**Amazon**

Hystorique :

En 2006, Amazon Web Services (AWS) a commencé à proposer des services d'infrastructure informatique destinés aux entreprises, sous la forme services web, plus communément connus sous le nom de cloud computing. Le cloud computing présente pour principal avantage de pouvoir remplacer les dépenses initiales en infrastructure par des coûts variables réduits, adaptés à votre activité. Grâce au cloud, les entreprises n'ont plus besoin de prévoir et d'acheter des serveurs et d'autres infrastructures informatiques des semaines, voire des mois, à l'avance. Elles peuvent, en revanche, accéder instantanément à des centaines ou des milliers de serveurs en quelques minutes, et générer plus rapidement des résultats. A l'heure actuelle, Amazon Web Services propose une plateforme d'infrastructure extrêmement fiable, évolutive et peu coûteuse dans le cloud, qui gère des centaines de milliers d'entreprises dans 190 pays partout dans le monde. Ce livre blanc présente la plateforme de cloud computing d'AWS. Il détaille les avantages du cloud computing, ainsi que les principes fondamentaux d'AWS. Il présente également les services d'AWS qui englobent la plateforme.

Infrastructure mondiale :

AWS gère plus d'1 million de clients actifs répartis dans plus de 240 pays et territoires. Nous élargissons constamment l'infrastructure mondiale pour aider nos clients à réduire la latence et à augmenter le débit, tout en veillant à ce que leurs données ne résident que dans la région AWS qu'ils ont spécifiée. Nos clients faisant croître leurs entreprises, AWS continue dans le même temps à fournir une infrastructure qui répond à leurs exigences globales.

L'infrastructure du Cloud AWS repose sur des régions AWS et des zones de disponibilité. Une région AWS est un emplacement physique dans le monde qui héberge plusieurs zones de disponibilité. Les zones de disponibilité consistent en un ou plusieurs petits centres de données, qui disposent chacun d'une alimentation, d'une mise en réseau et d'une connectivité redondantes, hébergés dans différents sites. Ces zones de disponibilité vous permettent d'exploiter des applications de production et des bases de données plus hautement disponibles, tolérantes aux pannes et évolutives que ce que peut vous offrir un centre de données unique. Le cloud AWS est présent dans 80 zones de disponibilité réparties dans 25 régions à travers le monde. Des zones de disponibilité et des régions supplémentaires devraient prochainement faire leur apparition. Pour de plus amples informations sur les régions AWS et les zones de disponibilité AWS, veuillez consulter Infrastructure mondiale AWS

Chaque région Amazon est conçue pour être complètement isolée des autres régions. Cela permet d'atteindre la plus grande tolérance aux pannes possible et une stabilité optimale. Chaque zone de disponibilité est isolée, mais les zones de disponibilité d'une région sont connectées par des liaisons à faible latence. AWS vous offre la liberté de placer des instances et de stocker des données dans plusieurs régions géographiques, mais aussi sur plusieurs zones de disponibilité dans chaque région AWS. Chaque zone de disponibilité est conçue sous la forme d'une zone de défaillance indépendante. En d'autres termes, les zones de disponibilité sont physiquement séparées au sein d'une région métropolitaine classique et se trouvent dans des zones à très faible risque d'inondation (les catégories de zones inondables spécifiques varient d'une région AWS à l'autre). En plus de disposer d'un système UPS séparé et d'installations de génération de secours sur site, les centres de données situés dans différentes zones de disponibilité sont conçus pour être alimentés par des sous-stations indépendantes afin de réduire le risque qu'un événement survenant sur le réseau électrique affecte plusieurs zones de disponibilité. Les zones de disponibilité sont toutes reliées de manière redondante à plusieurs fournisseurs de transit de niveau 1.

