Livre blanc AWS

Hébergement d'Application Web sur le Nuage AWS

Hébergement d'Application Web sur le Nuage AWS: Livre blanc AWS

Table of Contents

Hébergement d'Application Web sur le Nuage AWS

Date de publication : 20 août 2021 (Révisions du document)

Résumé

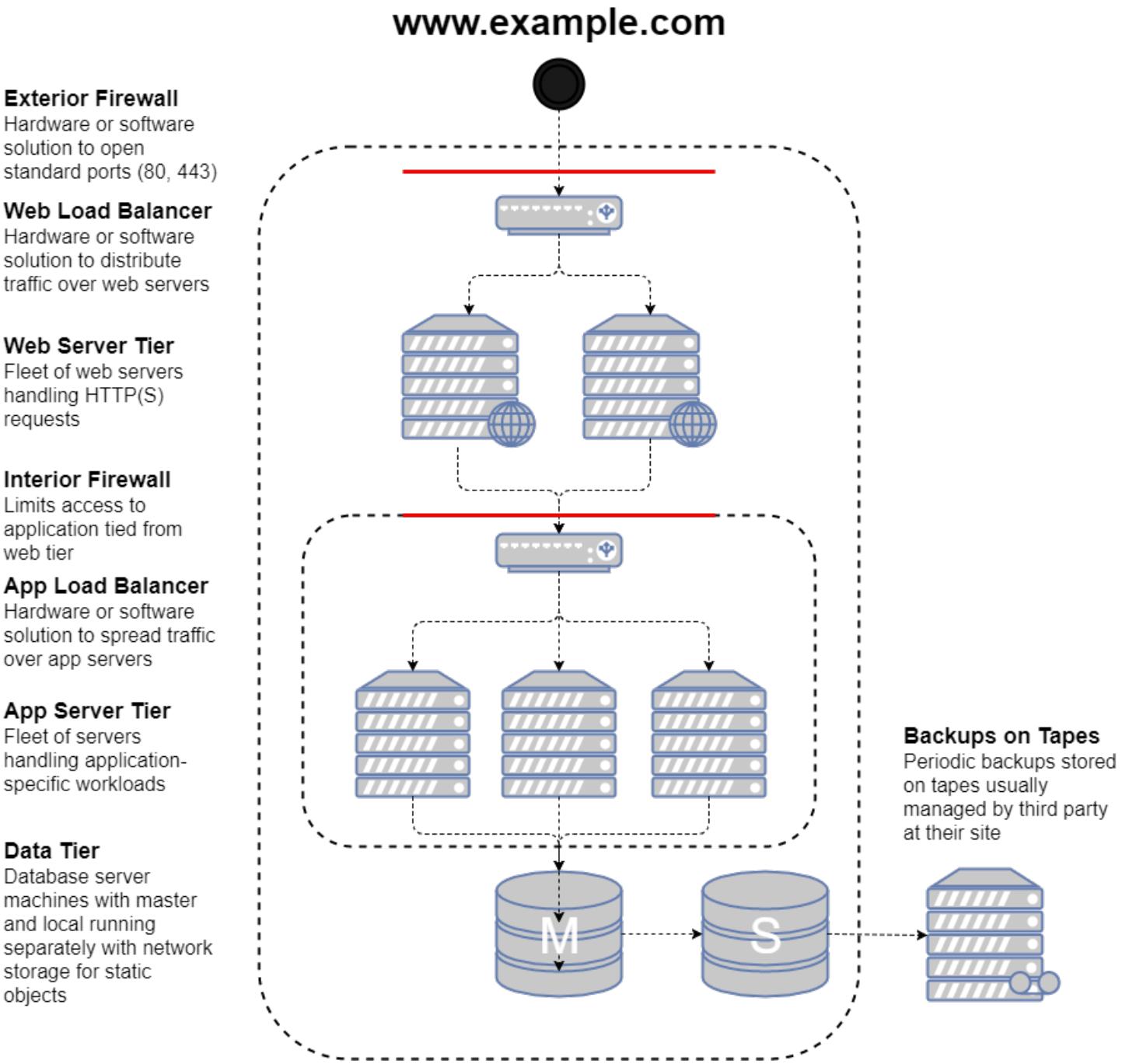
Les architectures web traditionnelles sur site nécessitent des solutions complexes et une prévision précise de la capacité réservée afin de garantir la fiabilité. Les périodes de pointe denses et les fluctuations brutales des modèles de trafic se traduisent par de faibles taux d'utilisation de matériels coûteux. Cela entraîne des coûts d'exploitation élevés de maintien du matériel inactif et une utilisation inefficace du capital pour le matériel sous-utilisé.

Amazon Web Services (AWS) offre une infrastructure fiable, évolutive, sécurisée et hautement performante pour les applications web les plus exigeantes. Cette infrastructure fait correspondre les coûts informatiques aux modèles de trafic des clients en temps quasi réel.

Ce livre blanc est destiné aux responsables informatiques et aux architectes système qui souhaitent comprendre comment exécuter des architectures web traditionnelles dans le cloud pour atteindre élasticité, capacité de mise à l'échelle et fiabilité.

Une présentation de l'hébergement web traditionnel

L'hébergement web évolutif est un problème bien connu. L'image suivante illustre une architecture d'hébergement web traditionnelle qui met en œuvre un modèle d'application web courant à trois niveaux. Dans ce modèle, l'architecture est divisée en couches de présentation, d'application et de permanence. La capacité de mise à l'échelle est assurée par l'ajout d'hôtes à ces couches. L'architecture possède également des fonctions intégrées de performance, de basculement et de disponibilité. L'architecture d'hébergement web traditionnelle est facilement transmise vers le cloud AWS avec seulement quelques modifications.



Une architecture d'hébergement web traditionnelle

Les sections suivantes expliquent pourquoi et comment une telle architecture doit être et pourrait être déployée dans le cloud AWS.

Hébergement d'applications web dans le cloud à l'aide d'AWS

La première question que vous devez vous poser concerne l'intérêt de déplacer une solution d'hébergement d'applications web vers le cloud AWS. Si vous décidez que le cloud vous convient, vous aurez besoin d'une architecture adaptée. Cette section vous aide à évaluer une solution cloud AWS. Elle compare le déploiement de votre application web dans le cloud à un déploiement sur site, présente une architecture cloud AWS pour l'hébergement de votre application et aborde les principaux composants de la solution d'architecture cloud AWS.

