AWS Livre blanc

Limites d'isolation des pannes AWS

Limites d'isolation des pannes AWS: AWS Livre blanc

Table of Contents

AWS Well-Architected

Date de publication : 16 novembre 2022 (Révisions du document)

Résumé

Amazon Web Services (AWS) fournit différentes limites d'isolation, telles que les zones de disponibilité (AZ), les régions, les plans de contrôle et les plans de données. Ce paper explique comment ces limites sont AWS utilisées pour créer des services zonaux, régionaux et mondiaux. Il inclut également des conseils prescriptifs sur la manière de prendre en compte les dépendances à l'égard de ces différents services et d'améliorer la résilience des charges de travail que vous créez à l'aide de ces services.

Well-Architected

Le AWSWell-Architected Framework vous aide à comprendre les avantages et les inconvénients des décisions que vous prenez lors de la création de systèmes dans le cloud. Les six piliers du Framework vous permettent d'apprendre les meilleures pratiques architecturales pour concevoir et exploiter des systèmes fiables, sécurisés, efficaces, rentables et durables. À l'aide du AWS Well-Architected Tool, disponible gratuitement dans le AWS Management Console, vous pouvez examiner vos charges de travail par rapport à ces meilleures pratiques en répondant à une série de questions pour chaque pilier.

Pour obtenir des conseils d'experts supplémentaires et les meilleures pratiques concernant votre architecture cloud (déploiements d'architecture de référence, diagrammes et livres blancs), consultez le Centre d'architecture. AWS

Introduction

AWSexploite une infrastructure mondiale pour fournir des services cloud qui aident les clients à déployer des charges de travail de manière flexible, sécurisée, évolutive et hautement disponible. L'AWSinfrastructure utilise plusieurs structures d'isolation des pannes pour aider les clients à atteindre leurs objectifs de résilience. Ces limites d'isolation des pannes permettent aux clients de concevoir leurs charges de travail de manière à tirer parti de l'étendue prévisible de la limitation des impacts qu'ils fournissent. Il est également important de comprendre comment les AWS services

sont conçus en fonction de ces limites afin de pouvoir faire des choix intentionnels concernant les dépendances que vous sélectionnez pour votre charge de travail.

Ce paper résumera d'abord l'infrastructure AWS mondiale et les limites d'isolation des pannes qu'elle fournit, ainsi que certains des modèles utilisés pour concevoir nos services. À l'aide de cette base de compréhension, le paper décrira ensuite les différents domaines de services AWS fournis : zonal, régional et mondial. Il présentera également les meilleures pratiques pour créer des architectures qui utilisent ces limites d'isolement et différentes étendues de service afin d'améliorer la résilience des charges de travail sur lesquelles vous exécutez. AWS Il fournit notamment des conseils prescriptifs sur la manière de gérer les dépendances vis-à-vis des services mondiaux tout en minimisant les points de défaillance uniques. Cela vous aidera à faire des choix éclairés concernant vos AWS dépendances et la manière dont vous concevez votre charge de travail pour la haute disponibilité (HA) et la reprise après sinistre (DR).

infrastructure AWS

Cette section présente un résumé de l'infrastructure AWS globale et des limites d'isolation des pannes qu'elle fournit. En outre, cette section fournira un aperçu des concepts de plans de contrôle et de plans de données, qui constituent des distinctions essentielles dans la AWS conception de ses services. Ces informations fournissent la base de référence pour comprendre comment les limites d'isolation des pannes ainsi que le plan de contrôle et le plan de données d'un AWS service s'appliquent aux types de services dont nous parlerons dans la section suivante.

Rubriques

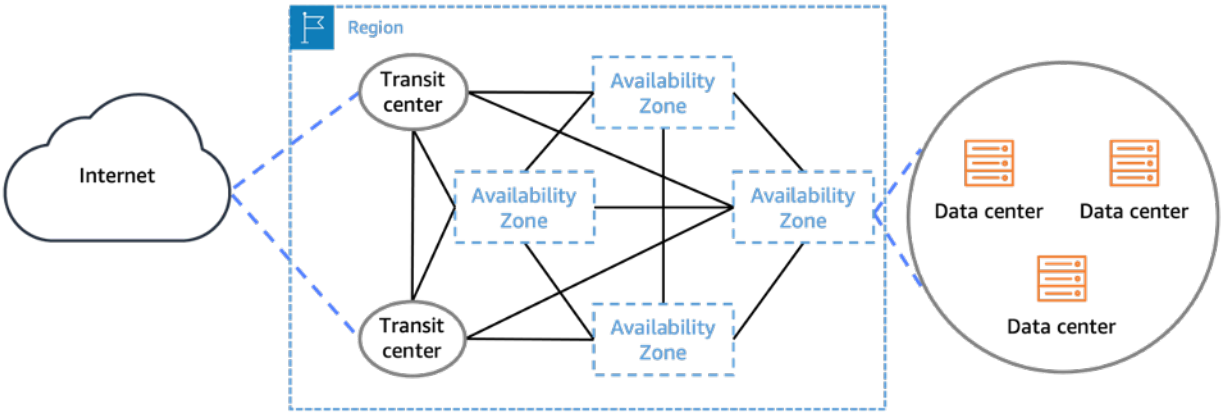
Zones de disponibilité

AWSgère plus de 100 zones de disponibilité dans plusieurs régions du monde (les chiffres actuels se trouvent ici : Infrastructure AWS mondiale). Une zone de disponibilité est un ou plusieurs centres de données discrets dotés d'une infrastructure électrique, d'un réseau et d'une connectivité indépendants et redondants au sein d'unRégion AWS. Les zones de disponibilité d'une région sont nettement distantes les unes des autres, jusqu'à 60 miles (~100 km) pour éviter les défaillances corrélées, mais suffisamment proches pour utiliser la réplication synchrone avec une latence d'un chiffre en millisecondes. Ils sont conçus pour ne pas être simultanément affectés par un scénario de destin partagé, tel que l'alimentation en électricité, les ruptures d'eau, l'isolation par fibre optique, les tremblements de terre, les incendies, les tornades ou les inondations. Les points de défaillance courants, tels que les générateurs et les équipements de refroidissement, ne sont pas partagés entre les zones de disponibilité et sont conçus pour être alimentés par des sous-stations électriques indépendantes. Lors du AWS déploiement de mises à jour de ses services, les déploiements vers les

zones de disponibilité d'une même région sont séparés dans le temps afin d'éviter toute défaillance corrélée.

Toutes les zones de disponibilité d'une région sont interconnectées par un réseau à bande passante élevée et à faible latence, via une fibre métropolitaine dédiée entièrement redondante. Chaque zone de disponibilité d'une région est connectée à Internet via deux centres de transit où se trouvent des AWS homologues de plusieurs fournisseurs Internet de niveau 1 (pour plus d'informations, reportez-vous à la section Présentation d'Amazon Web Services).

Ces fonctionnalités permettent d'isoler fortement les zones de disponibilité les unes des autres, ce que nous appelons l'indépendance des zones de disponibilité (AZI). La structure logique des zones de disponibilité et de leur connectivité à Internet est illustrée dans la figure suivante.



Les zones de disponibilité se composent d'un ou de plusieurs centres de données physiques connectés de manière redondante les uns aux autres et à Internet

Régions

Chacune Région AWS se compose de plusieurs zones de disponibilité indépendantes et physiquement séparées au sein d'une zone géographique. Toutes les régions disposent actuellement de trois zones de disponibilité ou plus. Les régions elles-mêmes sont isolées et indépendantes des autres régions, à quelques exceptions près indiquées plus loin dans ce document (voir les opérations mondiales à une seule région). Cette séparation entre les régions limite les défaillances de service, lorsqu'elles se produisent, à une seule région. Les opérations normales des autres régions ne sont pas affectées dans ce cas. En outre, les ressources et les données que vous créez dans une région n'existent dans aucune autre région, sauf si vous utilisez explicitement une fonctionnalité de réplication ou de copie proposée par un AWS service ou si vous répliquez vous-même la ressource.

Régions AWS actuelles et prévues en décembre 2022

AWSZones Locales

AWSLes Zones Locales sont un type de déploiement d'infrastructure qui place le calcul, le stockage, les bases de données et d'autres AWSservices sélectionnés à proximité de grands centres urbains et industriels. Vous pouvez utiliser AWS des services, tels que les services de calcul et de stockage, dans la zone locale pour exécuter des applications à faible latence en périphérie ou pour simplifier les migrations vers le cloud hybride. Les zones locales disposent d'une entrée et d'une sortie Internet locales pour réduire la latence, mais elles sont également connectées à leur région mère via le réseau privé redondant et à large bande passante d'Amazon, ce qui permet aux applications exécutées dans les zones AWS locales d'accéder rapidement, de manière sécurisée et fluide à la gamme complète de services.

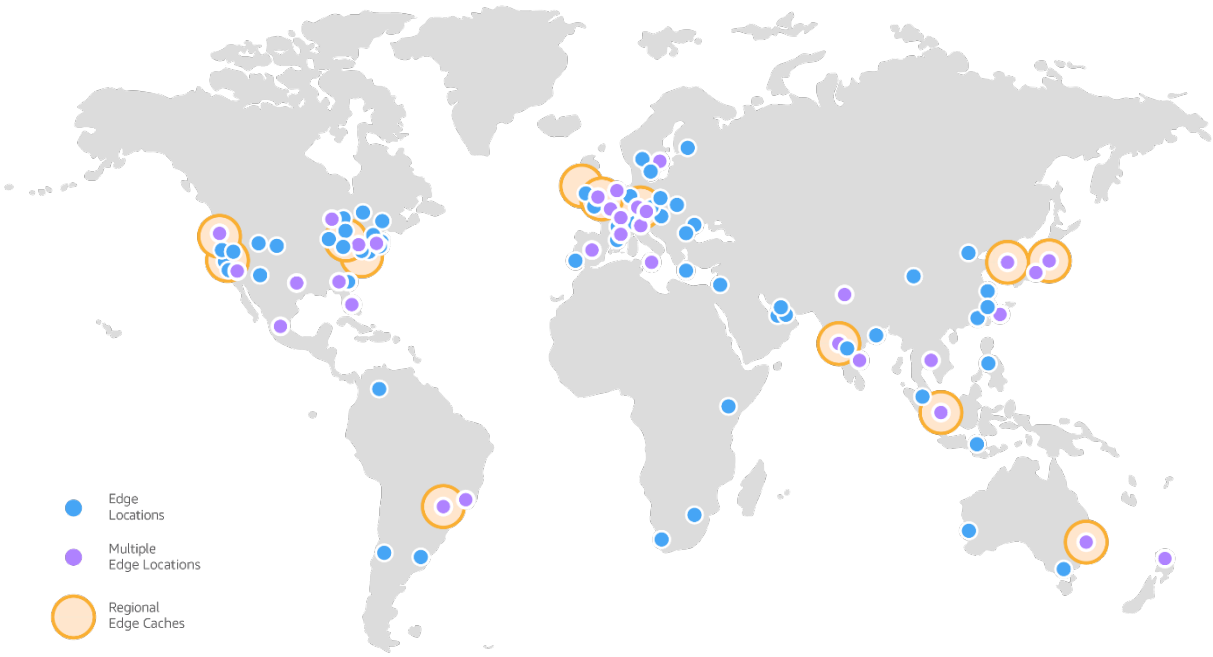
AWS Outposts

AWS Outpostsest une famille de solutions entièrement gérées fournissant une AWS infrastructure et des services à pratiquement tous les sites sur site ou en périphérie pour une expérience hybride véritablement cohérente. Les solutions Outposts vous permettent d'étendre et d'exécuter des AWS services natifs sur site. Elles sont disponibles sous différents formats, des serveurs Outposts 1U et 2U aux racks Outposts 42U, en passant par les déploiements en rack multiples.