Services de base :

AWS Identity and Access Management :

AWS Identity and Access Management (IAM) vous permet de contrôler de façon sécurisée l'accès aux services et ressources AWS pour vos utilisateurs. Avec IAM, vous pouvez créer et gérer des utilisateurs ainsi que des groupes AWS, et utiliser des autorisations afin de leur permettre ou non d'accéder aux ressources AWS. IAM vous permet d'effectuer les tâches suivantes :

• Gérer les utilisateurs IAM et leurs accès : vous pouvez créer des utilisateurs dans IAM, leur attribuer des informations d'identification de sécurité individuelles (clés d'accès, mots de passe et appareils d'authentification multifacteur) ou demander des informations d'identification de sécurité temporaires pour que ces utilisateurs aient accès aux services et ressources AWS. Vous pouvez gérer les autorisations pour contrôler quelles opérations un utilisateur peut effectuer.

• Gérer les rôles IAM et leurs autorisations : vous pouvez créer des rôles dans IAM et gérer les autorisations permettant de contrôler les opérations pouvant être effectuées en fonction de l'entité ou du service AWS qui endosse le rôle. Vous pouvez également définir l'entité qui est autorisée à endosser le rôle.

• Gérer les utilisateurs fédérés et leurs autorisations : vous pouvez activer la fédération d'identités pour permettre aux identités existantes (utilisateurs, groupes et rôles) de votre entreprise d'accéder à la Console de gestion AWS, d'appeler les API AWS et d'accéder aux ressources, sans avoir à créer un utilisateur IAM pour chaque identité.

Amazon VPC Amazon Virtual Private Cloud

(Amazon VPC) vous permet d'allouer une section logiquement isolée du cloud AWS dans laquelle vous pouvez lancer des ressources AWS dans un réseau virtuel que vous définissez. Vous disposez d'un contrôle total sur votre environnement de mise en réseau virtuel, notamment concernant la sélection de votre propre gamme d'adresses IP, la création de sous-réseaux et la configuration de tables de routage et de passerelles réseau. Dans votre VPC, vous pouvez utiliser aussi bien le protocole IPv4 qu'IPv6, pour un accès simple et sécurisé aux ressources et applications. Vous pouvez facilement adapter la configuration du réseau à votre instance VPC. Par exemple, vous pouvez créer un sous-réseau public pour vos serveurs web : un sous-réseau avec accès Internet qui place vos systèmes backend, comme des bases de données ou des serveurs d'application, dans un sous-réseau privé sans accès Internet. Vous pouvez exploiter plusieurs couches de sécurité, y compris les groupes de sécurité et les listes de contrôles d'accès au réseau, afin de renforcer le contrôle des accès aux instances EC2 dans chaque sous-réseau. De plus, vous pouvez établir une connexion matérielle VPN entre votre centre de données d'entreprise et votre VPC, et profiter du cloud AWS comme d'une extension de ce centre de données

Amazon EC2 Amazon Elastic Compute Cloud

(Amazon EC2) est un service web qui offre une capacité de calcul sécurisée et redimensionnable dans le cloud. Il est conçu pour faciliter l'informatique à l'échelle du web pour les développeurs. L'interface web simple d'Amazon EC2 vous permet d'obtenir et de configurer facilement la capacité. Elle fournit un contrôle complet de vos ressources informatiques et vous permet d'exécuter votre application dans l'environnement informatique éprouvé d'Amazon. Amazon EC2 réduit à quelques minutes le temps nécessaire pour obtenir et démarrer de nouvelles instances de serveurs (appelées instances Amazon EC2), ce qui vous permet de rapidement mettre à l'échelle la capacité, en l'augmentant et en la diminuant, au fur et à mesure de l'évolution de vos besoins informatiques. Amazon EC2 change l'aspect financier de l'informatique en vous permettant de ne payer que pour la capacité que vous utilisez réellement. Amazon EC2 fournit aux administrateurs système les outils nécessaires pour créer des applications résistant aux pannes tout en évitant les scénarios de défaillance les plus courants

Amazon Elastic Block Store Amazon Elastic Block Store

(Amazon EBS) fournit des volumes de stockage permanent au niveau bloc à utiliser avec des instances Amazon EC2 dans le cloud AWS. Chaque volume Amazon EBS est automatiquement répliqué au sein de sa zone de disponibilité, afin de vous protéger contre toute défaillance de composants, tout en garantissant une disponibilité et une durabilité élevées. Les volumes Amazon EBS offrent les performances homogènes, à faible latence, nécessaires pour exécuter vos charges de travail. Avec Amazon EBS, vous pouvez augmenter ou diminuer votre utilisation en quelques minutes, tout en payant pour ce que vous mettez en service à moindre coût.