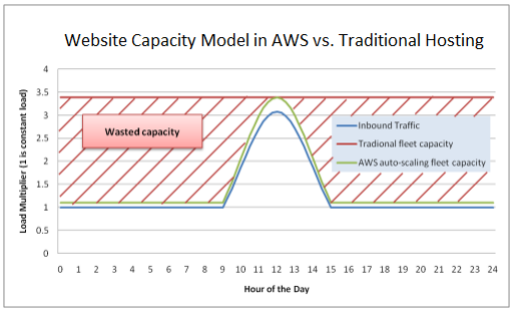
Comment AWS peut résoudre les problèmes courants d'hébergement d'applications web

Si vous êtes responsable de l'exécution d'une application web, vous pourriez être confronté à divers problèmes d'infrastructure et d'architecture pour lesquels AWS peut fournir des solutions transparentes et rentables. Les points suivants sont juste quelques uns des avantages à utiliser AWS plutôt qu'un modèle d'hébergement traditionnel.

Une alternative économique aux flottes surdimensionnées nécessaires aux pics de trafic

Dans le modèle d'hébergement traditionnel, vous devez allouer des serveurs pour gérer les pics de capacité. Les cycles non utilisés sont gaspillés en dehors des périodes de pointe. Les applications web hébergées par AWS peuvent tirer parti de l'approvisionnement à la demande de serveurs supplémentaires, afin que vous puissiez constamment ajuster la capacité et les coûts aux modèles de trafic réels.

Par exemple, le graphique suivant montre une application web avec un pic d'utilisation de 9 h 00 à 15 h 00 et une utilisation moindre pour le reste de la journée. Une approche de scalabilité automatique basée sur les tendances réelles du trafic, qui n'alloue des ressources que lorsque cela est nécessaire, entraînerait moins de gaspillage de capacité et une réduction des coûts de plus de 50 %.



Un exemple de capacité gâchée dans un modèle d'hébergement classique

Une solution évolutive pour gérer les pics inattendus de trafic

Une des conséquences encore plus terrible que la lenteur d'approvisionnement associée à un modèle d'hébergement traditionnel est l'incapacité à répondre à temps aux pics de trafic inattendus. Il existe un certain nombre de témoignages sur des applications web qui deviennent indisponibles en raison d'un pic de trafic inattendu après la mention du site dans des médias populaires. Dans le cloud AWS, la même capacité à la demande qui permet aux applications web de s'adapter aux pics de trafic réguliers peut également gérer une charge inattendue. De nouveaux hôtes peuvent être lancés et sont facilement disponibles en quelques minutes ; ils peuvent être mis hors ligne tout aussi rapidement lorsque le trafic revient à la normale.

Une solution à la demande pour un environnement de test, de charge, bêta, et de reproduction

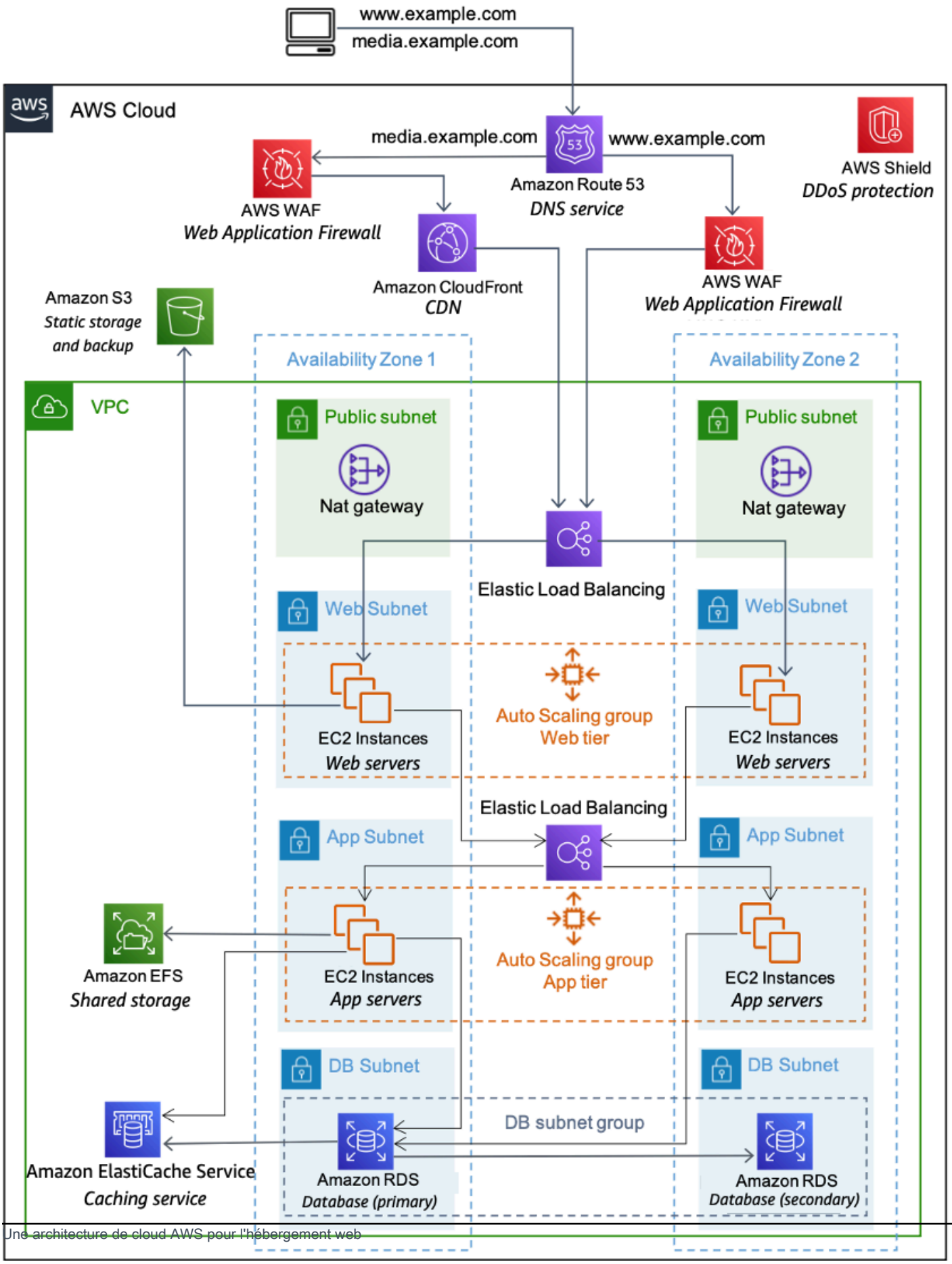
Les coûts du matériel associés à la construction et à la maintenance d'un environnement d'hébergement traditionnel pour la production d'une application web ne s'arrêtent pas à ceux de la flotte de production. Vous devez souvent créer des flottes de préproduction, de version bêta et de test pour garantir la qualité de l'application web à chaque étape du cycle de vie du développement. Bien que vous puissiez effectuer diverses optimisations pour garantir la meilleure utilisation possible de ce matériel de test, ces flottes parallèles ne sont pas toujours utilisées de manière optimale et de nombreux matériels coûteux restent inutilisés pendant de longues périodes.