AvecAWS Outposts, vous pouvez exécuter certains AWS services localement et vous connecter à un large éventail de services disponibles chez le parentRégion AWS. AWS Outpostssont des racks de calcul et de stockage entièrement gérés et configurables conçus avec du matériel AWS conçu qui permet aux clients d'exécuter des opérations de calcul et de stockage sur site, tout en se connectant facilement à AWS la vaste gamme de services dans le cloud.

Points de présence

Outre les zones de disponibilité Régions AWS et de disponibilité, exploite AWS également un réseau mondial de points de présence distribués (PoP). Ils PoPs hébergent Amazon CloudFront, un réseau de diffusion de contenu (CDN) ; Amazon Route 53, un service public de résolution du système de noms de domaine (DNS) ; et AWS Global Accelerator (AGA), un service d'optimisation des réseaux de pointe. Le réseau Edge mondial comprend actuellement plus de 410 PoPs, dont plus de 400 emplacements Edge, et 13 caches régionaux de milieu de gamme dans plus de 90 villes dans 48 pays (le statut actuel peut être consulté ici : Caractéristiques CloudFront principales d'Amazon).



Réseau périphérique CloudFront mondial Amazon

Chaque PoP est isolé des autres, ce qui signifie qu'une panne affectant un seul PoP ou une seule zone métropolitaine n'a pas d'impact sur le reste du réseau mondial. Le AWS réseau est compatible avec des milliers d'opérateurs de télécommunications de niveau 1/2/3 dans le monde entier, est bien connecté à tous les principaux réseaux d'accès pour des performances optimales et dispose de

centaines de térabits de capacité déployée. Les sites périphériques sont connectés Régions AWS via le backbone du AWS réseau, une fibre parallèle 100 GbE multiple entièrement redondante qui fait le tour du monde et est reliée à des dizaines de milliers de réseaux pour améliorer la récupération des données d'origine et l'accélération dynamique du contenu.

Partitions

AWSregroupe les régions en partitions. Chaque région se trouve exactement dans une partition, et chaque partition possède une ou plusieurs régions. Les partitions possèdent des instances indépendantes de AWS Identity and Access Management (IAM) et fournissent une limite stricte entre les régions des différentes partitions. AWSLes régions commerciales sont dans la `aws` partition, les régions en Chine sont dans la `aws-cn` partition et AWS GovCloud les régions sont dans la `aws-us-gov` partition. Certains AWS services sont conçus pour fournir des fonctionnalités interrégionales, comme Amazon S3 Cross-Region Replication ou le peering interrégional AWS Transit Gateway. Ces types de fonctionnalités ne sont pris en charge qu'entre les régions d'une même partition. Vous ne pouvez pas utiliser les informations d'identification IAM d'une partition pour interagir avec les ressources d'une autre partition.

Plans de contrôle et plans de données

AWSsépare la plupart des services entre les concepts de plan de contrôle et de plan de données. Ces termes proviennent du monde des réseaux, en particulier des routeurs. Le plan de données du routeur, qui est sa principale fonctionnalité, déplace les paquets selon des règles. Mais les politiques de routage doivent être créées et distribuées depuis quelque part, et c'est là qu'intervient le plan de contrôle.

Les plans de contrôle fournissent les API administratives utilisées pour créer, lire/décrire, mettre à jour, supprimer et répertorier les ressources (CRUDL). Par exemple, les actions suivantes concernent toutes le plan de contrôle : lancement d'une nouvelle instance Amazon Elastic Compute Cloud (Amazon EC2), création d'un bucket Amazon Simple Storage Service (Amazon S3) et description d'une file d'attente Amazon Simple Queue Service (Amazon SQS). Lorsque vous lancez une instance EC2, le plan de contrôle doit effectuer plusieurs tâches, telles que la recherche d'un hôte physique capable, l'allocation des interfaces réseau, la préparation d'un volume Amazon Elastic Block Store (Amazon EBS), la génération d'informations d'identification IAM, l'ajout des règles du groupe de sécurité, etc. Les plans de contrôle sont généralement des systèmes d'orchestration et d'agrégation complexes.

Le plan de données est la principale fonction du service. Par exemple, les éléments suivants constituent tous les éléments du plan de données pour chacun des services concernés : l'instance EC2 en cours d'exécution elle-même, la lecture et l'écriture sur un volume EBS, l'obtention et le placement d'objets dans un compartiment S3, et la Route 53 répondant aux requêtes DNS et effectuant des contrôles de santé.

Les plans de données sont volontairement moins compliqués, avec moins de pièces mobiles que les plans de contrôle, qui mettent généralement en œuvre un système complexe de flux de travail, de logique métier et de bases de données. Cela rend les événements de défaillance statistiquement moins susceptibles de se produire dans le plan de données par rapport au plan de contrôle. Bien que les données et le plan de contrôle contribuent au fonctionnement global et au succès du service, ils AWS sont considérés comme des composants distincts. Cette séparation présente à la fois des avantages en termes de performances et de disponibilité.

Stabilité statique

L'une des caractéristiques de résilience les plus importantes des AWS services est ce que l'on AWS appelle la stabilité statique. Ce terme signifie que les systèmes fonctionnent dans un état statique et continuent de fonctionner normalement sans qu'il soit nécessaire d'apporter des modifications en cas de défaillance ou d'indisponibilité des dépendances. Nous y parvenons notamment en empêchant les dépendances circulaires dans nos services qui pourraient empêcher le rétablissement réussi de l'un de ces services. Une autre façon d'y parvenir est de maintenir l'état existant. Nous prenons en compte le fait que les plans de contrôle sont statistiquement plus susceptibles de tomber en panne que les plans de données. Bien que le plan de données dépende généralement des données provenant du plan de contrôle, le plan de données conserve son état existant et continue de fonctionner même en cas de détérioration du plan de contrôle. L'accès aux ressources par le plan de données, une fois provisionné, ne dépend pas du plan de contrôle et n'est donc pas affecté par une quelconque altération du plan de contrôle. En d'autres termes, même si la capacité de créer, de modifier ou de supprimer des ressources est réduite, les ressources existantes restent disponibles. Cela rend AWS les plans de données statiquement stables en cas de détérioration du plan de contrôle. Vous pouvez implémenter différents modèles pour être statiquement stable face à différents types de défaillances de dépendance.

Un exemple de stabilité statique peut être trouvé dans Amazon EC2. Une fois qu'une instance EC2 a été lancée, elle est tout aussi disponible que le serveur physique d'un centre de données. Il ne dépend d'aucune API du plan de contrôle pour continuer à fonctionner ou pour recommencer à fonctionner après un redémarrage. La même propriété vaut pour d'autres AWS ressources telles que les VPC, les compartiments et objets Amazon S3 et les volumes Amazon EBS.

La stabilité statique est un concept profondément ancré dans la AWS conception de ses services, mais c'est également un modèle qui peut être utilisé par les clients. En fait, la majorité des meilleures pratiques pour utiliser les différents types de AWS services de manière résiliente consistent à implémenter la stabilité statique pour les environnements de production. Les mécanismes de rétablissement et d'atténuation les plus fiables sont ceux qui nécessitent le moins de changements pour réaliser le rétablissement. Au lieu de compter sur le plan de contrôle EC2 pour lancer de nouvelles instances EC2 afin de procéder à une restauration après une défaillance d'une zone de disponibilité, le préprovisionnement de cette capacité supplémentaire permet d'obtenir une stabilité statique. Ainsi, l'élimination des dépendances à l'égard des plans de contrôle (les API qui mettent en œuvre les modifications apportées aux ressources) dans votre processus de restauration permet de produire des charges de travail plus résilientes. Pour plus de détails sur la stabilité statique, les plans de contrôle et les plans de données, consultez l'article Static stability using Availability Zones de la bibliothèque Amazon Builders.

Récapitulatif

AWSutilise différents conteneurs de défauts dans notre infrastructure pour isoler les pannes. Les principaux conteneurs de défaillances de l'infrastructure sont les partitions, les régions, les zones de disponibilité, les plans de contrôle et les plans de données. Nous examinerons ensuite les différents types de AWS services, la manière dont ces conteneurs d'erreurs sont utilisés dans leur conception et la manière dont vous devez concevoir les charges de travail avec eux pour qu'elles soient résilientes.

