports Blog Post (Publication du blog sur les rapports des limites de services)

Livres blancs

- Backup Archive and Restore Approach Using AWS Whitepaper (Livre blanc sur l'archivage des sauvegardes et la restauration à l'aide d'AWS)
- Managing your AWS Infrastructure at Scale Whitepaper (Livre blanc sur la gestion de votre infrastructure AWS à l'échelle)
- AWS Disaster Recovery Whitepaper (Livre blanc sur la reprise après sinistre AWS)
- AWS Amazon VPC Connectivity Options Whitepaper (Livre blanc sur les options de connexion 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, de façon simple et 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, peut réduire aussi 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.

Définition

L'**efficacité des performances** dans le cloud se compose de quatre zones :

1. Calcul
2. Stockage
3. Base de données
4. Compromis espace-temps

Les considérations relatives à chacune de ces zones sont les suivantes : a) comment sélectionner l'approche et les ressources optimales, b) comment maintenir l'actualité de cette approche au vu de l'évolution des capacités du cloud, c) comment surveiller les performances en temps réel par rapport aux attentes, et, enfin, d) comment dimensionner les ressources par rapport à la demande.

Bonnes pratiques

Calcul

La configuration optimale d'un serveur pour une architecture particulière peut varier selon la conception de l'application, les modèles d'utilisation et les paramètres de configuration. De nombreux systèmes utilisent différentes configurations de serveur pour divers composants et activent différentes fonctions pour améliorer les performances. La sélection d'une configuration serveur incorrecte dans un cas d'utilisation donné peut conduire à une efficacité moindre des performances.

Dans AWS, les serveurs sont virtualisés et, par conséquent, vous pouvez modifier leurs capacités d'un simple clic ou appel d'API. Comme 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 SSD et GPU. Dans AWS, il est également possible d'effectuer des calculs sans serveur. Par exemple, AWS Lambda vous permet d'exécuter du code sans exécuter une instance.

Les questions suivantes sont axées sur des considérations relatives au calcul (pour obtenir la liste complète des questions, réponses et bonnes pratiques relatives à l'efficacité des performances, consultez l'Annexe) :

PERF 1.	**Comment sélectionnez-vous le type d'instance approprié pour votre système ?**
PERF 2.	**Comment savez-vous que vous continuez à avoir le type d'instance le plus approprié tandis que de nouveaux types d'instance et de nouvelles fonctionnalités sont introduites ?**
PERF 3.	**Comment surveillez-vous les instances après leur lancement pour vous assurer qu'elles se comportent comme prévu ?**
PERF 4.	**Comment garantissez-vous que la quantité de vos instances correspond à la demande ?**

Lors de la sélection des types d'instance à utiliser, il importe d'avoir des données de test qui montrent les types d'instance (ou approches sans serveur) correspondant le mieux à *cette* charge de travail. Ces tests doivent être reproductibles (idéalement dans le cadre d'un pipeline de livraison continue) de telle sorte que vous puissiez aisément tester de nouveaux types d'instance ou de nouvelles capacités au fur et à mesure qu'elles deviennent disponibles. D'un point de vue opérationnel, la supervision doit être en place pour que vous puissiez être informé de toute dégradation des performances.

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 portent essentiellement sur le stockage en termes d'efficacité des performances :

PERF 5. Comment sélectionnez-vous la solution de stockage appropriée pour votre système ?

PERF 6. Comment savez-vous que vous continuez à avoir la solution de stockage la plus appropriée tandis que de nouvelles fonctionnalités et solutions de stockage sont lancées ?

PERF 7. Comment surveillez-vous votre solution de stockage pour vous assurer qu'elle se comporte comme prévu ?

PERF 8. Comment garantissez-vous que la capacité et le débit de vos solutions de stockage correspondent à la demande ?

Lors de la sélection d'une solution de stockage, il importe d'avoir des données de test qui montrent quelle solution de stockage offrira la marge coût/valeur requise pour *cette* charge de travail. Ces tests doivent être reproductibles (idéalement dans le cadre d'un pipeline de livraison continue) de telle sorte que vous puissiez aisément tester de nouvelles solutions de stockage ou de nouvelles capacités au fur et à mesure qu'elles deviennent disponibles. Les types de stockage (EBS contre stockage d'instance ou HDD contre SSD) utilisés pour différentes instances peuvent substantiellement modifier l'efficacité des performances de votre système. D'un point de vue opérationnel, la supervision doit être en place pour que vous puissiez être informé de toute dégradation des performances.

Base de données

La solution de base de données optimale pour un système particulier peut varier en fonction des exigences de cohérence, de disponibilité, de tolérance des partitions et de latence. 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. Nous proposons également d'autres solutions de base de données et de stockage. 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 portent essentiellement sur les bases de données en termes d'efficacité des performances :

PERF 9. Comment sélectionnez-vous la solution de base de données appropriée pour votre système ?

PERF 10. Comment savez-vous que vous continuez à avoir la solution de base de données et les fonctionnalités les plus appropriées tandis que de nouvelles fonctionnalités et solutions de base de données sont lancées ?

PERF 11. Comment surveillez-vous vos bases de données pour vous assurer que les performances sont celles attendues ?

PERF 12. Comment garantissez-vous que la capacité et le débit de vos bases de données correspondent à la demande ?

Même si l'approche de base de données (RDBMS, NoSQL, etc.) d'une organisation a un impact réel sur l'efficacité des performances d'un système, il s'agit souvent d'un domaine choisi conformément aux valeurs par défaut de l'organisation plutôt qu'au travers d'une évaluation. Pendant le développement et le déploiement de votre solution de base de données, traitez la base de données comme du code pour lui permettre d'évoluer au fil du temps plutôt que comme une décision ponctuelle fixe. Utilisez les données de test pour identifier la solution de base de données correspondant le mieux à chaque charge de travail. Ces tests doivent être reproductibles (idéalement dans le cadre d'un pipeline de livraison continue) de telle sorte que vous puissiez aisément tester de nouvelles solutions de base de données ou de nouvelles capacités au fur et à mesure qu'elles deviennent disponibles. Par exemple, évaluez si les réplicas en lecture seule améliorent l'efficacité des performances sans enfreindre d'autres exigences non fonctionnelles. D'un point de vue opérationnel, la supervision doit être en place pour que vous puissiez être informé de toute dégradation des performances.