Dans le cloud AWS, vous pouvez allouer des flottes de test au fur et à mesure que vous en avez besoin. En plus d'éliminer le besoin d'allouer des ressources des jours ou des mois avant leur utilisation réelle, cela vous donne également la flexibilité de démolir des composants de l'infrastructure lorsque vous n'en avez pas besoin. En outre, vous pouvez simuler le trafic utilisateur sur le cloud AWS lors des tests de charge. Vous pouvez également utiliser ces flottes parallèles comme environnement de transit pour une nouvelle version de production. Cela permet de passer

rapidement de la production actuelle à une nouvelle version de l'application avec peu ou pas de pannes de service.

Une architecture de cloud AWS pour l'hébergement web

La figure suivante présente une autre vue d'une architecture classique d'application web et comment elle peut tirer profit de l'infrastructure de calcul du cloud AWS.


www.example.com
media.example.com
AWS Cloud
media.example.com
www.example.com
AWS Shield
DDoS protection
AWS WAF
Web Application Firewall
Amazon Route 53
DNS service
Amazon CloudFront
CDN
AWS WAF
Web Application Firewall
Amazon S3
Static storage and backup
Availability Zone 1
Availability Zone 2
VPC
Public subnet
Nat gateway
Public subnet
Nat gateway
Elastic Load Balancing
Web Subnet
Web Subnet
EC2 Instances
Web servers
Auto Scaling group
Web tier
EC2 Instances
Web servers
Elastic Load Balancing
App Subnet
App Subnet
Amazon EFS
Shared storage
EC2 Instances
App servers
Auto Scaling group
App tier
EC2 Instances
App servers
DB Subnet
DB Subnet
DB subnet group
Amazon ElastiCache Service
Caching service
Amazon RDS
Database (primary)
Amazon RDS
Database (secondary)

Un exemple d'une architecture d'hébergement web sur AWS

1. Services DNS avec Amazon Route 53 : fournit des services DNS pour simplifier la gestion des domaines.
2. Mise en cache périphérique avec Amazon CloudFront : la périphérie met en cache le contenu à volume élevé afin de réduire la latence pour les clients.
3. Sécurité périphérique pour Amazon CloudFront avec AWS WAF : filtre le trafic malveillant, y compris les scripts intersites (XSS) et l'injection SQL via des règles définies par le client.
4. Répartition de charge avec Elastic Load Balancing (ELB) : vous permet de répartir la charge sur plusieurs zones de disponibilité et groupes AWS Auto Scaling pour la redondance et le découplage des services.
5. Protection DDoS avec AWS Shield : protège automatiquement votre infrastructure contre les attaques DDoS les plus courantes au niveau du réseau et de la couche de transport.
6. Pare-feu avec groupes de sécurité : déplace la sécurité vers l'instance pour un pare-feu dynamique avec état au niveau de l'hôte pour les serveurs web et les serveurs d'applications.
7. Mise en cache avec Amazon ElastiCache : fournit des services de mise en cache avec Redis ou Memcached pour supprimer la charge de l'application et de la base de données, et réduire la latence pour les demandes fréquentes.
8. Base de données gérée avec Amazon Relational Database Service (Amazon RDS) : crée une architecture de base de données multi-AZ hautement disponible avec six moteurs de base de données possibles.
9. Stockage statique et sauvegardes avec Amazon Simple Storage Service (Amazon S3) : permet un stockage d'objets simple basé sur HTTP pour les sauvegardes et les ressources statiques telles que les images et les vidéos.

Composants clés d'une architecture d'hébergement web AWS

Les sections suivantes présentent certains des composants clés d'une architecture d'hébergement web déployée dans le cloud AWS et expliquent en quoi ils diffèrent d'une architecture d'hébergement web traditionnelle.

Gestion du réseau

Dans le cloud AWS, la possibilité de segmenter votre réseau par rapport à celui des autres clients permet une architecture plus sécurisée et évolutive. Alors que les groupes de sécurité fournissent

une sécurité au niveau de l'hôte (voir la section Sécurité de l'hôte), Amazon Virtual Private Cloud (Amazon VPC) vous permet de lancer des ressources dans un réseau virtuel et isolé logiquement que vous définissez.

Amazon VPC est un service qui vous donne un contrôle total sur les détails de votre configuration de réseaux dans AWS. Parmi les exemples de ce contrôle, citons la création de sous-réseaux publics pour les serveurs web et de sous-réseaux privés sans accès à Internet pour vos bases de données. En outre, Amazon VPC vous permet de créer des architectures hybrides en utilisant des réseaux VPN matériels et d'utiliser le cloud AWS comme extension de votre propre centre de données.

Amazon VPC prend également en charge IPv6, en plus de la prise en charge IPv4 traditionnelle pour votre réseau.