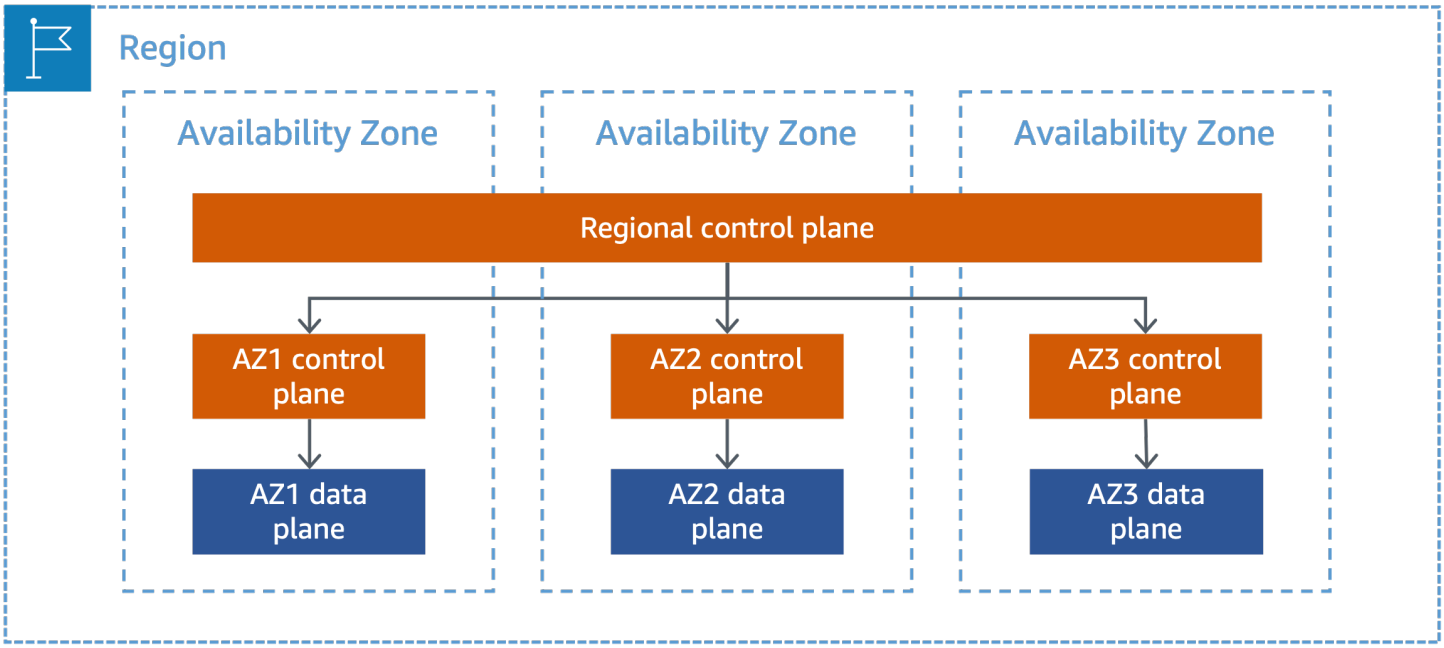
AWS types de services

AWS gère trois catégories de services différentes en fonction de leur limite d'isolation des pannes : zonal, régional et mondial. Cette section décrit plus en détail comment ces différents types de services ont été conçus afin que vous puissiez déterminer l'impact des défaillances d'un service d'un certain type de service sur votre charge de travail AWS. Il fournit également des conseils de haut niveau sur la manière de concevoir vos charges de travail afin d'utiliser ces services de manière résiliente. Pour les services mondiaux, ce document fournit également des conseils prescriptifs Annexe B - Guide de service global pour les réseaux Edge qui peuvent vous aider à éviter l'impact sur vos charges de travail des défaillances du plan de contrôle des AWS services, en vous aidant à assumer en toute sécurité votre dépendance à l'égard des services mondiaux tout en minimisant l'introduction de points de défaillance uniques. Annexe A - Directives relatives aux services partitionnés

Rubriques

Services zonaux

L'indépendance de la zone de disponibilité (AZI) permet AWS de proposer des services zonaux, tels qu'Amazon EC2 et Amazon EBS. Un service zonal est un service qui permet de spécifier dans quelle zone de disponibilité les ressources sont déployées. Ces services fonctionnent indépendamment dans chaque zone de disponibilité d'une région et, plus important encore, échouent également indépendamment dans chaque zone de disponibilité. Cela signifie que les composants d'un service dans une zone de disponibilité ne dépendent pas des composants d'autres zones de disponibilité. Nous pouvons le faire parce qu'un service zonal possède des plans de données zonaux. Dans certains cas, comme avec EC2, le service inclut également des plans de contrôle zonaux pour les opérations alignées par zone, telles que le lancement d'une instance EC2. Pour ces services, fournit AWS également un point de terminaison du plan de contrôle régional pour faciliter l'interaction avec le service. Le plan de contrôle régional fournit également des fonctionnalités à portée régionale et sert de couche d'agrégation et de routage au-dessus des plans de contrôle zonaux. Cela est illustré dans la figure suivante.



Un service zonal avec des plans de contrôle et des plans de données isolés par zone

Les zones de disponibilité permettent aux clients de gérer des charges de travail de production plus hautement disponibles, tolérantes aux pannes et évolutives que ce qui serait possible à partir d'un seul centre de données. Lorsqu'une charge de travail utilise plusieurs zones de disponibilité, les clients sont mieux isolés et protégés contre les problèmes qui affectent l'infrastructure physique d'une seule zone de disponibilité. Cela permet aux clients de créer des services redondants dans toutes les zones de disponibilité et, s'ils sont correctement architecturés, de rester opérationnels même en cas de défaillance d'une zone de disponibilité. Les clients peuvent tirer parti d'AZI pour créer des charges de travail hautement disponibles et résilientes. La mise en œuvre d'AZI dans votre architecture vous permet de récupérer rapidement après une défaillance isolée d'une zone de disponibilité, car vos ressources dans une zone de disponibilité minimisent ou éliminent l'interaction avec les ressources des autres zones de disponibilité. Cela permet de supprimer les dépendances entre les zones de disponibilité, ce qui simplifie l'évacuation des zones de disponibilité. Reportez-vous à la section Modèles de résilience multi-AZ avancés pour plus de détails sur la création de mécanismes d'évacuation des zones de disponibilité. En outre, vous pouvez tirer davantage parti des zones de disponibilité en suivant certaines des meilleures pratiques AWS utilisées pour ses propres services, telles que le déploiement des modifications sur une seule zone de disponibilité à la fois ou la suppression d'une zone de disponibilité du service si une modification de cette zone de disponibilité se passe mal.

La stabilité statique est également un concept important pour les architectures de zones de disponibilité multiple. L'un des modes de défaillance que vous devez prévoir avec les architectures

à zones de disponibilité multiples est la perte d'une zone de disponibilité, qui peut entraîner la perte de capacité d'une zone de disponibilité. Si vous n'avez pas préprovisionné suffisamment de capacité pour faire face à la perte d'une zone de disponibilité, votre capacité restante peut être submergée par la charge actuelle. En outre, vous devrez vous fier aux plans de contrôle des services zonaux que vous utilisez pour remplacer cette capacité perdue, ce qui peut être moins fiable qu'une conception statiquement stable. Dans ce cas, le pré-provisionnement d'une capacité supplémentaire suffisante peut vous aider à rester statiquement stable face à la perte d'un domaine de défaillance, tel qu'une zone de disponibilité, en étant en mesure de poursuivre vos opérations normales sans avoir besoin de modifications dynamiques.

Vous pouvez choisir d'utiliser un groupe d'instances EC2 à dimensionnement automatique déployées dans plusieurs zones de disponibilité pour effectuer une mise à l'échelle interne et descendante de manière dynamique, en fonction des besoins de votre charge de travail. La mise à l'échelle automatique fonctionne bien pour les changements graduels d'utilisation qui se produisent au fil des minutes, voire des dizaines de minutes. Cependant, le lancement de nouvelles instances EC2 prend du temps, en particulier si vos instances nécessitent un amorçage (par exemple, l'installation d'agents, de fichiers binaires d'applications ou de fichiers de configuration). Pendant ce temps, votre capacité restante pourrait être dépassée par la charge actuelle. En outre, le déploiement de nouvelles instances par le biais de l'autoscaling repose sur le plan de contrôle EC2. Cela présente un inconvénient : pour être statiquement stable face à la perte d'une seule zone de disponibilité, vous devez préapprovisionner suffisamment d'instances EC2 dans les autres zones de disponibilité pour gérer la charge qui a été transférée hors de la zone de disponibilité affectée, au lieu de vous fier à l'autoscaling pour approvisionner de nouvelles instances. Toutefois, le pré-provisionnement en capacité supplémentaire peut entraîner des coûts supplémentaires.

Par exemple, en fonctionnement normal, supposons que votre charge de travail nécessite six instances pour traiter le trafic client dans trois zones de disponibilité. Pour garantir une stabilité statique en cas de défaillance d'une seule zone de disponibilité, vous devez déployer trois instances dans chaque zone de disponibilité, pour un total de neuf. Si une seule instance correspondant à une zone de disponibilité tombait en panne, il vous en resterait six et vous pourriez continuer à traiter le trafic de vos clients sans avoir à approvisionner et à configurer de nouvelles instances en cas de panne. La stabilité statique de votre capacité EC2 entraîne un coût supplémentaire, car dans ce cas, vous exécutez 50 % d'instances supplémentaires. Tous les services pour lesquels vous pouvez préprovisionner des ressources n'entraîneront pas de coûts supplémentaires, tels que le préprovisionnement d'un compartiment S3 ou d'un utilisateur. Vous devrez évaluer les inconvénients liés à la mise en œuvre de la stabilité statique par rapport au risque de dépassement du temps de restauration souhaité pour votre charge de travail.

AWS Les Zones Locales et les Outposts rapprochent le plan de données de certains AWS services des utilisateurs finaux. Les plans de contrôle de ces services se trouvent dans la région parent. Votre instance de zone locale ou d'Outposts comportera des dépendances de plan de contrôle pour les services zonaux tels que EC2 et EBS sur la zone de disponibilité dans laquelle vous avez créé votre zone locale ou votre sous-réseau Outposts. Ils dépendront également des plans de contrôle régionaux pour les services régionaux tels qu'Elastic Load Balancing (ELB), les groupes de sécurité et le plan de contrôle Kubernetes géré par Amazon Elastic Kubernetes Service (Amazon EKS) (si vous utilisez EKS). Pour plus d'informations spécifiques à Outposts, reportez-vous à la documentation et aux FAQ sur le support et la maintenance. Mettez en œuvre la stabilité statique lorsque vous utilisez des Zones Locales ou des Outposts pour améliorer la résilience en cas de défaillance du plan de contrôle ou d'interruption de la connectivité réseau avec la région parent.

Services régionaux

Les services régionaux sont des services qui AWS s'appuient sur plusieurs zones de disponibilité afin que les clients n'aient pas à déterminer comment tirer le meilleur parti des services zonaux. Nous regroupons logiquement le service déployé dans plusieurs zones de disponibilité afin de présenter un point de terminaison régional unique aux clients. Amazon SQS et Amazon DynamoDB sont des exemples de services régionaux. Ils utilisent l'indépendance et la redondance des zones de disponibilité pour minimiser les défaillances de l'infrastructure en tant que catégorie de risque de disponibilité et de durabilité. Amazon S3, par exemple, répartit les demandes et les données entre plusieurs zones de disponibilité et est conçu pour effectuer une restauration automatique en cas de défaillance d'une zone de disponibilité. Toutefois, vous n'interagissez qu'avec le point de terminaison régional du service.