Compromis espace-temps

Lors de la conception architecturale des solutions, il existe un certain nombre de compromis où l'espace (espace mémoire ou espace de stockage) permet de réduire le temps de traitement (calcul), ou bien où le temps permet de réduire l'espace. Vous pouvez aussi placer les ressources ou les données mises en cache plus près des utilisateurs finaux afin de réduire le temps.

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.

Utilisez l'infrastructure globale d'AWS pour parvenir à diminuer la latence et à augmenter le débit, et vous assurer que vos données ne résident que dans les régions que vous spécifiez. Les solutions réseau comme AWS Direct Connect sont conçues pour fournir une latence prévisible entre votre réseau local et votre infrastructure AWS. AWS propose aussi des solutions de mise en cache, telles qu'Amazon ElastiCache, qui aident à améliorer l'efficacité, 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 13. Comment sélectionnez-vous les solutions appropriées de proximité et de mise en cache pour votre système ?

PERF 14. Comment savez-vous que vous continuez à avoir les solutions les plus appropriées de proximité et de mise en cache tandis que de nouvelles solutions sont lancées ?

PERF 15. Comment surveillez-vous vos solutions de proximité et de mise en cache pour vous assurer que les performances sont celles attendues ?

PERF 16. Comment garantissez-vous que vos solutions de proximité et de mise en cache correspondent à la demande ?

Les compromis espace-temps sont obligatoires pour garantir l'efficacité des performances et il est important d'avoir des données de test qui montrent quels compromis correspondent le mieux à *cette* charge de travail. Ces tests doivent être reproductibles (idéalement dans le cadre d'un pipeline de livraison continue) de telle sorte que vous puissiez aisément tester de nouvelles approches ou de nouvelles capacités au fur et à mesure qu'elles deviennent disponibles. Par exemple, procédez à des tests pour évaluer si l'utilisation d'Amazon ElastiCache comme cache en écriture directe améliore l'efficacité des performances sans enfreindre d'autres exigences non fonctionnelles. D'un point de vue opérationnel, la supervision doit être en place pour que vous puissiez être informé de toute dégradation des performances. L'architecture doit évoluer avec la demande et conserver sa marge.

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 quatre domaines d'efficacité des performances :

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 fournit un stockage à redondance réduite (RRS, Reduced-Redundancy Storage), des stratégies de cycle de vie pour Amazon Glacier (stockage d'archives) et une livraison de contenu sans serveur.

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.

Compromis espace-temps : AWS possède des régions à travers le monde, qui vous permettent de choisir l'emplacement optimal pour vos ressources, données et traitements. Utilisez Amazon CloudFront pour mettre en cache le contenu encore plus près de vos utilisateurs.

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

- Amazon S3 Performance Optimization Documentation (Documentation sur l'optimisation des performances Amazon S3)
- Amazon EBS Volume Performance Documentation (Documentation sur les performances des volumes Amazon EBS)

Pilier « Optimisation des coûts »

Utilisez le pilier **Optimisation des coûts** pour évaluer votre capacité à éviter ou à supprimer les coûts superflus ou les ressources sous-optimales, et à utiliser ces gains sur les avantages différenciés pour votre activité. Un système à coût optimisé vous permet de payer le prix le plus bas possible tout en continuant à atteindre vos objectifs métier et à satisfaire, ou dépasser, les principales exigences des autres piliers correctement architecturés. Vous pouvez parvenir à l'optimisation des coûts grâce à des techniques qui vous permettent de sélectionner l'architecture appropriée, de réduire les ressources non utilisées et de sélectionner l'approche la plus économique.

Principes de conception

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

- **Attribution transparente des dépenses :** le cloud facilite l'identification du coût d'un système et l'attribution des coûts informatiques à leurs responsables métier. L'identification du retour sur investissement en est simplifiée et, par conséquent, ces responsables bénéficient d'une incitation à optimiser leurs ressources et à 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. En outre, comme les services gérés interviennent à l'échelle du cloud, ils peuvent offrir un coût moindre par transaction ou service.
- **Echange des dépenses d'investissement contre les dépenses d'exploitation :** au lieu d'investir massivement dans les centres de données et les serveurs avant de savoir comment vous allez les utiliser, ne payez que les ressources informatiques que vous consommez, quand vous les consommez. Par exemple, comme les environnements de développement et de test ne sont généralement utilisés que huit heures par jour pendant la semaine de travail, vous pouvez arrêter ces ressources quand elles ne sont pas utilisées et profiter d'une économie de coût de 75 % (40 heures au lieu de 168 heures).
- **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.

Définition

L'optimisation des coûts dans le cloud se compose de quatre zones :

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

Comme pour les autres piliers, il y a des compromis à prendre en compte : par exemple, entre l'optimisation de la rapidité de mise sur le marché et l'optimisation des coûts. 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

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 le temps ou les files d'attente, ou guidées par les événements, 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 sont axées sur la correspondance entre l'offre et la demande à des fins d'optimisation des coûts (pour obtenir la liste complète des questions, réponses et bonnes pratiques relatives à l'optimisation des coûts, consultez l'Annexe) :

COST 1. Comment êtes-vous sûr que la capacité correspond à ce dont vous avez besoin, sans le dépasser de façon substantielle ?

COST 2. Comment optimisez-vous votre utilisation des services AWS ?