Diffusion de contenu

Lorsque votre trafic web est géographiquement dispersé, il n'est pas toujours possible et certainement pas économique de répliquer la totalité de votre infrastructure à travers le monde. Un réseau de diffusion de contenu (CDN) vous offre la possibilité d'utiliser son réseau mondial d'emplacements périphériques pour déployer une copie en cache du contenu web, comme des vidéos, des pages web, des images et autres, vers vos clients. Pour réduire le temps de réponse, le CDN utilise l'emplacement périphérique le plus proche du client ou de l'emplacement de la demande d'origine. Le débit augmente considérablement du fait que les ressources web sont déployées à partir du cache. Pour les données dynamiques, un grand nombre de CDN peuvent être configurés pour récupérer les données depuis les serveurs d'origine.

Vous pouvez utiliser CloudFront pour diffuser votre site web y compris les contenus dynamiques, statiques et diffusés en continu, à partir d'un réseau mondial d'emplacements périphériques. CloudFront achemine automatiquement les requêtes pour votre contenu vers l'emplacement périphérique le plus proche, de sorte que le contenu puisse être diffusé de manière optimale. CloudFront est optimisé pour fonctionner avec d'autres services AWS, tels qu'Amazon S3 et Amazon Elastic Compute Cloud (Amazon EC2). CloudFront fonctionne aussi de manière transparente avec n'importe quel serveur d'origine autre qu'AWS qui stocke les versions définitives et originales de vos fichiers.

Comme pour les autres services AWS, aucun contrat ni abonnement mensuel n'est requis pour utiliser CloudFront : vous ne payez que la quantité de contenu que vous diffusez par le biais de ce service.

En outre, toutes les solutions existantes de mise en cache périphérique dans votre infrastructure d'application web devraient fonctionner correctement dans le cloud AWS.

Gestion du DNS public

Le déplacement d'une application web vers le cloud AWS nécessite certaines modifications du système de noms de domaine (DNS). Pour vous aider à gérer le routage DNS, AWS fournit Amazon Route 53, un service web DNS cloud hautement disponible et évolutif. Route 53 est conçu pour donner aux développeurs et aux entreprises un moyen extrêmement fiable et rentable d'acheminer les utilisateurs finaux vers des applications Internet en remplaçant des noms, comme www.exemple.com, par des adresses IP telles que 192.0.2.1, que les ordinateurs utilisent pour se connecter l'un à l'autre. De plus, Route 53 est entièrement conforme au protocole IPv6.

Sécurité de l'hôte

Outre le filtrage du trafic réseau entrant en périphérie, AWS recommande également aux applications web d'appliquer le filtrage du trafic réseau au niveau de l'hôte. Amazon EC2 fournit une fonction nommée groupes de sécurité. Un groupe de sécurité est analogue à un pare-feu réseau entrant, pour lequel vous pouvez spécifier les protocoles, les ports et les plages d'adresses IP sources qui sont autorisés à atteindre vos instances EC2.

Vous pouvez attribuer un ou plusieurs groupes de sécurité à chaque instance EC2. Chaque groupe de sécurité autorise le trafic approprié vers chaque instance. Les groupes de sécurité peuvent être configurés de telle sorte que seuls des sous-réseaux, des adresses IP et des ressources spécifiques aient accès à une instance EC2. Ils peuvent également référencer d'autres groupes de sécurité pour limiter l'accès aux instances EC2 appartenant à des groupes spécifiques.

Dans l'architecture d'hébergement web AWS de la figure 3, le groupe de sécurité du cluster de serveurs web peut autoriser l'accès uniquement à partir de l'équilibreur de charge de la couche web et uniquement via TCP sur les ports 80 et 443 (HTTP et HTTPS). Le groupe de sécurité du serveur d'applications, en revanche, peut autoriser l'accès uniquement à partir de l'équilibreur de charge de la couche application. Dans ce modèle, vos ingénieurs de support doivent également accéder aux instances EC2, ce qui peut être réalisé avec AWS Systems Manager Session Manager. Pour une discussion plus approfondie sur la sécurité, veuillez consulter Sécurité dans le cloud AWS, qui contient des bulletins de sécurité, des informations de certification et des livres blancs sur la sécurité qui expliquent les fonctionnalités de sécurité d'AWS.

Répartition de charge entre les clusters

Les équilibreurs de charge matériels sont une appliance réseau courante utilisée dans les architectures traditionnelles d'application web. AWS fournit cette fonctionnalité par le biais du service Elastic Load Balancing (ELB). ELB distribue automatiquement le trafic d'application entrant sur plusieurs cibles, telles que les instances EC2, les conteneurs, les adresses IP, les fonctions AWS Lambda et les appliances virtuelles. Il peut gérer la charge variable du trafic de votre application dans une seule zone de disponibilité ou à travers plusieurs zones de disponibilité. Elastic Load Balancing fournit quatre types d'équilibreurs de charge offrant tous la haute disponibilité, la scalabilité automatique et la sécurité robuste nécessaires pour rendre vos applications tolérantes aux pannes.

Trouver d'autres hôtes et services

Dans l'architecture d'hébergement web traditionnelle, la plupart de vos hôtes ont des adresses IP statiques. Dans le cloud AWS, la plupart de vos hôtes ont des adresses IP dynamiques. Bien que chaque instance EC2 puisse avoir des entrées DNS publiques et privées et soit adressable sur Internet, les entrées DNS et les adresses IP sont attribuées dynamiquement lorsque vous lancez l'instance. Elles ne peuvent pas être assignées manuellement. Les adresses IP statiques (adresses IP élastiques dans la terminologie AWS) peuvent être attribuées à des instances en cours d'exécution après leur lancement. Les adresses IP élastiques devraient être utilisées pour les instances et les services qui nécessitent des points de terminaison constants, telles que les bases de données primaires, les serveurs de fichiers centraux ou les équilibreurs de charge hébergés sur EC2.

Mise en mémoire cache dans l'application web

Les caches d'applications en mémoire peuvent réduire la charge sur les services et améliorer les performances et la capacité de mise à l'échelle au niveau de la base de données en mettant en cache les informations fréquemment utilisées. Amazon ElastiCache est un service web qui facilite le déploiement, l'utilisation et la mise à l'échelle d'un cache en mémoire dans le cloud. Vous pouvez configurer le cache en mémoire que vous créez pour qu'il s'adapte automatiquement à la charge et remplace automatiquement les nœuds défaillants. ElastiCache est conforme au protocole avec Memcached et Redis, ce qui simplifie la migration à partir de votre solution sur site actuelle.