AWS estime que la plupart des clients peuvent atteindre leurs objectifs de résilience dans une seule région en utilisant des services régionaux ou des architectures multi-AZ qui s'appuient sur des services zonaux. Toutefois, certaines charges de travail peuvent nécessiter une redondance supplémentaire, et vous pouvez utiliser l'isolation de Régions AWS pour créer des architectures multirégionales à des fins de haute disponibilité ou de continuité des activités. La séparation physique et logique Régions AWS permet d'éviter les défaillances corrélées entre eux. En d'autres termes, comme si vous étiez un client EC2 et que vous pouviez bénéficier de l'isolation des zones de disponibilité en les déployant sur plusieurs d'entre elles, vous pouvez bénéficier des mêmes avantages pour les services régionaux en les déployant dans plusieurs régions. Cela nécessite que vous implémentiez une architecture multirégionale pour votre application, ce qui peut vous aider à résister à la détérioration d'un service régional.

Cependant, il peut être difficile de tirer parti des avantages d'une architecture multirégionale ; il faut travailler avec soin pour tirer parti de l'isolement régional sans rien annuler au niveau de l'application. Par exemple, si vous basculez une application entre régions, vous devez maintenir une séparation stricte entre vos piles d'applications dans chaque région, connaître toutes les dépendances entre les applications et basculer toutes les parties de l'application en même temps. Pour y parvenir, une architecture complexe basée sur des microservices comportant de nombreuses dépendances entre les applications nécessite une planification et une coordination entre de nombreuses équipes d'ingénierie et commerciales. Permettre aux charges de travail individuelles de prendre leurs propres décisions en matière de basculement simplifie la coordination, mais introduit un comportement modal en raison de la différence significative de latence entre les régions par rapport à l'intérieur d'une même région.

AWS ne fournit pas de fonctionnalité de réplication synchrone entre régions pour le moment. Lorsque vous utilisez une banque de données répliquée de manière asynchrone (fournie par AWS) entre les régions, il existe un risque de perte de données ou d'incohérence lorsque vous basculez votre application entre les régions. Pour atténuer les éventuelles incohérences, vous avez besoin d'un processus de rapprochement des données fiable dans lequel vous pouvez avoir confiance et qui peut avoir besoin d'opérer sur plusieurs magasins de données au sein de votre portefeuille de charges de travail, ou vous devez être prêt à accepter une perte de données. Enfin, vous devez pratiquer le failover pour savoir qu'il fonctionnera quand vous en aurez besoin. La rotation régulière de votre application entre les régions pour pratiquer le basculement sur incident représente un investissement considérable en temps et en ressources. Si vous décidez d'utiliser une banque de données répliquée de manière synchrone entre plusieurs régions pour prendre en charge vos applications exécutées simultanément à partir de plusieurs régions, les caractéristiques de performance et de latence d'une telle base de données qui s'étend sur des centaines ou des milliers de kilomètres sont très différentes de celles d'une base de données fonctionnant dans une seule région. Cela vous oblige à planifier votre pile d'applications à partir de zéro pour tenir compte de ce comportement. Cela rend également la disponibilité des deux régions fortement dépendante, ce qui peut entraîner une diminution de la résilience de votre charge de travail.

Services mondiaux

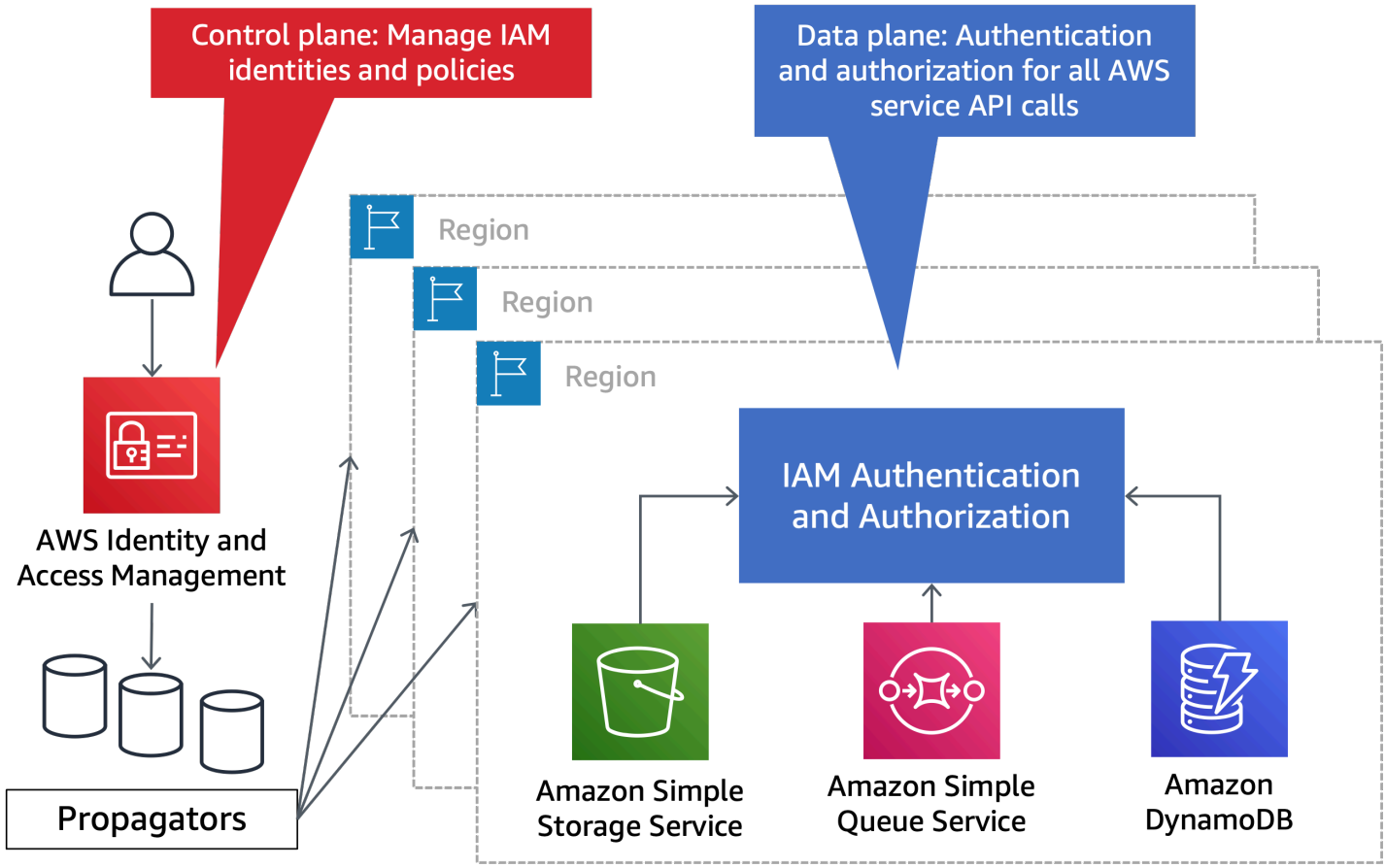
Outre les AWS services régionaux et zonaux, il existe un petit ensemble de AWS services dont les plans de contrôle et les plans de données n'existent pas indépendamment dans chaque région. Comme leurs ressources ne sont pas spécifiques à une région, elles sont communément appelées « mondiales ». Les AWS services globaux suivent toujours le schéma de AWS conception classique qui consiste à séparer le plan de contrôle du plan de données afin d'obtenir une stabilité statique.

La différence significative pour la plupart des services mondiaux est que leur plan de contrôle est hébergé dans un seul Région AWS, tandis que leur plan de données est distribué dans le monde entier. Il existe trois types différents de services globaux et un ensemble de services qui peuvent sembler globaux en fonction de la configuration que vous avez sélectionnée.

Les sections suivantes identifieront chaque type de service global et la manière dont leurs plans de contrôle et leurs plans de données sont séparés. Vous pouvez utiliser ces informations pour vous aider à créer des mécanismes fiables de haute disponibilité (HA) et de reprise après sinistre (DR) sans avoir à dépendre d'un plan de contrôle de service global. Cette approche permet d'éliminer les points de défaillance uniques de votre architecture et d'éviter les impacts potentiels entre régions, même lorsque vous opérez dans une région différente de celle où le plan de contrôle des services global est hébergé. Il vous aide également à mettre en œuvre en toute sécurité des mécanismes de basculement qui ne reposent pas sur des plans de contrôle de service mondiaux.

Des services globaux uniques par partition

Certains AWS services globaux existent dans chaque partition (désignés dans ce paper sous le nom de services partitionnels). Les services partitionnels fournissent leur plan de contrôle en une seule Région AWS fois. Certains services partitionnels, tels que AWS Network Manager, ne concernent que le plan de contrôle et orchestrent le plan de données d'autres services. D'autres services partitionnels, tels que IAM, possèdent leur propre plan de données qui est isolé et distribué sur l'ensemble de Régions AWS la partition. Les défaillances d'un service partitionnel n'ont aucune incidence sur les autres partitions. Dans la `aws` partition, le plan de contrôle du service IAM se trouve dans la `us-east-1` région, avec des plans de données isolés dans chaque région de la partition. Les services partitionnels disposent également de plans de contrôle et de plans de données indépendants dans les `aws-cn` partitions `aws-us-gov` et. La séparation du plan de contrôle et du plan de données pour IAM est illustrée dans le schéma suivant.



IAM dispose d'un plan de contrôle unique et d'un plan de données régionalisé

Les services partitionnels et l'emplacement de leur plan de contrôle dans la `aws` partition sont les suivants :

- AWS JE SUIS () `us-east-1`
- AWS Organizations (`us-east-1`)
- AWS Gestion de compte (`us-east-1`)
- Route 53 Application Recovery Controller (ARC) (`us-west-2`) - Ce service n'est présent que dans la `aws` partition
- AWS Gestionnaire de réseau (`us-west-2`)
- DNS privé Route 53 (`us-east-1`)

Si l'un de ces plans de contrôle de service présente un événement ayant une incidence sur la disponibilité, il se peut que vous ne puissiez pas utiliser les opérations de type CRUDL fournies par ces services. Ainsi, si votre stratégie de reprise dépend de ces opérations, un impact sur la

disponibilité du plan de contrôle ou de la région hébergeant le plan de contrôle réduira vos chances de réussite du rétablissement. Annexe A - Directives relatives aux services partitionnésfournit des stratégies pour supprimer les dépendances vis-à-vis des plans de contrôle de service globaux lors de la restauration.

> **Recommandation**
>
> Ne vous fiez pas aux plans de contrôle des services partitionnels dans votre processus de restauration. Fiez-vous plutôt aux opérations du plan de données de ces services. Consultez Annexe A - Directives relatives aux services partitionnés pour plus de détails sur la manière dont vous devez concevoir pour les services partitionnels.