Amazon Simple Storage Service Amazon Simple Storage Service

(Amazon S3) est un service de stockage d'objets qui offre une capacité de mise à l'échelle, une disponibilité des données, une sécurité et des performances de pointe. Cela signifie que les clients de toutes tailles et de tous secteurs peuvent l'utiliser pour stocker et protéger n'importe quelle quantité de données pour un éventail de cas d'utilisation, tels que les sites web, les applications mobiles, la sauvegarde et la restauration, l'archivage, les applications métier, les appareils IoT et l'analyse du Big Data. Amazon S3 fournit des fonctions de gestion faciles à utiliser qui vous permettent d'organiser vos données et de configurer des contrôles d'accès précis afin de répondre aux exigences spécifiques de votre entreprise, de votre organisation et de votre conformité. Amazon S3 offre une durabilité de 99,999999999 % et stocke des données pour des millions d'applications pour des entreprises du monde entier.

Services intégrés :

Elastic Load Balancing Elastic Load Balancing (ELB)

distribue automatiquement le trafic d'application entrant sur plusieurs cibles, comme des instances Amazon EC2, des conteneurs et les adresses IP. Il peut traiter 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.

• Application Load Balancer est adapté pour équilibrer les charges du trafic HTTP et HTTPS et fournit un routage de requête avancé pour la fourniture d'architectures d'applications modernes, incluant les microservices et les conteneurs. Fonctionnant au niveau des requêtes individuelles (couche 7), Application Load Balancer achemine le trafic vers des cibles au sein d'Amazon Virtual Private Cloud (Amazon VPC) en fonction du contenu de la requête.

• Network Load Balancer est parfaitement adapté pour la répartition de charge du trafic TCP, qui nécessite des performances extrêmes. Fonctionnant au niveau des connexions (couche 4), Network Load Balancer achemine le trafic vers des cibles au sein d'Amazon Virtual Private Cloud (Amazon VPC) et peut traiter des millions de requêtes par seconde, tout en maintenant des temps de latence extrêmement faibles. Network Load Balancer est aussi optimisé pour traiter les modèles de trafic soudains ou volatiles.

• Gateway Load Balancer facilite le déploiement, l'évolutivité et l'exécution des appliances de réseau virtuel de tiers. En proposant la répartition de charge et la scalabilité automatique pour les flottes d'appliances tierces, Gateway Load Balancer est transparent pour la source et la destination du trafic. Cette capacité en fait un outil bien adapté pour fonctionner avec des appliances tierces pour la sécurité, l'analyse de réseau et d'autres cas d'utilisation.

• Classic Load Balancer assure une répartition de charge de base entre plusieurs instances Amazon EC2 et fonctionne aussi bien au niveau des requêtes que des connexion. Classic Load Balancer est conçu pour les applications créées dans le réseau EC2-Classic

AWS Auto Scaling AWS

Auto Scaling permet de surveiller vos applications et d'ajuster automatiquement la capacité pour maintenir des performances stables et prévisibles au coût le plus bas possible. Grâce à AWS Auto Scaling, il est facile de configurer la mise à l'échelle de l'application pour diverses ressources réparties entre de multiples services en quelques minutes seulement. Le service est doté d'une interface utilisateur à la fois simple et performante qui vous permet de mettre en place des plans de mise à l'échelle pour les ressources, notamment les instances Amazon EC2 et les parcs d'instances Spot, les tâches Amazon ECS, les tables et les index Amazon DynamoDB et les réplicas Amazon Aurora. AWS Auto Scaling simplifie la mise à l'échelle avec des recommandations qui vous permettent d'optimiser les performances ou les coûts, ou d'établir un équilibre entre eux. Si vous utilisez déjà Amazon EC2 Auto Scaling pour mettre à l'échelle vos instance Amazon EC2 dynamiquement, vous pouvez désormais le combiner avec AWS Auto Scaling pour mettre à l'échelle des ressources supplémentaires pour d'autres services AWS. Avec AWS Auto Scaling, vos applications disposent toujours des ressources adéquates au bon moment

Amazon Route 53

Amazon Route 53 est un service web de système de noms de domaine (DNS) cloud hautement disponible et évolutif. Il 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 lisibles par les humains, 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, Amazon Route 53 est entièrement conforme au protocole IPv6. Amazon Route 53 connecte efficacement les demandes des utilisateurs à l'infrastructure s'exécutant dans AWS, et notamment aux instances EC2, aux équilibreurs de charge Elastic Load Balancing ou aux compartiments Amazon S3.