Configuration, sauvegarde et basculement de la base de données

De nombreuses applications web contiennent une certaine forme de permanence, généralement sous la forme d'une base de données relationnelle ou non relationnelle. AWS propose des services de base de données relationnelle et non relationnelle. Vous pouvez également déployer votre propre

logiciel de base de données sur une instance EC2. Le tableau suivant résume ces options, qui sont abordées plus en détail dans cette section.

Tableau 1 - Solutions de bases de données relationnelles et non relationnelles

	Solutions de bases de données relationnelles	Solutions NoSQL
Service de base de données gérée	Amazon RDS for MySQL, Oracle, SQL Server, MariaDB, PostgreSQL, Amazon Aurora	Amazon DynamoDB , Amazon Keyspaces, Amazon Neptune, Amazon QLDB, Amazon Timestream
Auto-gérée	Hébergement d'un système de gestion d'une base de données relationnelle (SGBD) sur une instance Amazon EC2	Hébergement d'une solution de base de données non relationnelle sur une instance EC2

Amazon RDS

Amazon Relational Database Service (Amazon RDS) vous donne accès aux fonctionnalités d'un moteur de base de données MySQL, PostgreSQL, Oracle et Microsoft SQL Server familier. Le code, les applications et les outils que vous utilisez déjà peuvent être utilisés avec Amazon RDS. Amazon RDS corrige automatiquement le logiciel de base de données et sauvegarde votre base de données, en stockant les sauvegardes pendant une période de rétention définie par l'utilisateur. Il prend également en charge la restauration à un instant dans le passé. Vous profitez d'une grande flexibilité vous permettant de mettre à l'échelle les ressources de calcul ou les capacités de stockage associées à votre instance de base de données relationnelle via un simple appel d'API.

Les déploiements multi-AZ Amazon RDS augmentent la disponibilité de votre base de données et protègent votre base de données contre les pannes imprévues. Les réplicas en lecture Amazon RDS fournissent des réplicas en lecture seule de votre base de données, afin que vous puissiez augmenter la capacité de déploiement d'une base de données unique pour les charges de travail de base de données à lecture intensive. Comme pour tous les services AWS, aucun investissement initial n'est requis et vous ne payez que pour les ressources que vous utilisez.

Hébergement d'un système de gestion de base de données relationnelle (SGBDR) sur une instance Amazon EC2

Outre l'offre Amazon RDS gérée, vous pouvez installer le SGBDR de votre choix (tel que MySQL, Oracle, SQL Server ou DB2) sur une instance EC2 et le gérer vous-même. Les clients AWS hébergeant une base de données sur Amazon EC2 utilisent avec succès divers modèles primaires/de sauvegarde et de réplication, notamment la mise en miroir pour les copies en lecture seule et l'envoi de journaux pour les esclaves passifs toujours prêts.

Lorsque vous gérez votre propre logiciel de base de données directement sur Amazon EC2, vous devez également prendre en compte la disponibilité d'un stockage permanent et tolérant aux pannes. À cette fin, nous recommandons que les bases de données exécutées sur Amazon EC2 utilisent des volumes Amazon Elastic Block Store (Amazon EBS), qui sont similaires au stockage rattaché au réseau.

Pour les instances EC2 exécutant une base de données, vous devez placer toutes les données et tous les journaux de base de données sur des volumes EBS. Ils restent disponibles même en cas de défaillance de l'hôte de la base de données. Cette configuration permet un scénario de basculement simple, dans lequel une nouvelle instance EC2 peut être lancée en cas de défaillance d'un hôte, et les volumes EBS existants peuvent être attachés à la nouvelle instance. La base de données peut alors reprendre là où elle s'est arrêtée.

Les volumes EBS fournissent automatiquement une redondance au sein de la zone de disponibilité. Si les performances d'un seul volume EBS ne sont pas suffisantes pour les besoins de vos bases de données, les volumes peuvent être agrégés par bandes pour augmenter les performances des opérations d'entrée/sortie par seconde (IOPS) de votre base de données.

Pour les charges de travail exigeantes, vous pouvez également utiliser des IOPS provisionnés EBS, où vous spécifiez les IOPS requis. Si vous utilisez Amazon RDS, le service gère son propre stockage afin que vous puissiez vous concentrer sur la gestion de vos données.

Bases de données non relationnelles

Outre la prise en charge des bases de données relationnelles, AWS propose également un certain nombre de bases de données non relationnelles gérées :

- Amazon DynamoDB est un service de base de données NoSQL entièrement géré, offrant des performances exceptionnelles et prévisibles en termes de capacité de mise à l'échelle transparente. À l'aide de la AWS Management Console ou de l'API DynamoDB, vous pouvez

augmenter ou diminuer la capacité sans temps d'arrêt ni dégradation des performances. Puisque DynamoDB s'occupe des charges administratives liées au fonctionnement et à la mise à l'échelle des bases de données distribuées vers AWS, vous n'avez pas à vous soucier de l'approvisionnement, du paramétrage et de la configuration du matériel, de la réplication, ni des correctifs logiciels ou de la mise à l'échelle des clusters.

- Amazon DocumentDB (compatible avec MongoDB) est un service de base de données qui est spécifiquement conçu pour la gestion des données JSON à grande échelle, entièrement géré et intégré à AWS et représente une solution professionnelle avec une durabilité élevée.
- Amazon Keyspaces (pour Apache Cassandra) est un service de base de données évolutif, hautement disponible et géré, compatible avec Apache Cassandra. Avec Amazon Keyspaces, vous pouvez exécuter vos charges de travail Cassandra sur AWS à l'aide du même code d'application Cassandra et des mêmes outils pour développeur que ceux que vous utilisez aujourd'hui.
- Amazon Neptune est un service de base de données orientée graphe fiable, rapide et entièrement géré qui facilite la création et l'exécution d'applications utilisant des jeux de données hautement connectés. Le cœur d'Amazon Neptune est un moteur de base de données orientée graphe haute performance optimisé pour stocker des milliards de relations et interroger le graphique avec une latence de quelques millisecondes.
- Amazon Quantum Ledger Database (Amazon QLDB) est une base de données de registre entièrement gérée qui fournit un journal de transaction transparent, inaltérable et vérifiable par chiffrement appartenant à une autorité de confiance. QLDB peut être utilisé pour suivre chaque modification de données d'application et maintient un historique complet et vérifiable des modifications au fil du temps.
- Amazon Timestream est un service de base de données de séries temporelles rapide, évolutif et sans serveur pour l'IoT et les applications opérationnelles. Il facilite le stockage et l'analyse de milliers de milliards d'événements par jour, jusqu'à 1 000 fois plus rapidement que les bases de données relationnelles et pour un dixième du coût.