Des services mondiaux dans le réseau périphérique

L'ensemble de AWS services mondiaux suivant possède un plan de contrôle dans la `aws` partition et héberge ses plans de données dans l'infrastructure des points de présence mondiaux (PoP) (et potentiellement Régions AWS aussi). Les plans de données hébergés dans sont PoPs accessibles à partir des ressources de n'importe quelle partition ainsi que sur Internet. Par exemple, Route 53 exploite son plan de contrôle dans la `us-east-1` région, mais son plan de données est distribué sur des centaines de sites dans le PoPs monde entier, ainsi que sur chacun d'entre eux Région AWS (pour prendre en charge les DNS publics et privés de Route 53 dans la région). Les contrôles de santé de la Route 53 font également partie du plan de données et sont effectués à partir de huit Régions AWS points de la `aws` partition. Les clients peuvent résoudre le DNS à l'aide des zones hébergées publiques Route 53 depuis n'importe où sur Internet, y compris d'autres partitions GovCloud, ainsi que depuis un AWS Virtual Private Cloud (VPC). Les services de réseau de périphérie mondiaux et l'emplacement de leur plan de contrôle dans la `aws` partition sont les suivants :

- DNS public Route 53 (`us-east-1`)
- Amazon CloudFront (`us-east-1`)
- AWS WAF Classique pour CloudFront (`us-east-1`)
- AWS WAF pour CloudFront (`us-east-1`)
- Amazon Certificate Manager (ACM) pour CloudFront (`us-east-1`)
- AWS Global Accelerator (AGA) (`us-west-2`)
- AWS Shield Advanced (`us-east-1`)

Si vous utilisez les contrôles de santé AGA pour les instances EC2 ou les adresses IP élastiques, ceux-ci utilisent les contrôles de santé Route 53. La création ou la mise à jour des bilans de santé AGA dépendrait du plan de contrôle de la Route 53 intégré`us-east-1`. L'exécution des bilans de santé AGA utilise le plan de données des bilans de santé Route 53.

Lors d'une panne affectant la région hébergeant les plans de contrôle de ces services, ou d'une panne affectant le plan de contrôle lui-même, il se peut que vous ne puissiez pas utiliser les opérations de type CRUDL fournies par ces services. Si vous avez intégré des dépendances à ces opérations dans votre stratégie de restauration, celle-ci a moins de chances de réussir que si vous vous fiez uniquement au plan de données de ces services.

> **Recommandation**
>
> Ne vous fiez pas au plan de contrôle des services du réseau périphérique dans votre processus de restauration. Fiez-vous plutôt aux opérations du plan de données de ces services. Consultez Annexe B - Guide de service global pour les réseaux Edge pour plus de détails sur la façon de concevoir des services mondiaux dans le réseau de périphérie.

Opérations mondiales dans une seule région

La dernière catégorie est composée d'opérations de plan de contrôle spécifiques au sein d'un service ayant une portée globale, et non de services complets comme dans les catégories précédentes. Lorsque vous interagissez avec les services zonaux et régionaux de la région que vous spécifiez, certaines opérations dépendent d'une seule région différente de celle où se trouve la ressource. Ils sont différents des services qui ne sont fournis que dans une seule région ; consultez la liste de ces services. Annexe C - Services pour une seule région

Lors d'une panne ayant un impact sur la dépendance globale sous-jacente, il se peut que vous ne puissiez pas utiliser les actions de type CRUDL des opérations dépendantes. Si vous avez intégré des dépendances à ces opérations dans votre stratégie de restauration, celle-ci a moins de chances de réussir que si vous vous fiez uniquement au plan de données de ces services. Vous devez éviter de dépendre de ces opérations dans le cadre de votre stratégie de restauration.

Voici une liste de services dont d'autres services peuvent dépendre et qui ont une portée globale :

- Route 53

Plusieurs AWS services créent des ressources qui fournissent un ou plusieurs noms DNS spécifiques à la ressource. Par exemple, lorsque vous configurez un Elastic Load Balancer (ELB), le service crée des enregistrements DNS publics et des contrôles de santé dans Route 53 pour l'ELB. Cela repose sur le plan de contrôle de la Route 53 intégré`us-east-1`. Les autres services que vous utilisez peuvent également avoir besoin de configurer un ELB, de créer des enregistrements DNS Route 53 publics ou de créer des bilans de santé Route 53 dans le cadre de leurs flux de travail sur le plan de contrôle. Par exemple, le provisionnement d'une ressource d'API REST Amazon API Gateway, d'une base de données Amazon Relational Database Service (Amazon RDS) ou d'un domaine OpenSearch Amazon Service entraîne la création d'enregistrements DNS dans Route 53. Vous trouverez ci-dessous une liste de services dont le plan de contrôle dépend du plan de contrôle Route 53 `us-east-1` pour créer, mettre à jour ou supprimer des enregistrements DNS, des zones hébergées et/ou créer des bilans de santé Route 53. Cette liste n'est pas exhaustive ; elle vise à mettre en évidence certains des services les plus couramment utilisés dont les actions du plan de contrôle pour créer, mettre à jour ou supprimer des ressources dépendent du plan de contrôle Route 53 :

- API REST et HTTP Amazon API Gateway
- Instances Amazon RDS
- Bases de données Amazon Aurora
- Équilibreurs de charge Amazon ELB
- AWS PrivateLink Points de terminaison VPC
- AWS Lambda URL
- Amazon ElastiCache
- Amazon OpenSearch Service
- Amazon CloudFront
- Amazon MemoryDB for Redis
- Amazon Neptune
- Amazon DynamoDB Accelerator (DAX)
- AGA
- Amazon Elastic Container Service (Amazon ECS) avec Service Discovery basé sur le DNS (qui utilise AWS Cloud Map l'API pour gérer le DNS Route 53)
- Plan de contrôle Amazon EKS Kubernetes

Il est important de noter que le service DNS VPC pour les noms d'hôtes des instances EC2 existe indépendamment dans chacune d'elles Région AWS et ne dépend pas du plan de contrôle Route 53. Les enregistrements AWS créés pour les instances EC2 dans le service DNS VPC, `ip-10-0-10.ec2.internal` tels que,`ip-10-0-1-5.compute.us-west-2.compute.internal`, `i-0123456789abcdef.us-west-2.compute.internal` et`i-0123456789abcdef.ec2.internal`, ne reposent pas sur le plan de contrôle Route 53 dans. `us-east-1`

> **Recommandation**
>
> Ne vous fiez pas à la création, à la mise à jour ou à la suppression de ressources qui nécessitent la création, la mise à jour ou la suppression d'enregistrements de ressources Route 53, de zones hébergées ou de bilans de santé dans votre parcours de restauration. Préapprovisionnez ces ressources, telles que les ELB, pour éviter de dépendre du plan de contrôle Route 53 dans votre parcours de reprise.

- Amazon S3

Les opérations du plan de contrôle Amazon S3 suivantes ont une dépendance sous-jacente sur `us-east-1` la `aws` partition. Une panne affectant Amazon S3 ou d'autres services `us-east-1` pourrait perturber les actions de ces plans de contrôle dans d'autres régions :

```
PutBucketCors
DeleteBucketCors
PutAccelerateConfiguration
PutBucketRequestPayment
PutBucketObjectLockConfiguration
PutBucketTagging
DeleteBucketTagging
PutBucketReplication
DeleteBucketReplication
PutBucketEncryption
DeleteBucketEncryption
PutBucketLifecycle
DeleteBucketLifecycle
PutBucketNotification
PutBucketLogging
```

```
DeleteBucketLogging
PutBucketVersioning
PutBucketPolicy
DeleteBucketPolicy
PutBucketOwnershipControls
DeleteBucketOwnershipControls
PutBucketAcl
PutBucketPublicAccessBlock
DeleteBucketPublicAccessBlock
```

Le plan de contrôle des points d'accès multirégionaux (MRAP) Amazon S3 est hébergé uniquement dans cette région `us-west-2` et les demandes de création, de mise à jour ou de suppression de MRAP ciblent directement cette région. Le plan de contrôle du MRAP a également des dépendances sous-jacentes à l'entrée AGA`us-west-2`, à l'entrée Route 53 et à `us-east-1` l'ACM dans chaque région à partir de laquelle le MRAP est configuré pour diffuser du contenu. Vous ne devez pas dépendre de la disponibilité du plan de contrôle MRAP dans votre chemin de restauration ou dans les plans de données de vos propres systèmes. Cela se distingue des contrôles de basculement du MRAP qui sont utilisés pour spécifier l'état de routage actif ou passif pour chacun de vos compartiments dans le MRAP. Ces API sont hébergées en cinq Régions AWS et peuvent être utilisées pour transférer efficacement le trafic à l'aide du plan de données du service.

En outre, les noms des compartiments Amazon S3 sont uniques au niveau mondial `CreateBucket` et tous les appels aux `DeleteBucket` API et dépendentus`-east-1`, dans la `aws` partition, de l'unicité des noms, même si l'appel d'API est dirigé vers la région spécifique dans laquelle vous souhaitez créer le compartiment. Enfin, si vos flux de travail de création de compartiments sont essentiels, vous ne devez pas vous fier à la disponibilité d'une orthographe spécifique pour le nom d'un bucket, en particulier ceux qui suivent un schéma perceptible.

> **Recommandation**
>
> Ne vous fiez pas à la suppression ou à la création de nouveaux compartiments S3 ou à la mise à jour de configurations de compartiments S3 dans le cadre de votre processus de restauration. Préapprovisionnez tous les compartiments S3 requis avec les configurations nécessaires afin de ne pas avoir à apporter de modifications pour récupérer après une panne. Cette approche s'applique également aux MRAP.

- CloudFront

  Amazon API Gateway fournit des points de terminaison d'API optimisés pour les périphériques. La création de ces points de terminaison dépend du plan CloudFront de contrôle utilisé `us-east-1` pour créer la distribution devant le point de terminaison de la passerelle.

  > **Recommandation**
  >
  > Ne vous fiez pas à la création de nouveaux points de terminaison API Gateway optimisés pour les périphériques dans le cadre de votre processus de reprise. Préapprovisionnez tous les points de terminaison API Gateway requis.

  Toutes les dépendances abordées dans cette section sont des actions du plan de contrôle, et non des actions du plan de données. Si vos charges de travail sont configurées pour être statiquement stables, ces dépendances ne devraient pas avoir d'impact sur votre chemin de restauration, en gardant à l'esprit que la stabilité statique nécessite du travail ou des services supplémentaires à mettre en œuvre.