Ce service permet également d'acheminer les utilisateurs vers une infrastructure extérieure à AWS. Vous pouvez utiliser Amazon Route 53 pour configurer les surveillances de l'état DNS qui vous permettront d'acheminer le trafic vers des points de terminaison sains ou de contrôler de manière indépendante l'état de votre application et de ses points de terminaison. Le flux de trafic Amazon Route 53 facilite la gestion du trafic à l'échelle mondiale via plusieurs types de routage, dont le routage basé sur la latence, le DNS géographique et la fonctionnalité WRR (Weighted Round Robin ou technique du tourniquet pondéré).

Tous ces types peuvent être combinés avec le basculement DNS afin de créer plusieurs architectures à faible latence et tolérantes aux pannes. Grâce à l'éditeur visuel simple du flux de trafic Amazon Route 53, vous pouvez facilement gérer la façon dont les utilisateurs finaux sont acheminés vers les points de terminaison de votre application, dans une région AWS unique ou distribuée aux quatre coins du monde. Amazon Route 53 propose également un enregistrement de noms de domaine. Vous pouvez acheter et gérer des noms de domaine, comme example.com, et Amazon Route 53 configure automatiquement les paramètres DNS de ces domaines.

Amazon RDS on VMware Amazon

Relational Database Service (Amazon RDS) on VMware vous permet de déployer des bases de données gérées dans des environnements VMware sur site à l'aide de la technologie Amazon RDS dont bénéficient des centaines de milliers de clients AWS.

Amazon RDS offre une capacité économique et ajustable ainsi qu'une automatisation des tâches d'administration chronophages, telles que l'allocation du matériel, la configuration des bases de données, l'application de correctifs et les sauvegardes, vous permettant ainsi de vous concentrer sur vos applications. RDS on VMware apporte les mêmes avantages à vos déploiements sur site, en facilitant la configuration, l'exploitation et la mise à l'échelle des bases de données dans les centres de données privés VMware vSphere ou leur migration vers AWS.

Amazon RDS on VMware vous permet d'utiliser la même interface simple de gestion des bases de données dans les environnements VMware sur site que celle que vous utiliseriez dans AWS.

Vous pouvez facilement répliquer des bases de données RDS on VMware vers des instances RDS dans AWS, vous permettant ainsi d'effectuer des déploiements hybrides peu coûteux pour la reprise après sinistre, la transmission en paquets de réplicas en lecture et la conservation de sauvegardes à long terme en option dans Amazon Simple Storage Service (Amazon S3).

AWS Lambda

AWS Lambda vous permet d'exécuter du code sans avoir à allouer ni gérer des serveurs. Vous payez uniquement le temps de calcul utilisé et ne déboursez rien quand votre code ne s'exécute pas. Avec Lambda, vous pouvez exécuter le code pour quasiment n'importe quel type d'application ou service backend, sans avoir à vous préoccuper de leur administration. Il vous suffit de télécharger votre code et Lambda s'occupe de tout ce qui est nécessaire à l'exécution et à la mise à l'échelle de votre code en garantissant une haute disponibilité.

Vous pouvez configurer le code de manière à le déclencher automatiquement depuis d'autres services AWS ou l'appeler directement à partir de n'importe quelle application web ou mobile

AWS Elastic Beanstalk

AWS Elastic Beanstalk est un service simple à utiliser pour déployer et mettre à l'échelle des applications et services web développés avec Java, .NET, PHP, Node.js, Python, Ruby, Go et Docker sur des serveurs connus, tels qu'Apache, Nginx, Passenger et Internet Information Services (IIS). Il vous suffit de télécharger votre code pour qu'AWS Elastic Beanstalk gère automatiquement les étapes du déploiement, de la mise en service des capacités à la répartition de charge, en passant par 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.