En outre, vous pouvez utiliser Amazon EC2 pour héberger d'autres technologies de base de données non relationnelles avec lesquelles vous êtes susceptible de travailler.

Stockage et sauvegarde de données et de ressources

Le cloud AWS offre de nombreuses options de stockage, d'accès et de sauvegarde des données et des ressources de votre application web. Amazon S3 fournit un magasin d'objets redondant et hautement disponible. Amazon S3 est une excellente solution de stockage pour les objets relativement statiques et peu changeant comme les images, les vidéos et d'autres médias statiques.

Amazon S3 prend également en charge la mise en cache périphérique et le streaming de ces ressources en interagissant avec CloudFront.

Pour le stockage similaire à un système de fichiers rattaché, les instances EC2 peuvent être associées à des volumes EBS. Ils agissent comme des disques montables pour exécuter des instances EC2. Amazon EBS est parfaitement adapté aux données qui doivent être accessibles en stockage par bloc et qui nécessitent une permanence en dehors de la vie de l'instance en cours d'exécution, comme par exemple les partitions d'une base de données et les journaux de l'application.

En plus d'avoir une durée de vie indépendante de l'instance EC2, vous pouvez prendre des instantanés des volumes EBS et les stocker dans Amazon S3. Puisque les instantanés EBS ne sauvegardent que les modifications apportées depuis l'instantané précédent, des instantanés plus fréquents peuvent en réduire la durée. Vous pouvez également utiliser un instantané EBS comme référence pour répliquer des données sur plusieurs volumes EBS et attacher ces volumes à d'autres instances en cours d'exécution.

Les volumes EBS peuvent atteindre 16 To, et plusieurs volumes EBS peuvent être agrégés par bandes pour des volumes encore plus importants ou pour des performances d'entrée/sortie (I/O) accrues. Pour optimiser les performances de vos applications gourmandes en I/O, vous pouvez utiliser des volumes IOPS provisionnés. Les volumes d'IOPS provisionnés sont conçus pour satisfaire les besoins des charges de travail gourmandes en I/O, notamment les charges de travail de base de données sensibles aux performances de stockage et à l'homogénéité du débit d'I/O.

Vous spécifiez un taux d'IOPS lorsque vous créez le volume et Amazon EBS alloue ce taux pour la durée de vie du volume. Amazon EBS prend actuellement en charge l'IOPS par volume allant de 16 000 (pour tous les types d'instances) à 64 000 (pour les instances construites sur Nitro System). Vous pouvez agréger plusieurs volumes afin de pouvoir livrer des milliers d'IOPS par instance à votre application. En dehors de cela, pour un débit plus élevé et des charges de travail stratégiques nécessitant une latence inférieure à la milliseconde, vous pouvez utiliser le type de volume express par bloc io2 qui peut prendre en charge jusqu'à 256 000 IOPS avec une capacité de stockage maximale de 64 To.

Scalabilité automatique de la flotte

L'une des principales différences entre l'architecture cloud AWS et le modèle d'hébergement traditionnel est qu'AWS peut automatiquement mettre à l'échelle la flotte d'applications web à la demande pour gérer les changements de trafic. Dans le modèle d'hébergement traditionnel, les

modèles de prévisions de trafic sont généralement utilisés pour allouer des hôtes à l'avance. Dans AWS, les instances peuvent être allouées à la volée en fonction d'un ensemble de déclencheurs permettant de mettre à l'échelle la flotte qui sort et revient.

Le service Auto Scaling peut créer une capacité de groupes de serveurs pouvant croître ou réduire à la demande. Auto Scaling fonctionne également directement avec CloudWatch pour les données métriques et avec Elastic Load Balancing pour ajouter et supprimer des hôtes pour la distribution de la charge. Par exemple, si les serveurs web indiquent une utilisation du CPU de plus de 80 % sur une certaine période de temps, un serveur web supplémentaire peut alors être rapidement déployé, et être ensuite ajouté automatiquement à l'équilibreur de charge pour une inclusion immédiate dans la rotation de la répartition de charge.

Comme montré dans le modèle d'architecture d'hébergement web AWS, plusieurs groupes Auto Scaling peuvent être créés pour différentes couches de l'architecture afin de permettre à chaque couche une mise à l'échelle indépendante. Par exemple, le groupe Auto Scaling du serveur web peut déclencher une augmentation ou une réduction en réponse à des modifications des I/O réseau, tandis que le groupe Auto Scaling du serveur d'applications peut augmenter et réduire en fonction de l'utilisation du CPU. Vous pouvez définir des minimums et des maximums pour garantir une disponibilité 24 h/24 et 7 j/7, et limiter l'utilisation au sein d'un groupe.

Les déclencheurs Auto Scaling peuvent être configurés pour augmenter et réduire la flotte totale à une couche donnée afin d'adapter l'utilisation des ressources à la demande réelle. Outre le service Auto Scaling, vous pouvez mettre à l'échelle les flottes Amazon EC2 directement via l'API Amazon EC2, qui permet de lancer, de mettre fin et d'inspecter des instances.

Fonctions de sécurité supplémentaires

Le nombre et le degré de sophistication des attaques par déni de service distribué (DDoS) augmentent. Traditionnellement, ces attaques sont difficiles à repousser. Elles finissent souvent par être coûteuses en termes de temps d'atténuation et d'énergie dépensée, ainsi que de coût d'opportunité en cas de perte de visites sur votre site web pendant l'attaque. Un certain nombre de facteurs et de services AWS peuvent vous aider à vous défendre contre de telles attaques. L'un d'eux est la mise à l'échelle du réseau AWS. L'infrastructure AWS est assez importante et vous permet de tirer parti de notre mise à l'échelle pour optimiser votre défense. Plusieurs services, dont Elastic Load Balancing, Amazon CloudFront et Amazon Route 53, permettent une mise à l'échelle efficace de votre application web en réponse à une forte augmentation du trafic.