Les outils de supervision et les tests réguliers peuvent vous aider à parvenir à une bien meilleure utilisation des ressources. La flexibilité de l'informatique à la demande, Auto Scaling et autres mécanismes de déploiement automatisés permettent un plus haut degré d'optimisation, en garantissant que vous n'allouez que les ressources dont vous avez besoin et que vous êtes à même de dimensionner horizontalement.

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 économiques, 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. 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 :

COST 3. Avez-vous sélectionné les types de ressources appropriés pour répondre à vos cibles de coût ?

COST 4. Avez-vous sélectionné le modèle de tarification approprié pour répondre à vos cibles de coût ?

COST 5. Existe-t-il des services gérés (services de niveau plus élevés qu'Amazon EC2, Amazon EBS et Amazon S3) que vous utilisez pour améliorer votre retour sur investissement ?

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. Vous pouvez également tirer parti des services AWS gérés, comme Amazon RDS, Amazon Elastic MapReduce (EMR) et Amazon DynamoDB, qui peuvent réduire les coûts de gestion et par élément. Pensez aux solutions CDN telles qu'Amazon CloudFront pour réduire potentiellement vos coûts associés au trafic réseau.

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 :

COST 6.	**Quels contrôles d'accès et procédures avez-vous en place pour régir les coûts AWS ?**
COST 7.	**Comment surveillez-vous l'utilisation et les dépenses ?**
COST 8.	**Comment mettez-vous hors service les ressources dont vous n'avez plus besoin ou arrêtez-vous celles qui ne sont pas nécessaires temporairement ?**
COST 9.	**Comment considérez-vous les charges de transfert des données lors de la conception de votre architecture ?**

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.

Avec cette visibilité des coûts par rapport aux ressources balisées, il devient plus facile 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 réévaluer 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 :

COST 10. Comment gérez-vous et/ou envisagez-vous l'adoption de nouveaux services ?

En réévaluant régulièrement votre déploiement, il est souvent possible d'utiliser les nouveaux services AWS pour diminuer vos coûts. De même, évaluez l'applicabilité de services plus récents pour vous aider à réaliser des économies : par exemple, AWS RDS pour Aurora peut contribuer 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 :

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

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.

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 AWS

Outils

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

Conclusion

L'infrastructure AWS correctement architecturée fournit les bonnes pratiques à travers quatre piliers permettant de concevoir dans le cloud des systèmes fiables, sécurisés et économiques. L'infrastructure AWS documente un ensemble de questions qui vous permettent d'évaluer 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 programmes senior, Amazon Web Services
- Callum Hughes, architecte de solutions, Amazon Web Services
- Max Ramsay, architecte principal de solutions de sécurité, Amazon Web Services
- Scott Paddock, architecte de solutions de sécurité, Amazon Web Services

Historique du document

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

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 »

SEC 1. 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, etc.

Bonnes pratiques :

- 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.).
- Les données au repos sont chiffrées à l'aide de techniques côté client.
- Solution d'AWS Marketplace ou d'un partenaire APN.

SEC 2. 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 :

- Les API AWS SSL sont utilisées de manière appropriée.
- Le protocole SSL ou équivalent est utilisé pour la communication.
- Solution basée sur VPN.
- Connexion privée (par exemple, AWS Direct Connect).
- La solution AWS Marketplace est utilisée.

SEC 3. 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 dispositif matériel d'authentification multifacteur (MFA) ; de cette façon, la seule utilisation possible de l'identité racine s'effectue via AWS Management Console et le recours aux appels d'API (Application Programming Interface) n'est pas autorisé. 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 :

- Les informations d'identification du compte racine ne sont utilisées que pour les activités minimales obligatoires.
- Il existe un dispositif matériel à authentification multifacteur MFA associé au compte racine AWS.
- La solution AWS Marketplace est utilisée.

SEC 4. 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 :

- Utilisateurs et groupes IAM
- Intégration SAML
- Fédération des identités Web
- AWS Security Token Service (STS)
- Rôles IAM pour l'accès entre comptes
- Solution d'AWS Marketplace (par exemple, Okta, Ping Identity) ou d'un partenaire APN
- Les stratégies de cycle de vie des employés sont définies et appliquées
- Les utilisateurs, les groupes et les rôles sont clairement définis et ne leur sont accordés que les privilèges minimum nécessaires pour accomplir les besoins métier

SEC 5. Comment limitez-vous l'accès automatique aux ressources AWS ? (par exemple, applications, scripts et/ou outil ou service tiers)

L'accès systématique doit être défini de manière similaire à celle dont 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 :

- Rôles IAM pour Amazon EC2
- Les informations d'identification IAM sont utilisées, mais ne sont pas codées en dur dans les scripts et les applications
- Intégration SAML
- AWS Security Token Service (STS)
- Les contrôles propres au système d'exploitation sont utilisés pour les instances EC2.
- La solution AWS Marketplace est utilisée.

SEC 6. Comment gérez-vous les clés et les informations d'identification ?

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

Bonnes pratiques :

- Une stratégie appropriée de rotation des clés et des informations d'identification est utilisée.
- Utilisez AWS CloudHSM.
- Les techniques AWS côté serveur sont utilisées avec les clés gérées AWS (par exemple, chiffrement côté serveur Amazon S3, volumes chiffrés Amazon EBS, etc.).
- Solutions AWS Marketplace (par exemple, SafeNet, TrendMicro, etc.).