Services utilisant des points de terminaison globaux par défaut

Dans certains cas, les AWS services fournissent un point de terminaison global par défaut, tel que le AWS Security Token Service (AWS STS). D'autres services peuvent utiliser ce point de terminaison global par défaut dans leur configuration par défaut. Cela signifie qu'un service régional que vous utilisez peut avoir une dépendance globale à l'égard d'un service unique Région AWS. Les informations suivantes expliquent comment supprimer les dépendances involontaires sur les points de terminaison globaux par défaut afin de vous aider à utiliser le service de manière régionale.

AWS STS : STS est un service Web qui vous permet de demander des informations d'identification temporaires à privilèges limités pour les utilisateurs IAM ou pour les utilisateurs que vous authentifiez (utilisateurs fédérés). L'utilisation du STS à partir du kit de développement AWS logiciel (SDK) et de l'interface de ligne de commande (CLI) est définie par défaut sur. `us-east-1` Le service STS fournit également des points de terminaison régionaux. Ces points de terminaison sont activés par défaut dans les régions qui le sont également par défaut. Vous pouvez en tirer parti à tout moment en configurant votre SDK ou votre CLI en suivant les instructions suivantes : Points de terminaison régionaux AWS STS. L'utilisation de SigV4a nécessite également des informations d'identification

temporaires demandées à un point de terminaison STS régional. Vous ne pouvez pas utiliser le point de terminaison STS global pour cette opération.

> **Recommandation**
>
> Mettez à jour la configuration de votre SDK et de votre CLI pour utiliser les points de terminaison STS régionaux.

Connexion au langage SAML (Security Assertion Markup Language) : les services SAML existent partout. Régions AWS Pour utiliser ce service, choisissez le point de terminaison SAML régional approprié, tel que https://us-west-2.signin.aws.amazon.com/saml. Vous devez mettre à jour les configurations de vos politiques de confiance et de votre fournisseur d'identité (IdP) pour utiliser les points de terminaison régionaux. Reportez-vous à la documentation AWS SAML pour plus de détails.

Si vous utilisez un IdP qui est également hébergé sur AWS, il existe un risque qu'il soit également impacté en cas de AWS panne. Cela peut vous empêcher de mettre à jour la configuration de votre IdP ou de vous opposer à une fédération complète. Vous devez pré-approvisionner les utilisateurs « break-glass » au cas où votre IdP serait défaillant ou indisponible. Reportez-vous à Annexe A - Directives relatives aux services partitionnés pour plus de détails sur la façon de créer des utilisateurs break-glass de manière statiquement stable.

> **Recommandation**
>
> Mettez à jour vos politiques de confiance en matière de rôles IAM pour accepter les connexions SAML provenant de plusieurs régions. En cas de panne, mettez à jour la configuration de votre IdP pour utiliser un point de terminaison SAML régional différent si votre point de terminaison préféré est altéré. Créez un ou plusieurs utilisateurs au cas où votre IdP serait défaillant ou indisponible.

AWS IAM Identity Center : Identity Center est un service basé sur le cloud qui facilite la gestion centralisée de l'accès par authentification unique aux applications cloud Comptes AWS et aux applications du client. Identity Center doit être déployé dans une seule région de votre choix. Toutefois, le comportement par défaut du service consiste à utiliser le point de terminaison SAML global (https://signin.aws.amazon.com/saml), qui est hébergé dans`us-east-1`. Si vous avez déployé Identity Center dans un autre Région AWS, vous devez mettre à jour l'URL relaystate de chaque ensemble d'autorisations afin de cibler le même point de terminaison de console régional que votre

déploiement d'Identity Center. Par exemple, si vous avez déployé Identity Center dans`us-west-2`, vous devez mettre à jour l'état de relais de vos ensembles d'autorisations pour utiliser https://us-west-2.console.aws.amazon.com. Cela supprimera toute dépendance `us-east-1` à l'égard de votre déploiement d'Identity Center.

En outre, étant donné qu'IAM Identity Center ne peut être déployé que dans une seule région, vous devez pré-approvisionner les utilisateurs « révolutionnaires » au cas où votre déploiement serait perturbé. Reportez-vous à Annexe A - Directives relatives aux services partitionnés pour plus de détails sur la façon de créer des utilisateurs break-glass de manière statiquement stable.

> Recommandation
>
> Définissez l'URL relaystate de vos ensembles d'autorisations dans IAM Identity Center pour qu'elle corresponde à la région dans laquelle le service est déployé. Créez un ou plusieurs utilisateurs exceptionnels au cas où le déploiement de votre IAM Identity Center ne serait pas disponible.

Amazon S3 Storage Lens : Storage Lens fournit un tableau de bord par défaut appelé default-account-dashboard. La configuration du tableau de bord et ses métriques associées sont stockées dans`us-east-1`. Vous pouvez créer des tableaux de bord supplémentaires dans d'autres régions en spécifiant la région d'origine pour la configuration du tableau de bord et les données métriques.

> Recommandation
>
> Si vous avez besoin de données provenant du tableau de bord par défaut de S3 Storage Lens lors d'une panne affectant le service`us-east-1`, créez un tableau de bord supplémentaire dans une autre région d'origine. Vous pouvez également dupliquer tout autre tableau de bord personnalisé que vous avez créé dans d'autres régions.

Résumé des services mondiaux

Les plans de données pour les services mondiaux appliquent des principes d'isolation et d'indépendance similaires à ceux AWS des services régionaux. Une panne affectant le plan de données IAM dans une région n'affecte pas le fonctionnement du plan de données IAM dans une autre. Région AWS De même, une panne affectant le plan de données de la Route 53 dans un PoP n'affecte pas le fonctionnement du plan de données Route 53 dans le reste du PoPs. Par

conséquent, nous devons prendre en compte les événements de disponibilité du service qui affectent la région dans laquelle le plan de contrôle opère ou qui affectent le plan de contrôle lui-même. Comme il n'existe qu'un seul plan de contrôle pour chaque service global, une défaillance affectant ce plan de contrôle peut avoir des effets interrégionaux sur les opérations de type CRUDL (qui sont les opérations de configuration généralement utilisées pour installer ou configurer un service par opposition à l'utilisation directe du service).

La méthode la plus efficace pour concevoir des charges de travail afin d'utiliser les services globaux de manière résiliente consiste à utiliser la stabilité statique. Lors d'un scénario de panne, concevez votre charge de travail de manière à ne pas avoir à apporter de modifications à un plan de contrôle pour atténuer l'impact ou à basculer vers un autre emplacement. Consultez Annexe A - Directives relatives aux services partitionnés et obtenez Annexe B - Guide de service global pour les réseaux Edge des conseils prescriptifs sur la manière d'utiliser ces types de services globaux afin de supprimer les dépendances au plan de contrôle et d'éliminer les points de défaillance uniques. Si vous avez besoin des données d'une opération du plan de contrôle à des fins de restauration, mettez ces données en cache dans un magasin de données accessible via son plan de données, tel qu'un paramètre du magasin de paramètres AWS Systems Manager (SSM Parameter Store), une table DynamoDB ou un compartiment S3. Pour des raisons de redondance, vous pouvez également choisir de stocker ces données dans une région supplémentaire. Par exemple, conformément aux meilleures pratiques pour Route 53 Application Recovery Controller (ARC), vous devez coder en dur ou ajouter vos cinq points de terminaison de cluster régional à vos favoris. En cas de panne, il se peut que vous ne puissiez pas accéder à certaines opérations d'API, notamment les opérations d'API Route 53 ARC qui ne sont pas hébergées sur le cluster de plans de données extrêmement fiable. Vous pouvez répertorier les points de terminaison de vos clusters Route 53 ARC à l'aide de l'opération `DescribeCluster` API.

Voici un résumé de certaines des erreurs de configuration ou des anti-modèles les plus courants qui introduisent des dépendances sur les plans de contrôle des services globaux :

- Apporter des modifications aux enregistrements Route 53, par exemple en mettant à jour la valeur d'un enregistrement A ou en modifiant les poids d'un ensemble d'enregistrements pondéré, pour effectuer un basculement.
- Création ou mise à jour de ressources IAM, notamment de rôles et de politiques IAM, lors d'un basculement. Ce n'est généralement pas intentionnel, mais cela peut être le résultat d'un plan de basculement non testé.
- S'appuyer sur IAM Identity Center pour permettre aux opérateurs d'accéder aux environnements de production en cas de panne.

- En vous basant sur la configuration par défaut d'IAM Identity Center pour utiliser la console `us-east-1` lorsque vous avez déployé Identity Center dans une autre région.
- Modification de la pondération des numéros de trafic AGA pour effectuer manuellement un basculement régional.
- Mettre à jour la configuration d'origine d'une CloudFront distribution pour qu'elle échoue à éliminer une origine altérée.
- Provisionner des ressources de reprise après sinistre (DR), telles que les ELB et les instances RDS en cas de panne, qui dépendent de la création d'enregistrements DNS dans Route 53.

Ce qui suit est un résumé des recommandations fournies dans cette section pour utiliser les services mondiaux de manière résiliente afin de prévenir les anciens modèles anti-modèles courants.

Résumé des recommandations

Ne vous fiez pas aux plans de contrôle des services partitionnels dans votre processus de restauration. Fiez-vous plutôt aux opérations du plan de données de ces services. Consultez Annexe A - Directives relatives aux services partitionnés pour plus de détails sur la manière dont vous devez concevoir pour les services partitionnels.
Ne vous fiez pas au plan de contrôle des services du réseau périphérique dans votre processus de restauration. Fiez-vous plutôt aux opérations du plan de données de ces services. Consultez Annexe B - Guide de service global pour les réseaux Edge pour plus de détails sur la façon de concevoir des services mondiaux dans le réseau de périphérie.
Ne vous fiez pas à la création, à la mise à jour ou à la suppression de ressources qui nécessitent la création, la mise à jour ou la suppression d'enregistrements de ressources Route 53, de zones hébergées ou de bilans de santé dans votre parcours de restauration. Préapprovisionnez ces ressources, telles que les ELB, pour éviter de dépendre du plan de contrôle Route 53 dans votre parcours de reprise.
Ne vous fiez pas à la suppression ou à la création de nouveaux compartiments S3 ou à la mise à jour de configurations de compartiments S3 dans le cadre de votre processus de restauration. Préapprovisionnez tous les compartiments S3 requis avec les configurations nécessaires afin de ne pas avoir à apporter de modifications pour récupérer après une panne. Cette approche s'applique également aux MRAP.
Ne vous fiez pas à la création de nouveaux points de terminaison API Gateway optimisés pour les périphériques dans le cadre de votre processus de reprise. Préapprovisionnez tous les points de terminaison API Gateway requis.