Les services de protection de l'infrastructure, en particulier, contribuent à votre stratégie de défense :

- AWS Shield est un service de protection DDoS géré qui permet de se protéger contre diverses formes de vecteurs d'attaque DDoS. L'offre standard d'AWS Shield est gratuite et automatiquement active sur l'ensemble de votre compte. Cette offre standard permet de vous défendre contre les attaques les plus courantes sur les couches réseaux et de transport. En plus de ce niveau, l'offre avancée propose des niveaux de protection plus élevés pour votre application web en vous offrant une visibilité quasiment en temps réel d'une attaque en cours, ainsi qu'une intégration à des niveaux plus élevés avec les services mentionnés précédemment. En outre, vous avez accès à l'équipe d'intervention contre les attaques DDoS (DRT) d'AWS pour vous aider à atténuer les attaques sophistiquées à grande échelle contre vos ressources.
- AWS WAF (pare-feu d'applications Web) est conçu pour protéger vos applications web contre les attaques susceptibles de compromettre la disponibilité ou la sécurité, ou de consommer des ressources excessives. AWS WAF fonctionne conformément à CloudFront ou Application Load Balancer, ainsi qu'à vos règles personnalisées, pour vous protéger contre les attaques telles que le scripting intersites, l'injection SQL et les attaques DDoS. Comme pour la plupart des services AWS, AWS WAF est livré avec une API complète qui peut vous aider à automatiser la création et la modification des règles pour votre instance AWS WAF à mesure que vos besoins en matière de sécurité évoluent.
- AWS Firewall Manager est un service de gestion de la sécurité qui vous permet de configurer et de gérer de manière centralisée les règles de pare-feu dans vos comptes et vos applications dans AWS Organizations. Lorsque de nouvelles applications sont créées, AWS Firewall Manager facilite également la mise en conformité des nouvelles applications et ressources en appliquant un ensemble commun de règles de sécurité.

Basculement avec AWS

Un autre avantage clé d'AWS par rapport à l'hébergement web traditionnel réside dans les zones de disponibilité qui vous permettent d'accéder facilement aux emplacements de déploiement redondants. Les zones de disponibilité sont des emplacements distincts conçus pour être isolés des échecs dans d'autres zones de disponibilité. Elles fournissent une connectivité réseau économique à faible latence à d'autres zones de disponibilité de la même région AWS. Comme le montre le diagramme d'architecture d'un hébergement web AWS, AWS vous recommande de déployer des hôtes EC2 sur plusieurs zones de disponibilité afin de rendre votre application web plus tolérante aux pannes.

Il est important de garantir qu'il y ait des allocations pour migrer des points d'accès simples sur des zones de disponibilité en cas d'échec. Par exemple, vous devez configurer une base de données

de sauvegarde dans une deuxième zone de disponibilité afin que la permanence des données reste cohérente et hautement disponible, même en cas de scénario d'échec peu probable. Vous pouvez le faire sur Amazon EC2 ou Amazon RDS d'un simple clic.

Bien que certains changements architecturaux soient souvent nécessaires lors du déplacement d'une application web existante vers le cloud AWS, des améliorations significatives en termes de capacité de mise à l'échelle, de fiabilité et de rentabilité font que l'utilisation du cloud AWS en vaut la peine. La section suivante aborde ces améliorations.

Considérations clé lors de l'utilisation d'AWS pour l'hébergement web

Il existe certaines différences clé entre le cloud AWS et un modèle d'hébergement d'applications web traditionnel. La section précédente a mis en avant de nombreux domaines clé à considérer lors du déploiement d'une application web sur le cloud. Cette section présente certains changements architecturaux clé que vous devez prendre en compte lorsque vous importez une application dans le cloud.

Plus aucune appliance réseau physique

Vous ne pouvez pas déployer d'appliance réseau physique dans AWS. Par exemple, les pare-feu, les routeurs et les équilibreurs de charge de vos applications AWS ne peuvent plus prendre la forme d'appareils physiques et doivent être remplacés par des solutions logicielles. Il existe une grande variété de solutions logicielles d'entreprise, que ce soit pour la répartition de charge ou l'établissement d'une connexion VPN. Il ne s'agit pas d'une limitation de ce qui peut être exécuté sur le cloud AWS, mais d'une modification architecturale de votre application si vous utilisez ces appareils aujourd'hui.

Pare-feu omniprésents

Là où vous n'aviez précédemment qu'une simple zone démilitarisée (DMZ) et des communications ouvertes entre vos hôtes dans un modèle d'hébergement traditionnel, AWS adopte un modèle plus sécurisé où chaque hôte est fermé. L'une des étapes de la planification d'un déploiement AWS est l'analyse du trafic entre les hôtes. Cette analyse guidera les décisions concernant exactement les ports qui doivent être ouverts. Vous pouvez créer des groupes de sécurité pour chaque type d'hôte de votre architecture. Vous pouvez également créer une grande variété de modèles de sécurité simples et hiérarchisés pour permettre un accès minimum entre les hôtes de votre architecture. L'utilisation de listes de contrôle d'accès au réseau au sein d'Amazon VPC peut vous aider à verrouiller votre réseau au niveau du sous-réseau.

Tenir compte de la disponibilité de plusieurs centres de données

Considérez les zones de disponibilité au sein d'une région AWS comme plusieurs centres de données. Les instances EC2 dans différentes zones de disponibilité sont à la fois logiquement et

physiquement séparées et fournissent un modèle facile d'utilisation pour déployer votre application au travers de centres de données pour allier haute disponibilité et grande fiabilité. Amazon VPC en tant que service régional vous permet d'exploiter les zones de disponibilité tout en conservant toutes vos ressources dans le même réseau logique.

Traitement éphémère et dynamique des hôtes

Probablement le changement les plus important dans la façon dont vous allez peut-être concevoir votre architecture d'application AWS, il faut considérer les hôtes Amazon EC2 comme éphémères et dynamiques. Toute application créée pour le cloud AWS ne doit pas présumer qu'un hôte sera toujours disponible et doit être conçue en sachant que toutes les données des magasins instantanés EC2 seront perdues en cas de défaillance d'une instance EC2.