SEC 7. 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 avec 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 :

- Les groupes de sécurité avec les autorisations minimales permettent d'appliquer l'accès basé sur les rôles.
- Le système s'exécute dans un ou plusieurs VPC.
- L'accès à un VPC fiable s'exécute via un mécanisme privé (par exemple, réseau privé virtuel (VPN), tunnel IPsec, AWS Direct Connect, solution AWS Marketplace, etc.).
- Les sous-réseaux et les ACL réseau sont utilisés en conséquence.
- Les pare-feux basés sur les hôtes avec les autorisations minimales sont utilisés.
- Les contrôles d'accès spécifiques aux services sont utilisés (par exemple, stratégies de compartiment).
- La connexion privée à un VPC est utilisée (par exemple, VPN, AWS Direct Connect, homologation VPC, etc.)
- La technique d'hôte bastion permet de gérer les instances.
- Les tests de sécurité sont exécutés régulièrement.
- Les contrôles AWS Trusted Advisor sont vérifiés régulièrement.

SEC 8. Comment appliquez-vous la protection au niveau des services AWS ?

Une autre bonne pratique consiste à contrôler l'accès aux ressources. AWS Identity and Access Management (IAM) permet de définir différents contrôles au niveau des ressources (par exemple, utilisation du chiffrement, heure du jour, IP source, etc.) et divers services permettent l'utilisation de techniques supplémentaires (par exemple, stratégies de compartiment Amazon S3, etc.). En outre, les clients peuvent utiliser différentes techniques au sein de leurs instances Amazon EC2.

Bonnes pratiques :

- Informations d'identification configurées avec le moindre privilège.
- Séparation des fonctions.
- Audit régulier des autorisations.
- Les exigences des ressources sont définies pour les appels d'API sensibles, telles que l'obligation d'authentification multifacteur (MFA) et le chiffrement.
- Les exigences spécifiques au service sont définies et utilisées.
- La solution AWS Marketplace est utilisée.

SEC 9. 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 EC2 à l'aide des techniques traditionnelles basées sur les hôtes (par exemple, OSSEC, Tripwire, Trend Micro Deep Security, etc.).

Bonnes pratiques :

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

SEC 10. Comment capturez-vous et analysez-vous les journaux AWS ?

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 :

- AWS CloudTrail.
- Journaux Amazon CloudWatch.
- Journaux Elastic Load Balancing (ELB).
- Journaux de filtre Amazon Virtual Private Cloud (VPC).
- Journaux de compartiment Amazon S3.

- Autres sources de journaux spécifiques aux services AWS.
- Journaux du système d'exploitation ou des applications tierces.
- La solution AWS Marketplace est utilisée.

Pilier « Fiabilité »

REL 1. Comment gérez-vous les limites des services AWS pour votre compte ?

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 limites de service fixes :** connaissez les limites de service non modifiables et de leur architecture.

REL 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 hautement disponibles à AWS :** plusieurs circuits DX, plusieurs tunnels VPN, appliances AWS Marketplace.
- **Connexion hautement disponible au système :** 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 doivent 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.

REL 3. Disposez-vous d'un chemin de réaffectation pour traiter les problèmes techniques ?

Les clients doivent tirer parti d'AWS Support ou d'un partenaire AWS. Une interaction régulière permettra de traiter et de prévenir les problèmes identifiés, les écarts de connaissance et les difficultés de conception. Le risque d'implémentation de défaillances, ainsi que de pannes à grande échelle, s'en trouve réduit.

Bonnes pratiques :

- **Planification :** engagement/relation continu(e) avec AWS Support ou un partenaire APN.
- **Exploiter les API AWS Support :** intégrez l'API AWS Support à vos systèmes de tickets et de supervision interne.

REL 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.

REL 5. 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.
- **Vérification :** exécutez des vérifications régulières du système basées sur les événements significatifs pour évaluer l'architecture.

REL 6. Comment exécutez-vous la gestion des modifications ?

La gestion des modifications 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 contrôlée.

Bonnes pratiques :

- **Gestion des modifications automatique :** automatisez les déploiements/correctifs.

REL 7. 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 :

- **Sauvegarde des données :** sauvegardez les données importantes avec Amazon S3, les instantanés Amazon EBS ou les logiciels tiers pour satisfaire aux objectifs du point de récupération.
- **Sauvegardes automatiques :** utilisez les fonctions AWS, les solutions AWS Marketplace ou les logiciels tiers pour automatiser les sauvegardes.
- **Sécurisation et/ou chiffrement des sauvegardes :** consultez le livre blanc sur les bonnes pratiques de sécurité AWS.
- **Test régulier de la récupération :** validez que l'implémentation du processus de sauvegarde satisfait à l'objectif de délai de récupération et aux objectifs du point de récupération via un test de récupération.

REL 8. 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 :

- **Répartition de charge :** utilisez un répartiteur de charge devant un groupe de ressources.
- **Multi-AZ/région :** répartissez les applications entre plusieurs zones de disponibilité/régions.
- **Réparation automatique :** utilisez les fonctions automatiques pour détecter les défaillances et exécuter une action de correction.
- **Supervision :** surveillez en permanence l'état de votre système.
- **Notification :** prévoyez de recevoir des notifications pour tout événement significatif.

REL 9. Comment planifiez-vous la récupération ?

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.
- **Limites de service :** demandez une augmentation des limites de service auprès du site de reprise après sinistre pour accueillir le basculement.
- **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 »

PERF 1. Comment sélectionnez-vous le type d'instance approprié pour votre système ?

Amazon EC2 fournit un vaste éventail de types d'instances optimisés pour différents cas d'utilisation. Les types d'instances se composent de différentes combinaisons d'unité centrale, de mémoire, de stockage et de capacité réseau, et vous permettent de bénéficier de flexibilité dans le choix de l'association de ressources convenant à vos applications. Chaque type d'instance inclut une ou plusieurs tailles d'instance, ce qui vous permet de dimensionner vos ressources en fonction des exigences de votre charge de travail cible. AWS prend en charge les architectures sans serveur, telles qu'AWS Lambda, qui peuvent modifier radicalement l'efficacité des performances d'une charge de travail.