Mettez à jour la configuration de votre SDK et de votre CLI pour utiliser les points de terminaison STS régionaux.
Mettez à jour vos politiques de confiance en matière de rôles IAM pour accepter les connexions SAML provenant de plusieurs régions. En cas de panne, mettez à jour la configuration de votre IdP pour utiliser un point de terminaison SAML régional différent si votre point de terminaison préféré est altéré. Créez des utilisateurs Break Glass au cas où votre IdP serait défaillant ou indisponible.
Définissez l'URL relaystate de vos ensembles d'autorisations dans IAM Identity Center pour qu'elle corresponde à la région dans laquelle le service est déployé. Créez un ou plusieurs utilisateurs exceptionnels au cas où le déploiement de votre Identity Center ne serait pas disponible.
Si vous avez besoin de données provenant du tableau de bord par défaut de S3 Storage Lens lors d'une panne affectant le service`us-east-1`, créez un tableau de bord supplémentaire dans une autre région d'origine. Vous pouvez également dupliquer tout autre tableau de bord personnalisé que vous avez créé dans d'autres régions.

Conclusion

AWSfournit plusieurs structures différentes pour les limites d'isolation des pannes. Vous devez réfléchir à la manière dont vous concevez les services zonaux, régionaux et mondiaux, ainsi qu'aux impacts potentiels sur votre charge de travail et sur la capacité de votre charge de travail à se rétablir en cas de défaillance du plan de contrôle. La stabilité statique est l'un des principaux moyens d'éviter les dépendances du plan de contrôle et de créer des mécanismes HA et DR fiables et résilients lorsque vous utilisez AWS des services.

Annexe A - Directives relatives aux services partitionnés

Pour les services partitionnels, vous devez implémenter la stabilité statique afin de maintenir la résilience de votre charge de travail en cas de défaillance du plan AWS de contrôle des services. Ce qui suit fournit des conseils prescriptifs sur la manière de prendre en compte les dépendances à l'égard des services partitionnés ainsi que sur ce qui fonctionnera et ne fonctionnera pas en cas de déficience du plan de contrôle.

AWS Identity and Access Management (IAM)

Le plan de contrôle AWS Identity and Access Management (IAM) comprend toutes les API IAM publiques (y compris Access Advisor mais pas Access Analyzer ou IAM Roles Anywhere). Cela inclut des actions telles que `CreateRole` `AttachRolePolicy` `ChangePassword`,`UpdateSAMLProvider`, et`UpdateLoginProfile`. Le plan de données IAM fournit une authentification et une autorisation aux principaux IAM de chacun d'entre eux. Région AWS Lors d'une panne du plan de contrôle, les opérations de type CRUDL pour IAM peuvent ne pas fonctionner, mais l'authentification et l'autorisation pour les mandants existants continueront de fonctionner. STS est un service réservé au plan de données qui est distinct de l'IAM et ne dépend pas du plan de contrôle IAM.

Cela signifie que lorsque vous planifiez des dépendances à l'égard de l'IAM, vous ne devez pas vous fier au plan de contrôle IAM pour votre chemin de restauration. Par exemple, une conception statiquement stable pour un utilisateur administrateur « révolutionnaire » consisterait à créer un utilisateur doté des autorisations appropriées, à définir le mot de passe et à fournir la clé d'accès et la clé d'accès secrète, puis à verrouiller ces informations d'identification dans un coffre-fort physique ou virtuel. Lorsque cela est nécessaire en cas d'urgence, récupérez les informations d'identification de l'utilisateur dans le coffre-fort et utilisez-les selon vos besoins. Une non-statically-stable solution consisterait à approvisionner l'utilisateur en cas de panne ou à le préapprovisionner, tout en joignant la politique d'administration uniquement lorsque cela est nécessaire. Ces approches dépendraient du plan de contrôle IAM.

AWS Organizations

Le plan de AWS Organizations contrôle comprend toutes les API Organizations publiques telles que `AcceptHandshake` `AttachPolicy` `CreateAccount`,`CreatePolicy`, et`ListAccounts`. Il n'existe pas de plan de données pourAWS Organizations. Il orchestre le plan de données pour d'autres services tels que IAM. Lors d'une défaillance du plan de contrôle, les opérations

de type CRUDL pour Organizations peuvent ne pas fonctionner, mais les politiques, telles que les politiques de contrôle des services (SCP) et les politiques de balisage, continueront de fonctionner et seront évaluées dans le cadre du processus d'autorisation IAM. Les fonctionnalités d'administration déléguée et les fonctionnalités multicomptes AWS des autres services pris en charge par Organizations continueront également de fonctionner.

Cela signifie que lorsque vous planifiez des dépendances surAWS Organizations, vous ne devez pas vous fier au plan de contrôle de l'Organizations pour votre processus de reprise. Mettez plutôt en œuvre la stabilité statique dans votre plan de restauration. Par exemple, une non-statically-stable approche peut consister à mettre à jour les SCP afin de supprimer les restrictions relatives à l'autorisation Régions AWS via la `aws:RequestedRegion` condition, ou à activer les autorisations d'administrateur pour des rôles IAM spécifiques. Cela dépend du plan de contrôle de l'Organizations pour effectuer ces mises à jour. Une meilleure approche serait d'utiliser des balises de session pour accorder l'utilisation des autorisations d'administrateur. Votre fournisseur d'identité (IdP) peut inclure des balises de session qui peuvent être évaluées en fonction de cette `aws:PrincipalTag` condition, ce qui vous permet de configurer de manière dynamique les autorisations pour certains mandants tout en aidant vos SCP à rester statiques. Cela supprime les dépendances par rapport aux plans de contrôle et utilise uniquement les actions du plan de données.

Gestion de compte AWS

Le plan de contrôle de la gestion des AWS comptes est hébergé dans us-east-1 et comprend toutes les API publiques permettant de gérer unCompte AWS, telles que et. `GetContactInformation` `PutContactInformation` Cela inclut également la création ou la fermeture d'un nouveau Compte AWS via la console de gestion. Les API pour`CloseAccount`, `CreateAccountCreateGovCloudAccount`, et `DescribeAccount` font partie du plan de AWS Organizations contrôle, qui est également hébergé dans us-east-1. De plus, la création d'un GovCloud compte en dehors de AWS Organizations repose sur le plan de contrôle Compte AWS de gestion dans us-east-1. De plus, GovCloud les comptes doivent être liés de manière 1:1 à un Compte AWS dans la `aws` partition. La création de comptes dans la `aws-cn` partition ne repose pas sur us-east-1. Le plan de données Comptes AWS concerne les comptes eux-mêmes. Lors d'une panne du plan de contrôle, les opérations de type CRUDL (comme la création d'un nouveau compte ou l'obtention et la mise à jour des informations de contact) Comptes AWS peuvent ne pas fonctionner. Les références au compte dans les politiques IAM continueront de fonctionner.

Cela signifie que lorsque vous planifiez des dépendances à l'égard de la gestion des AWS comptes, vous ne devez pas vous fier au plan de contrôle de la gestion des comptes pour votre processus

de restauration. Bien que le plan de contrôle de la gestion des comptes ne fournisse pas les fonctionnalités directes que vous utiliseriez habituellement dans une situation de restauration, il se peut que vous le fassiez parfois. Par exemple, une conception statiquement stable consisterait à préprovisionner tout ce dont Comptes AWS vous avez besoin pour le basculement. Une solution non-statically-stable consisterait à en créer de nouvelles Comptes AWS lors d'un événement de panne pour héberger vos ressources de reprise après sinistre.

Application Recovery Controller Amazon Route 53

Le plan de contrôle de Route 53 ARC comprend les API nécessaires au contrôle et à la préparation de la restauration, comme indiqué sur : Points de terminaison et quotas d'Amazon Route 53 Application Recovery Controller. Vous gérez les vérifications de disponibilité, les contrôles de routage et les opérations du cluster à l'aide du plan de contrôle. Le plan de données d'ARC est votre cluster de restauration, qui gère les valeurs de contrôle de routage demandées par les contrôles de santé de Route 53 et met également en œuvre les règles de sécurité. La fonctionnalité du plan de données de Route 53 ARC est accessible via les API de votre cluster de restauration, telles que`https://aaaaaaaa.route53-recovery-cluster.eu-west-1.amazonaws.com`.

Cela signifie que vous ne devez pas vous fier au plan de contrôle ARC de la Route 53 pour vous rétablir. Deux bonnes pratiques permettent de mettre en œuvre ce guide :

- Commencez par ajouter à vos favoris ou codez en dur les cinq points de terminaison du cluster régional. Cela évite d'avoir à utiliser l'opération du plan DescribeCluster de contrôle lors d'un scénario de basculement pour découvrir les valeurs des points de terminaison.
- Ensuite, utilisez les API du cluster Route 53 ARC en utilisant l'interface de ligne de commande ou le SDK pour mettre à jour les contrôles de routage et non les. AWS Management Console Cela supprime la console de gestion en tant que dépendance de votre plan de basculement et garantit qu'il dépend uniquement des actions du plan de données.

AWS Network Manager

Le service AWS Network Manager est principalement un système réservé au plan de contrôle hébergé sur us-west-2. Son objectif est de gérer de manière centralisée la configuration de votre AWS Cloud Réseau central WAN et votre Réseau AWS Transit Gateway à travers Comptes AWS les régions et emplacements sur site. Il regroupe également les métriques de votre cloud WAN dans us-west-2, qui est également accessible via le plan de données. CloudWatch Si Network Manager est défaillant, le plan de données des services qu'il orchestre ne sera pas affecté. Les CloudWatch

métriques relatives au Cloud WAN sont également disponibles dans us-west-2. Si vous souhaitez obtenir des données statistiques historiques, telles que le nombre d'octets entrants et sortants par région, afin de comprendre le volume de trafic susceptible d'être transféré vers d'autres régions en cas de panne affectant us-west-2, ou à d'autres fins opérationnelles, vous pouvez exporter ces mesures sous forme de données CSV directement depuis la CloudWatch console ou en utilisant cette méthode : Publier les CloudWatch métriques Amazon dans un fichier CSV. Les données se trouvent sous l'`AWS/Network Manager`espace de noms et vous pouvez effectuer cette opération selon un calendrier de votre choix et les stocker dans S3 ou dans un autre magasin de données que vous sélectionnez. Pour mettre en œuvre un plan de restauration statiquement stable, n'utilisez pas AWS Network Manager pour mettre à jour votre réseau et ne vous fiez pas aux données issues des opérations de son plan de contrôle pour la saisie du basculement.

DNS privé Route 53

Les zones hébergées privées de Route 53 sont prises en charge dans chaque partition ; toutefois, les considérations relatives aux zones hébergées privées et aux zones hébergées publiques dans Route 53 sont les mêmes. Reportez-vous à Amazon Route 53 dans l'Annexe B - Guide de service global relatif au réseau Edge.