Lorsqu'un nouvel hôte est créé, vous ne devez pas émettre d'hypothèses concernant l'adresse IP ou l'emplacement dans une zone de disponibilité de l'hôte. Votre modèle de configuration doit être flexible et votre approche en matière d'action d'amorçage d'un hôte doit tenir compte de la nature dynamique du cloud. Ces techniques sont essentielles pour créer et exécuter une application hautement évolutive et tolérante aux pannes.

Envisager les conteneurs et un modèle sans serveur

Ce livre blanc se concentre principalement sur une architecture web plus traditionnelle. Cependant, pensez à moderniser vos applications web en passant aux technologies de Conteneurs et Sans serveur, en tirant parti de services tels que AWS Fargate et AWS Lambda afin de vous permettre de faire abstraction des machines virtuelles pour effectuer des tâches de calcul. Avec le calcul sans serveur, les tâches de gestion de l'infrastructure telles que l'approvisionnement en capacité et l'application de correctifs sont gérées par AWS, ce qui vous permet de créer des applications plus agiles afin d'innover et de répondre aux changements plus rapidement.

Envisager un déploiement automatisé

- Amazon Lightsail est un serveur privé virtuel (VPS) facile à utiliser qui vous offre tout ce dont vous avez besoin pour créer une application ou un site web, en plus d'offrir un forfait mensuel rentable. Lightsail est idéal pour les charges de travail plus simples, les déploiements rapides et pour démarrer sur AWS. Il est conçu pour vous aider à débuter modestement, pour ensuite vous mettre à l'échelle au fur et à mesure de votre croissance.

- AWS Elastic Beanstalk est un service simple à utiliser pour déployer et mettre à l'échelle des applications et des services web développés avec Java, .NET, PHP, Node.js, Python, Ruby, Go et Docker sur des serveurs courants, tels qu'Apache, Nginx, Passenger et IIS. Il vous suffit de télécharger votre code pour qu'Elastic Beanstalk effectue automatiquement le déploiement, l'approvisionnement en capacité, la répartition de charge, la scalabilité automatique et la surveillance de l'état de l'application. Ce faisant, vous conservez la maîtrise totale des ressources AWS alimentant votre application et pouvez accéder aux ressources sous-jacentes à tout moment.
- AWS App Runner est un service entièrement géré qui permet aux développeurs de déployer facilement et rapidement des applications web et des API conteneurisées, à grande échelle et sans aucune expérience préalable en matière d'infrastructure. Commencez avec votre code source ou une image de conteneur. App Runner crée et déploie automatiquement l'application web et équilibre la charge du trafic avec chiffrement. App Runner se charge également de l'augmentation ou de la réduction automatique du trafic pour répondre à vos besoins.
- AWS Amplify est un ensemble d'outils et de services qui peuvent être utilisés ensemble ou séparément, pour aider les développeurs front-end web et mobiles à créer des applications complètes et évolutives à technologie AWS. Avec Amplify, vous pouvez configurer les backends d'applications et connecter votre application en quelques minutes, déployer des applications web statiques en quelques clics et gérer facilement le contenu des applications en dehors de la AWS Management Console.

Conclusion et contributeurs

Conclusion

Il existe de nombreuses considérations architecturales et conceptuelles à prendre en compte lorsque vous envisagez de migrer votre application web vers le cloud AWS. Les avantages d'une infrastructure rentable, hautement évolutive et tolérante aux pannes qui évolue avec votre entreprise dépassent de loin les efforts de migration vers le cloud AWS.

Participants

Les personnes et organisations suivantes ont participé à la préparation du présent document :

- Amir Khairalomoum, architecte de solutions principal, AWS
- Dinesh Subramani, architecte de solutions principal, AWS
- Jack Hemion, architecte de solutions principal, AWS
- Jatin Joshi, ingénieur support cloud, AWS
- Jorge Fonseca, architecte de solutions principal, AWS
- Shinduri K S, architecte de solutions, AWS

Autres lectures

- Déploiement d'une application basée sur Django sur Amazon Lightsail
- Déploiement d'un site web Drupal haute disponibilité sur Elastic Beanstalk
- Déploiement d'une application PHP haute disponibilité sur Elastic Beanstalk
- Déploiement d'une application Node.js avec DynamoDB vers Elastic Beanstalk
- Commencer à utiliser les applications web Linux dans le cloud AWS
- Héberger un site web statique
- Hébergement d'un site web statique à l'aide d'Amazon S3
- Tutoriel : Déploiement d'une application ASP.NET Core avec Elastic Beanstalk
- Tutoriel : Comment déployer un exemple d'application .NET avec Elastic Beanstalk

Révisions du document

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

update-history-change	update-history-description	update-history-date
Livre blanc mis à jour	Plusieurs sections et diagrammes mis à jour avec de nouveaux services et fonctions, et des limites de service mises à jour.	20 août 2021
Livre blanc mis à jour	Étiquette d'icône mise à jour pour « Mise en cache avec ElastiCache » dans la figure 3.	29 septembre 2019
Livre blanc mis à jour	Plusieurs sections ont été ajoutées et mises à jour pour les nouveaux services. Diagrammes mis à jour pour plus de clarté et de services. Ajout du VPC en tant que méthode de réseaux standard dans AWS dans « Gestion du réseau ». Ajout d'une section sur la protection et l'atténuation des attaques DDoS dans « Fonctions de sécurité supplémentaires ». Ajout d'une petite section sur les architectures sans serveur pour l'hébergement web.	1er juillet 2017
Livre blanc mis à jour	Plusieurs sections mises à jour pour plus de clarté. Diagrammes mis à jour pour utiliser les icônes AWS. Ajout	1er septembre 2012

	de la section « Gestion du DNS public » pour plus de détails sur Amazon Route 53. Mise à jour de la section « Trouver d'autres hôtes et services » pour plus de clarté. La section « Configuration, sauvegarde et basculement de la base de données » a été mise à jour pour plus de clarté et avec DynamoDB. La section « Stockage et sauvegarde de données et de ressources » a été étendue pour couvrir les volumes IOPS provisionnés EBS.	
Publication initiale	Livre blanc publié.	1er mai 2010

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