Bonnes pratiques :

- **Stratégie/architecture de référence :** sélectionnez le type et la taille d'instance en fonction des besoins en ressources prévisibles selon les normes internes de gouvernance.
- **Coût/budget :** sélectionnez le type et la taille d'instance en fonction des besoins en ressources prévisibles selon les contrôles internes de coût.
- **Tests :** effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection (comparaison entre le test d'une performance connue et celui d'une charge de travail connue).
- **Instructions d'AWS ou d'un membre d'AWS Partner Network (APN) :** effectuez vos sélections selon les conseils relatifs aux bonnes pratiques.
- **Test de charge :** déployez la version la plus récente de votre système sur AWS à l'aide de différents types et tailles d'instance, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 2. Comment savez-vous que vous continuez à avoir le type d'instance le plus approprié tandis que de nouveaux types d'instance et de nouvelles fonctionnalités sont introduites ?

AWS écoute les commentaires des clients et continue à innover avec les nouveaux types et tailles d'instance, en fournissant de nouvelles combinaisons d'unité centrale, de mémoire, de stockage et de capacité réseau. Cela signifie qu'un nouveau type d'instance peut être proposé qui offre une meilleure efficacité des performances que celle que vous aviez sélectionnée à l'origine.

Bonnes pratiques :

- **Révision :** resélectionnez de façon cyclique les nouveaux types et tailles d'instance en fonction des besoins en ressources prévisibles.
- **Tests :** après la publication de chaque nouveau type d'instance, effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection.

- **Test de charge :** après la publication de chaque nouveau type d'instance, déployez la version la plus récente de votre système sur AWS, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 3. Comment surveillez-vous les instances 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 Amazon CloudWatch :** utilisez CloudWatch pour surveiller les instances.
- **Supervision tierce :** utilisez les outils tiers pour surveiller les systèmes.
- **Vérification régulière :** consultez régulièrement vos tableaux de bord de supervision.
- **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.

PERF 4. Comment garantissez-vous que la quantité de vos instances correspond à la demande ?

La quantité de la demande placée sur un système varie souvent sur différents cycles : cycle de vie du produit, comme le lancement ou la croissance ; cycles temporels, comme une heure de la journée, un jour de la semaine ou un mois ; les cycles non prévisibles, comme la visibilité des réseaux sociaux ; et les cycles prévisibles, comme les épisodes d'une série télévisée. Des instances insuffisantes pour satisfaire votre charge de travail peuvent dégrader l'expérience utilisateur et, au pire, conduire à une défaillance du système.

Bonnes pratiques :

- **Planification :** planifiez selon des métriques et/ou des événements prévus.
- **Automatisation - Scripts :** utilisez les outils de gestion automatique.
- **Automatisation - Auto Scaling :** utilisez Auto Scaling pour la gestion automatique.

PERF 5. Comment sélectionnez-vous la solution de stockage appropriée pour votre système ?

AWS est conçu pour fournir un stockage de données à bas coût, avec une durabilité et une disponibilité élevées. AWS offre un vaste choix d'options de stockage pour la sauvegarde, l'archivage et la reprise après sinistre, ainsi que le stockage de blocs, de fichiers et d'objets.

Bonnes pratiques :

- **Stratégie/architecture de référence :** sélectionnez les fonctionnalités et les solutions de stockage en fonction des besoins en ressources prévisibles selon les normes internes de gouvernance.
- **Coût/budget :** sélectionnez les fonctionnalités et les solutions de stockage en fonction des besoins en ressources prévisibles selon les contrôles internes de coût.
- **Tests :** effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection (comparaison entre le test d'une performance connue et celui d'une charge de travail connue).
- **Instructions d'AWS ou d'un partenaire APN :** sélectionnez une solution en fonction des conseils relatifs aux bonnes pratiques.
- **Test de charge :** déployez la version la plus récente de votre système sur AWS à l'aide de différentes solutions de stockage, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 6. Comment savez-vous que vous continuez à avoir la solution de stockage la plus appropriée tandis que de nouvelles fonctionnalités et solutions de stockage sont lancées ?

AWS écoute les commentaires des clients et continue à innover avec les nouvelles fonctionnalités et solutions de stockage, en fournissant de nouvelles combinaisons de capacité, de débit et de durabilité. Cela signifie qu'une nouvelle solution de stockage peut être proposée qui offre une meilleure efficacité des performances que celle que vous aviez sélectionnée à l'origine.

Bonnes pratiques :

- **Révision :** resélectionnez de façon cyclique les nouvelles fonctionnalités et solutions de stockage en fonction des besoins en ressources prévisibles.
- **Tests :** après la publication de chaque nouvelle fonctionnalité ou solution de stockage, effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection.
- **Test de charge :** après la publication de chaque nouvelle solution de stockage, déployez la version la plus récente de votre système sur AWS, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 7. Comment surveillez-vous votre solution de stockage pour vous assurer qu'elle se comporte comme prévu ?

Les performances du système peuvent se dégrader au fil du temps, ou sur certaines périodes, en raison de facteurs internes ou externes. La supervision des performances des systèmes vous permet d'identifier cette dégradation et de corriger les facteurs internes ou externes.

Bonnes pratiques :

- **Supervision Amazon CloudWatch :** utilisez CloudWatch pour surveiller les systèmes de stockage.
- **Supervision tierce :** utilisez les outils tiers pour surveiller les systèmes de stockage.
- **Vérification régulière :** consultez régulièrement vos tableaux de bord de supervision.
- **Vérification basée sur les alarmes :** planifiez vos systèmes de supervision pour qu'ils vous alertent automatiquement si les métriques excèdent les limites sécurisées.

- **Actions basées sur les déclencheurs :** planifiez les alarmes pour qu'elles déclenchent des actions automatiques pour corriger un problème ou le réaffecter.

PERF 8. Comment garantissez-vous que la capacité et le débit de vos solutions de stockage correspondent à la demande ?