Annexe B - Guide de service global pour les réseaux Edge

Pour les services globaux du réseau Edge, vous devez implémenter la stabilité statique afin de maintenir la résilience de votre charge de travail en cas de défaillance du plan de contrôle des AWS services.

Route 53

Le plan de contrôle de Route 53 comprend toutes les API publiques de Route 53 couvrant les fonctionnalités relatives aux zones hébergées, aux enregistrements, aux contrôles de santé, aux journaux de requêtes DNS, aux ensembles de délégations réutilisables, aux politiques de trafic et aux balises de répartition des coûts. Il est hébergé sur l'us-east-1. Le plan de données est le service DNS faisant autorité qui s'exécute sur plus de 200 emplacements PdPRégion AWS, répondant aux requêtes DNS en fonction de vos zones hébergées et vos données de surveillance de l'état. En outre, Route 53 dispose d'un plan de données pour les bilans de santé, qui est également un service distribué dans le monde entier sur plusieurs sites. Régions AWS Ce plan de données effectue les surveillances de l'état, agrège les résultats et les transmet aux plans de données du DNS public et privé de Route 53 et à. Lors d'une défaillance du plan de contrôle, les opérations de type CRUDL pour Route 53 peuvent ne pas fonctionner, mais la résolution et les contrôles d'intégrité du DNS, ainsi que les mises à jour du routage résultant de modifications des contrôles de santé, continueront de fonctionner.

Cela signifie que lorsque vous planifiez des dépendances sur Route 53, vous ne devez pas vous fier au plan de contrôle de Route 53 pour votre chemin de restauration. Par exemple, une conception statiquement stable consisterait à utiliser l'état des contrôles de santé pour effectuer des basculements entre régions ou pour évacuer une zone de disponibilité. Vous pouvez utiliser les contrôles de routage ARC (Application Recovery Controller) Route 53 pour modifier manuellement l'état des vérifications de santé et modifier les réponses aux requêtes DNS. Il existe des modèles similaires à ceux fournis par ARC que vous pouvez implémenter en fonction de vos besoins. Certains de ces modèles sont décrits dans la section Création de mécanismes de reprise après sinistre à l'aide de Route 53 et dans la section consacrée au contrôle de l'état des disjoncteurs sur les modèles de résilience multi-AZ avancés. Si vous avez choisi d'utiliser un plan DR multirégional, préprovisionnez les ressources qui nécessitent la création d'enregistrements DNS, telles que les ELB et les instances RDS. Une non-statically-stable solution consisterait à mettre à jour la valeur d'un enregistrement de ressource Route 53 via l'`ChangeResourceRecordSets`API, à modifier

le poids d'un enregistrement pondéré ou à créer de nouveaux enregistrements pour effectuer un basculement. Ces approches dépendent du plan de contrôle de la Route 53.

Amazon CloudFront

Le plan de CloudFront contrôle Amazon comprend toutes les CloudFront API publiques permettant de gérer les distributions et est hébergé sur us-east-1. Le plan de données est la distribution elle-même desservie depuis PoPs le réseau périphérique. Il gère les requêtes, assure le routage et la mise en cache de votre contenu d'origine. Lors d'une altération du plan de contrôle, les opérations de type CRUDL CloudFront (y compris les demandes d'invalidation) peuvent ne pas fonctionner, mais votre contenu continuera d'être mis en cache et diffusé, et les basculements d'origine continueront de fonctionner.

Cela signifie que lorsque vous planifiez des dépendances surCloudFront, vous ne devez pas vous fier au plan de CloudFront contrôle pour votre chemin de restauration. Par exemple, une conception statiquement stable consisterait à utiliser des basculements d'origine automatisés pour atténuer l'impact d'une panne sur l'une de vos origines. Vous pouvez également choisir de créer un équilibrage de charge d'origine ou un basculement à l'aide de Lamda @Edge. Reportez-vous à Trois modèles de conception avancés pour les applications à haut niveau de disponibilité à l'aide d'Amazon CloudFront et à Utilisation d'Amazon CloudFront et d'Amazon S3 pour créer des applications de géo-proximité actives multirégions pour plus de détails sur ce modèle. Une non-statically-stable solution consisterait à mettre à jour manuellement la configuration de votre distribution en réponse à une défaillance d'origine. Cette approche dépendrait du plan CloudFront de contrôle.

Certificate Manager

Si vous utilisez des certificats personnalisés avec votre CloudFront distribution, vous êtes également dépendant d'ACM. L'utilisation de certificats personnalisés avec votre distribution d'CloudFront dépend du plan de contrôle ACM dans la région us-east-1. En cas de panne du plan de contrôle, vos certificats existants configurés dans votre distribution continueront de fonctionner, de même que les renouvellements automatiques des certificats. Ne vous fiez pas à la modification de la configuration de la distribution ou à la création de nouveaux certificats dans le cadre de votre chemin de restauration.

AWSPare-feu d'applications Web (WAF) et WAF Classic

Si vous l'utilisez AWS WAF avec votre CloudFront distribution, vous êtes dépendant du plan de contrôle WAF, qui est également hébergé dans la région us-east-1. Lors d'une défaillance du plan de

contrôle, les listes de contrôle d'accès Web (ACL) configurées et leurs règles associées continuent de fonctionner. Ne vous fiez pas à la mise à jour de vos listes de contrôle d'accès Web WAF dans le cadre de votre processus de restauration.

AWS Global Accelerator

Le plan de contrôle AGA comprend toutes les API AGA publiques et est hébergé sur us-west-2. Le plan de données est le routage réseau des adresses IP anycast fournies par AGA vers vos points de terminaison enregistrés. AGA utilise également les bilans de santé de Route 53 pour déterminer l'état de santé de vos points de terminaison AGA, qui font partie du plan de données Route 53. En cas de panne d'un avion de contrôle, les opérations de type CRUDL pour l'AGA peuvent ne pas fonctionner. Le routage vers vos points de terminaison existants, ainsi que les bilans de santé, les numéros de trafic et les configurations de pondération des points de terminaison utilisées pour acheminer ou transférer le trafic vers d'autres points de terminaison et groupes de points de terminaison, continueront de fonctionner.

Cela signifie que lorsque vous planifiez des dépendances à l'égard de l'AGA, vous ne devez pas vous fier au plan de contrôle AGA pour votre chemin de restauration. Par exemple, une conception statiquement stable consisterait à utiliser l'état des contrôles de santé configurés pour éliminer les points de terminaison défectueux. Reportez-vous à la section Déploiement d'applications multirégionales à AWS l'aide de AWS Global Accelerator pour des exemples de cette configuration. Une non-statically-stable conception consisterait à modifier les pourcentages de numérotation du trafic AGA, à modifier les groupes de points de terminaison ou à supprimer un point de terminaison d'un groupe de points de terminaison en cas de déficience. Ces approches dépendraient du plan de contrôle AGA.

Shield

Le plan de contrôle Amazon Shield Advanced comprend toutes les API publiques de Shield Advanced et est hébergé sur us-east-1. Cela inclut des fonctionnalités telles que `CreateProtection` `CreateProtectionGroupAssociateHealthCheck`,`DesribeDRTAccess`, et`ListProtections`. Le plan de données est la protection DDoS fournie par Shield Advanced ainsi que la création des métriques Shield Advanced. Shield Advanced utilise également les contrôles de santé de Route 53 (qui font partie du plan de données Route 53), si vous les avez configurés. Lors d'une défaillance du plan de contrôle, les opérations de type CRUDL pour Shield Advanced peuvent ne pas fonctionner, mais la protection DDoS configurée pour vos ressources, ainsi que les réponses aux modifications des bilans de santé, continueront de fonctionner.

Cela signifie que vous ne devez pas vous fier au plan de contrôle Shield Advanced pour vous rétablir. Bien que le plan de contrôle Shield Advanced ne fournisse pas les fonctionnalités directes que vous utiliseriez habituellement dans une situation de restauration, il se peut que vous le fassiez parfois. Par exemple, une conception statiquement stable consisterait à configurer vos ressources DR de manière à ce qu'elles fassent partie d'un groupe de protection et soient associées à des contrôles de santé, au lieu de configurer cette protection après la survenue de la panne. Cela évite de dépendre du plan de contrôle Shield Advanced pour la restauration.

Annexe C - Services pour une seule région

Vous trouverez ci-dessous une liste des services, ou des fonctionnalités spécifiques de ce service (qui sont répertoriées entre parenthèses après le nom du service), qui ne sont disponibles que dans une seule région. Les instructions relatives à la mise en œuvre de la stabilité statique fournies pour d'autres services globaux s'appliquent à ces services lorsque vous devez planifier les dépendances sur leurs plans de contrôle et leurs plans de données.

- Alexa for Business
- AWS Marketplace(API de AWS Marketplace catalogue, analyse AWS Marketplace du commerce, AWS Marketplace service d'autorisation)
- Billing and Cost Management (AWS Cost Explorerrapports sur les AWS coûts et l'utilisation, AWS budgets, Savings Plans)
- AWS BugBust
- Amazon Mechanical Turk
- Amazon Chime
- SDK Amazon Chime (audio PSTN, messagerie, identité)
- AWSChatbot
- AWS DeepRacer
- AWSDevice Farm
- Amazon GameSparks
- Amazon Honeycode

Collaborateurs

Les contributeurs à ce document incluent :

- Michael Haken, architecte de solutions principal, Amazon Web Services

Révisions du document

Pour recevoir les notifications des mises à jour de ce livre blanc, abonnez-vous au flux RSS.

Modification	Description	Date
Révision mineure	Guide mis à jour pour s'aligner sur les bonnes pratiques IAM. Pour de plus amples informations, veuillez consulter Bonnes pratiques de sécurité dans IAM.	9 février 2023
Publication initiale	Livre blanc publié.	16 novembre 2022

Glossaire AWS

Pour connaître la terminologie la plus récente d’AWS, consultez le Glossaire AWS dans la Référence Glossaire AWS.

Avis

Les clients sont tenu de réaliser leur propre évaluation indépendante des informations contenues dans ce document. Ce document : (a) est fourni à titre informatif uniquement, (b) représente les offres de AWS produits et les pratiques actuelles, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ni aucune garantie de la part AWS de ses filiales, fournisseurs ou concédants de licence. AWSles produits ou services sont fournis « tels quels » sans garantie, représentation ou condition d'aucune sorte, qu'elle soit expresse ou implicite. Les responsabilités et obligations AWS de ses clients sont régies par AWS des accords, et le présent document ne fait partie d'aucun accord entre AWS et ses clients et ne le modifie pas.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.