Amazon Simple Notification Service

Amazon Simple Notification Service (Amazon SNS) est un service de messagerie pub/sub hautement disponible, sécurisé, durable et entièrement géré qui vous permet de découpler des microservices, des systèmes distribués et des applications sans serveur. Amazon SNS fournit des rubriques pour des messages à haut débit, en mode « push », de type « plusieurs à plusieurs ».

Grâce aux rubriques Amazon SNS, vos systèmes de publication peuvent déployer des messages vers un grand nombre de points de terminaison d'abonnés pour un traitement parallèle, y compris des files d'attente Amazon SQS, des fonctions AWS Lambda et des webhooks HTTP/S. De plus, SNS peut être utilisé pour diffuser des notifications aux utilisateurs finaux à l'aide du push mobile, des SMS et des e-mails.

Amazon CloudWatch

Amazon CloudWatch est un service de surveillance et de gestion conçu pour les développeurs, les opérateurs système, les techniciens de fiabilité des sites (SRE) et les informaticiens. CloudWatch vous offre des données et des informations qui vous permettent de surveiller vos applications, de comprendre les variations de performances systémiques et d'y répondre, d'optimiser l'utilisation des ressources et d'obtenir une vue unifiée de l'état opérationnel.

CloudWatch collecte les données opérationnelles et de surveillance sous forme de journaux, de métriques et d'événements pour vous offrir une vue unifiée des ressources, des applications et des services AWS exécutés sur AWS et les serveurs sur site. CloudWatch permet de définir des alarmes haute résolution, de visualiser côte à côte les journaux et les métriques, d'effectuer des actions automatisées, de résoudre des problèmes et d'acquérir des connaissances pour optimiser vos applications et assurer leur bon fonctionnement.

Amazon CloudFront

Amazon CloudFront est un réseau de diffusion de contenu (CDN) rapide qui distribue en toute sécurité des données, des vidéos, des applications et des API à vos utilisateurs, avec une faible latence et des vitesses de transfert élevées, le tout dans un environnement convivial pour les développeurs. CloudFront est intégré à AWS, deux emplacements physiques directement connectés à l'infrastructure mondiale AWS, ainsi qu'à d'autres services AWS.

CloudFront fonctionne de manière transparente avec certains services, notamment AWS Shield pour l'atténuation des attaques par déni de service (DDoS), Amazon S3, Elastic Load Balancing ou Amazon EC2 qui sont à l'origine de vos applications, et Lambda@Edge pour exécuter un code personnalisé plus proche des utilisateurs des clients et pour personnaliser l'expérience utilisateur.

Vous pouvez démarrer avec le réseau de diffusion de contenu en quelques minutes, en utilisant les outils AWS que vous connaissez déjà : des API, la Console de gestion AWS, AWS CloudFormation, les interfaces de ligne de commande et les kits SDK. Le réseau de diffusion de contenu d'Amazon offre un modèle de paiement à l'utilisation simple sans frais initiaux ni engagement à long terme. La prise en charge du réseau de diffusion de contenu est incluse dans votre abonnement AWS Support existant.

AWS CloudFormation

AWS CloudFormation permet aux développeurs et aux administrateurs système de créer et de gérer facilement un ensemble de ressources AWS associées, de les allouer et de les actualiser de manière ordonnée et prévisible. Vous pouvez utiliser des exemples de modèles AWS CloudFormation ou créer vos propres modèles pour décrire vos ressources AWS, ainsi que les dépendances ou paramètres d'exécution associés nécessaires à l'exécution de votre application.

Vous n'avez pas besoin de déterminer l'ordre d'approvisionnement des services AWS ou de connaître les subtilités de fonctionnement de ces dépendances. CloudFormation s'en charge à votre place. Une fois les ressources AWS déployées, vous pouvez les modifier et les mettre à jour de manière prévisible et contrôlée, et assurer ainsi un contrôle de version de votre infrastructure AWS similaire à celui que vous réalisez pour vos logiciels.

Vous pouvez également visualiser vos modèles sous forme de schémas et les modifier à l'aide d'une interface glisser-déposer avec AWS CloudFormation Designer