La quantité de la demande placée sur un système varie souvent sur différents cycles : cycle de vie du produit, comme le lancement ou la croissance ; cycles temporels, comme une heure de la journée, un jour de la semaine ou un mois ; les cycles non prévisibles, comme la visibilité des réseaux sociaux ; et les cycles prévisibles, comme les épisodes d'une série télévisée. Une capacité de stockage et un débit insuffisants pour satisfaire votre charge de travail peuvent dégrader l'expérience utilisateur et, au pire, conduire à une défaillance du système.

Bonnes pratiques :

- **Réaction :** gérez manuellement en fonction des métriques.
- **Planification :** planifiez le débit et la capacité futurs selon des métriques et/ou des événements prévus.
- **Automatisation :** automatisez par rapport aux métriques.

PERF 9. Comment sélectionnez-vous la solution de base de données appropriée pour votre système ?

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

Bonnes pratiques :

- **Stratégie/architecture de référence :** sélectionnez les fonctionnalités et les solutions de base de données en fonction des besoins en ressources prévisibles selon les normes internes de gouvernance.

- **Coût/budget :** sélectionnez les fonctionnalités et les solutions de base de données en fonction des besoins en ressources prévisibles selon les contrôles internes de coût.
- **Tests :** effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection (comparaison entre le test d'une performance connue et celui d'une charge de travail connue).
- **Instructions d'AWS ou d'un partenaire APN :** sélectionnez une solution en fonction des conseils relatifs aux bonnes pratiques.
- **Test de charge :** déployez la version la plus récente de votre système sur AWS à l'aide de différentes solutions de base de données, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 10. Comment savez-vous que vous continuez à avoir la solution de base de données et les fonctionnalités les plus appropriées tandis que de nouvelles fonctionnalités et solutions de base de données sont lancées ?

AWS écoute les commentaires des clients et continue à innover avec de nouvelles fonctionnalités et solutions de base de données, en fournissant de nouvelles combinaisons de cohérence, de disponibilité, de tolérance des partitions et de latence. Cela signifie qu'une nouvelle fonctionnalité ou solution de base de données peut être proposée qui offre une meilleure efficacité des performances que celle que vous aviez sélectionnée à l'origine.

Bonnes pratiques :

- **Révision :** resélectionnez de façon cyclique les nouvelles fonctionnalités et solutions de base de données en fonction des besoins en ressources prévisibles.
- **Tests :** après la publication de chaque nouvelle fonctionnalité ou solution de base de données, effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection.
- **Test de charge :** après la publication de chaque nouvelle fonctionnalité ou solution de base de données, déployez la version la plus récente de votre système sur AWS, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 11. Comment surveillez-vous vos bases de données pour vous assurer que les performances sont celles attendues ?

Les performances du système peuvent se dégrader au fil du temps en raison de facteurs internes ou externes. La supervision des performances des systèmes vous permet d'identifier cette dégradation et de corriger les facteurs internes ou externes.

Bonnes pratiques :

- **Supervision Amazon CloudWatch :** utilisez CloudWatch pour surveiller les bases de données.
- **Supervision tierce :** utilisez les outils tiers pour surveiller les bases de données.
- **Vérification régulière :** consultez régulièrement vos tableaux de bord de supervision.
- **Notifications basées sur les alarmes :** planifiez vos systèmes de supervision pour qu'ils vous alertent automatiquement si les métriques excèdent les limites sécurisées.
- **Actions basées sur les déclencheurs :** planifiez les alarmes pour qu'elles déclenchent des actions automatiques pour corriger un problème ou le réaffecter.

PERF 12. Comment garantissez-vous que la capacité et le débit de vos bases de données correspondent à la demande ?

La quantité de la demande placée sur un système varie souvent sur différents cycles : cycle de vie du produit, comme le lancement, la croissance, etc. ; cycles temporels, comme une heure de la journée, un jour de la semaine ou un mois, etc. ; les cycles non prévisibles, comme la visibilité des réseaux sociaux ; et les cycles prévisibles, comme les épisodes d'une série télévisée. Une capacité de base de données et un débit insuffisants pour satisfaire votre charge de travail peuvent dégrader l'expérience utilisateur et, au pire, conduire à une défaillance du système.

Bonnes pratiques :

- **Planification :** planifiez le débit et la capacité futurs selon des métriques et/ou des événements prévus.
- **Automatisation :** automatisez par rapport aux métriques.

PERF 13. Comment sélectionnez-vous les solutions appropriées de proximité et de mise en cache pour votre système ?

L'éloignement physique, la distance réseau ou les demandes de longue durée peuvent entraîner des retards système. La latence non traitée peut occuper les ressources système plus longtemps que requis et entraîner une dégradation des performances aussi bien internes qu'externes. Pour réduire la latence, considérez les performances de bout en bout de la totalité de votre système du point de vue de l'utilisateur final, et recherchez les opportunités d'ajuster la proximité physique de vos ressources ou solutions de cache.

Bonnes pratiques :

- **Stratégie/architecture de référence :** sélectionnez les solutions de proximité et de mise en cache en fonction des besoins en ressources prévisibles selon les normes internes de gouvernance.
- **Coût/budget :** sélectionnez les solutions de proximité et de mise en cache des besoins en ressources prévisibles selon les contrôles internes de coût.
- **Tests :** effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection (comparaison entre le test d'une performance connue et celui d'une charge de travail connue).
- **Instructions d'AWS ou d'un partenaire APN :** sélectionnez une solution de proximité et de mise en cache en fonction des conseils relatifs aux bonnes pratiques.
- **Test de charge :** déployez la version la plus récente de votre système sur AWS à l'aide de différentes solutions de proximité et de mise en cache, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.

PERF 14. Comment savez-vous que vous continuez à avoir les solutions les plus appropriées de proximité et de mise en cache tandis que de nouvelles solutions sont lancées ?

AWS écoute les commentaires des clients et continue à innover avec de nouvelles fonctionnalités et solutions de proximité et de mise en cache, en fournissant de nouvelles combinaisons de proximité, de mise en cache et de latence. Cela signifie que de nouvelles solutions de proximité et de mise en cache peuvent être proposées, qui offrent une meilleure efficacité des performances que celle que vous aviez sélectionnée à l'origine. Recherchez les opportunités de réduire la latence et d'augmenter les performances d'un bout à l'autre du système. Par exemple, avez-vous procédé à une optimisation ponctuelle ou continuez-vous à optimiser votre système tandis que la demande évolue au fil du temps ?

Bonnes pratiques :

- **Révision :** resélectionnez de façon cyclique les solutions de proximité et de mise en cache en fonction des besoins en ressources prévisibles.
- **Tests :** après la publication de chaque nouvelle solution de proximité et de mise en cache, effectuez un test de charge d'une charge de travail connue sur AWS et utilisez-le pour déterminer la meilleure sélection.
- **Test de charge :** après la publication de chaque nouvelle solution de proximité et de mise en cache, déployez la version la plus récente de votre système sur AWS, utilisez la supervision pour capturer les métriques de performance, puis effectuez une sélection basée sur un calcul de performance/coût.
- **Supervision proactive – Surveillance Amazon Cloud Watch :** utilisez Amazon CloudWatch pour surveiller les solutions de proximité et de mise en cache.
- **Supervision proactive – surveillance tierce :** utilisez les outils tiers pour surveiller les solutions de proximité et de mise en cache.
- **Notification basée sur les alarmes :** planifiez vos systèmes de supervision pour qu'ils vous alertent automatiquement si les métriques excèdent les limites sécurisées.
- **Actions basées sur les déclencheurs :** planifiez les alarmes pour qu'elles déclenchent des actions automatiques pour corriger un problème ou le réaffecter.

PERF 15. Comment surveillez-vous vos solutions de proximité et de mise en cache pour vous assurer que les performances sont celles attendues ?

Les performances du système peuvent se dégrader au fil du temps en raison de facteurs internes ou externes. La supervision des performances des systèmes vous permet d'identifier cette dégradation et de corriger les facteurs internes ou externes.

Bonnes pratiques :

- **Supervision Amazon CloudWatch :** utilisez CloudWatch pour surveiller les instances.
- **Supervision tierce :** utilisez les outils tiers pour surveiller les systèmes.
- **Vérification régulière :** consultez régulièrement vos tableaux de bord de supervision.
- **Notifications basées sur les alarmes :** planifiez vos systèmes de supervision pour qu'ils vous alertent automatiquement si les métriques excèdent les limites sécurisées.
- **Actions basées sur les déclencheurs :** planifiez les alarmes pour qu'elles déclenchent des actions automatiques pour corriger un problème ou le réaffecter.

PERF 16. Comment garantissez-vous que vos solutions de proximité et de mise en cache correspondent à la demande ?

La quantité de la demande placée sur un système varie souvent sur différents cycles : cycle de vie du produit, comme le lancement, la croissance, etc ; cycles temporels, comme une heure de la journée, un jour de la semaine ou un mois, etc. ; les cycles non prévisibles, comme la visibilité des réseaux sociaux ; et les cycles prévisibles, comme les épisodes d'une série télévisée. Des solutions de proximité et de mise en cache inappropriées pour satisfaire votre charge de travail peuvent dégrader l'expérience utilisateur et, au pire, conduire à une défaillance du système. Cela est particulièrement vrai si vous avez, ou prévoyez d'avoir, une base mondiale d'utilisateurs.

Bonnes pratiques :

- **Planification :** planifiez solutions de proximité et de mise en cache selon des métriques et/ou des événements prévus.
- **Surveillance :** surveillez l'utilisation du cache et la demande au fil du temps.
- **Vérification régulière :** vérifiez l'utilisation du cache et la demande au fil du temps.

Pilier « Optimisation des coûts »

COST 1. 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 surutilisation) 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 les files d'attente :** exécutez votre propre file d'attente Amazon Simple Queue Service (SQS) et faites tourner ou arrêtez-les instances en fonction de la demande.
- **Approche basée sur le temps :** par exemple, suivre le soleil, désactiver les instances de développement/test le weekend, suivre des planifications trimestrielles ou annuelles (par exemple, le Black Friday).
- **Provisionnement approprié :** provisionnez de façon appropriée le débit, le dimensionnement et le stockage de services tels qu'Amazon DynamoDB, Amazon EBS (IOPS provisionnées), Amazon RDS, Amazon EMR, etc.

COST 2. Comment optimisez-vous votre utilisation des services AWS ?

Si vous utilisez les services de niveau application, veillez à bien les utiliser. Par exemple, introduisez les stratégies de cycle de vie pour contrôler l'utilisation d'Amazon S3 ou tirez parti de services tels qu'Amazon RDS et Amazon DynamoDB pour bénéficier d'une extraordinaire flexibilité. Les contrôles d'utilisation appropriée incluent la vérification des déploiements multi-AZ pour Amazon RDS ou la vérification que les IOPS provisionnées sont applicables dans vos tables Amazon DynamoDB.

Bonnes pratiques :

- **Optimisations propres au service :** les exemples incluent la réduction des E/S pour Amazon EBS, le non-chargement d'un trop grand nombre de petits fichiers dans Amazon S3 ; l'utilisation extensive d'instances ponctuelles pour Amazon EMR ; etc.

COST 3. Avez-vous sélectionné les ressources appropriées pour répondre à vos cibles de coût ?

Assurez-vous que les instances Amazon EC2 que vous sélectionnez sont adaptées à la tâche en cours. AWS encourage l'utilisation d'évaluations comparatives afin que vous vous assuriez que le type d'instance que vous choisissez est optimisé pour sa charge de travail.

Bonnes pratiques :

- **Mise en correspondance du profil d'instance en fonction du besoin :** par exemple, mise en correspondance basée sur la charge de travail et la description de l'instance (calcul, mémoire ou stockage intensif).
- **Produits tiers** Par exemple, utilisez les produits tiers tels que CopperEgg ou New Relic pour déterminer les types d'instance appropriés.
- **Amazon CloudWatch :** utilisez CloudWatch pour déterminer la charge du processeur.
- **Métriques personnalisées :** chargez en mémoire les scripts personnalisés et examinez l'utilisation de la mémoire avec CloudWatch.
- **Applications profilées :** profilez vos applications de façon à savoir quand utiliser quel type d'Amazon EBS (magnétique, à visée générale (SSD), IOPS provisionnées). N'utilisez les instances optimisées pour EBS qu'en cas de nécessité.

COST 4. Avez-vous sélectionné le modèle de tarification approprié pour répondre à 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 :

- **Instances ponctuelles :** utilisez les instances ponctuelles pour sélectionner les charges de travail.
- **Analyse de l'utilisation :** analysez régulièrement l'utilisation et achetez les instances réservées en conséquence.
- **Vente des instances réservées :** au fur et à mesure que vos besoins évoluent, vendez les instances réservées dont vous n'avez plus besoin sur le Marketplace des instances réservées, et achetez-en d'autres.
- **Action automatique :** votre architecture doit vous permettre de désactiver les instances non utilisées (par exemple, choisissez Auto Scaling pour diminuer les instances en dehors des heures de travail).
- **Considérations de coût :** prenez les coûts en compte dans la sélection de la région.

COST 5. Existe-t-il des services gérés (services de niveau plus élevés qu'Amazon EC2, Amazon EBS et Amazon S3) que vous utilisez pour améliorer votre retour sur investissement ?

Amazon EC2, Amazon EBS et Amazon S3 sont tous trois des services AWS de « blocs de construction ». Les services gérés tels qu'Amazon RDS et Amazon DynamoDB sont des services AWS « de niveau supérieur ». En utilisant ces 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 :

- **Analyse des services :** analysez les services de niveau application pour voir ceux que vous pouvez utiliser.
- **Prise en compte des 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.

- **Prise en compte des 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.
- **Prise en compte d'AWS CloudFormation, AWS Elastic Beanstalk ou AWS Opsworks :** utilisez les modèles AWS CloudFormation / AWS Elastic Beanstalk/AWS OpsWorks pour profiter des avantages de la standardisation et du contrôle des coûts.

COST 6. 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.)
- **Suivi du 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.

COST 7. 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 :

- **Balisage de toutes les ressources :** vous pourrez ainsi relier les modifications de facturation et les modifications d'infrastructure et d'utilisation.
- **Vérification des rapports de facturation détaillés :** disposez d'un processus standard pour charger et interpréter les rapports de facturation détaillés.

- **Architecture économique :** disposez d'un plan à la fois pour l'utilisation et pour les dépenses (par unité : par exemple, utilisateur, gigaoctet de données).
- **Supervision :** surveillez régulièrement l'utilisation et les dépenses à l'aide d'Amazon CloudWatch ou d'un fournisseur tiers (par exemple : Cloudability, CloudCheckr).
- **Notifications :** permettez aux membres clés de notre équipe de savoir si nos dépenses excèdent des limites bien définies.
- **Utiliser l'Explorateur de coûts AWS**
- **Méthode de refacturation orientée finances :** utilisez cette méthode pour allouer les instances et les ressources aux centres de coûts (par exemple, balisage).

COST 8. Mettez-vous hors service les ressources dont vous n'avez plus besoin ou arrêtez-vous celles qui ne sont pas nécessaires temporairement ?

Assurez-vous de ne payer que pour les services que vous utilisez. 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.

Bonnes pratiques :

- Concevez votre système de façon à bien gérer la terminaison d'une instance tandis que vous identifiez et mettez hors service les instances non critiques ou non requises, ou les ressources avec une faible utilisation.
- Ayez un processus en place pour identifier et mettre hors service les ressources orphelines.
- Rapprochez les ressources mises hors service en fonction du système ou du processus.

COST 9. 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 :

- Utiliser un réseau de distribution de contenu
- Concevez une architecture qui permet d'optimiser le transfert de données (conception d'application, accélération WAN, etc.).
- Analysez la situation et utilisez AWS Direct Connect pour économiser de l'argent et améliorer les performances.
- Equilibrez les coûts de transfert des données de votre architecture et vos besoins en fiabilité et haute disponibilité.

COST 10. Comment gérez-vous et/ou envisagez-vous l'adoption de nouveaux services ?

Chez AWS, notre objectif est de vous aider à concevoir une architecture aussi optimale et économique que possible. Les nouveaux services et fonctionnalités peuvent directement réduire vos coûts. Un bon exemple en est Amazon Glacier, qui offre une solution de stockage à bas coût et « à froid » pour les données auxquelles vous accédez de manière occasionnelle, mais qui doivent être conservées pour des raisons professionnelles ou juridiques. Autre exemple, RRS (Reduced Redundancy Storage) pour Amazon S3, qui vous permet de choisir d'avoir un moins grand nombre de copies de vos objets Amazon S3 (niveaux inférieurs de redondance) pour un prix réduit. Certaines implications sont à prendre en compte lors de la prise de telles décisions, comme, par exemple : « Qu'est-ce que cela implique si j'ai moins de copies de mes données ? » ou « Aurai-je besoin d'accéder plus souvent à ces données que je ne l'imagine ? »

Bonnes pratiques :

- Rencontrez régulièrement votre architecte ou consultant de solutions AWS, ou l'équipe de votre compte, et envisagez les nouveaux services ou les nouvelles fonctionnalités que vous pourriez adopter pour économiser de l'argent.