AWS Framework Well-Architected

Reprise après sinistre des charges de travail sur AWS : restauration dans le cloud

Reprise après sinistre des charges de travail sur AWS : restauration dans le cloud: AWS Framework Well-Architected

Table of Contents

Reprise après sinistre des charges de travail sur AWS : restauration dans le cloud

Date de publication : 12 février 2021 (Historique de la documentation)

La reprise après sinistre est le processus de préparation et de reprise après un sinistre. Un événement qui empêche une charge de travail ou un système d'atteindre ses objectifs commerciaux sur son site de déploiement principal est considéré comme un sinistre. Ce paper décrit les meilleures pratiques pour planifier et tester la reprise après sinistre pour toute charge de travail déployée AWS, et propose différentes approches pour atténuer les risques et atteindre les objectifs de temps de restauration (RTO) et de point de restauration (RPO) pour cette charge de travail.

Ce livre blanc explique comment implémenter la reprise après sinistre pour les charges de travail sur. AWS Reportez-vous à la section Reprise après sinistre des applications sur site AWS pour obtenir des informations sur l'utilisation en AWS tant que site de reprise après sinistre pour les charges de travail sur site.

Introduction

Votre charge de travail doit exécuter la fonction prévue correctement et de manière cohérente. Pour y parvenir, vous devez concevoir une architecture axée sur la résilience. La résilience est la capacité d'une charge de travail à récupérer après une interruption d'infrastructure, de service ou d'application, à acquérir dynamiquement des ressources informatiques pour répondre à la demande et à atténuer les perturbations, telles que les mauvaises configurations ou les problèmes de réseau transitoires.

La reprise après sinistre (DR) est un élément important de votre stratégie de résilience et concerne la manière dont votre charge de travail réagit en cas de sinistre (un sinistre est un événement qui a de graves répercussions négatives sur votre entreprise). Cette réponse doit être basée sur les objectifs commerciaux de votre organisation, qui spécifient la stratégie de votre charge de travail pour éviter la perte de données, connue sous le nom d'objectif de point de restauration (RPO), et pour réduire les temps d'arrêt lorsque votre charge de travail n'est pas disponible, connue sous le nom d'objectif de temps de restauration (RTO). Vous devez donc implémenter la résilience dans la conception de vos charges de travail dans le cloud afin d'atteindre vos objectifs de restauration (RPO et RTO) en cas de sinistre ponctuel donné. Cette approche aide votre organisation à maintenir la continuité des activités dans le cadre de la planification de la continuité des activités (BCP).

Ce paper explique comment planifier, concevoir et implémenter des architectures AWS répondant aux objectifs de reprise après sinistre de votre entreprise. Les informations partagées ici sont destinées aux personnes occupant des postes technologiques, tels que les directeurs de la technologie (CTOs), les architectes, les développeurs, les membres de l'équipe opérationnelle et les personnes chargées d'évaluer et d'atténuer les risques.

Reprise après sinistre et disponibilité

La reprise après sinistre peut être comparée à la disponibilité, qui constitue un autre élément important de votre stratégie de résilience. Alors que la reprise après sinistre mesure les objectifs pour des événements ponctuels, les objectifs de disponibilité mesurent les valeurs moyennes sur une période donnée.

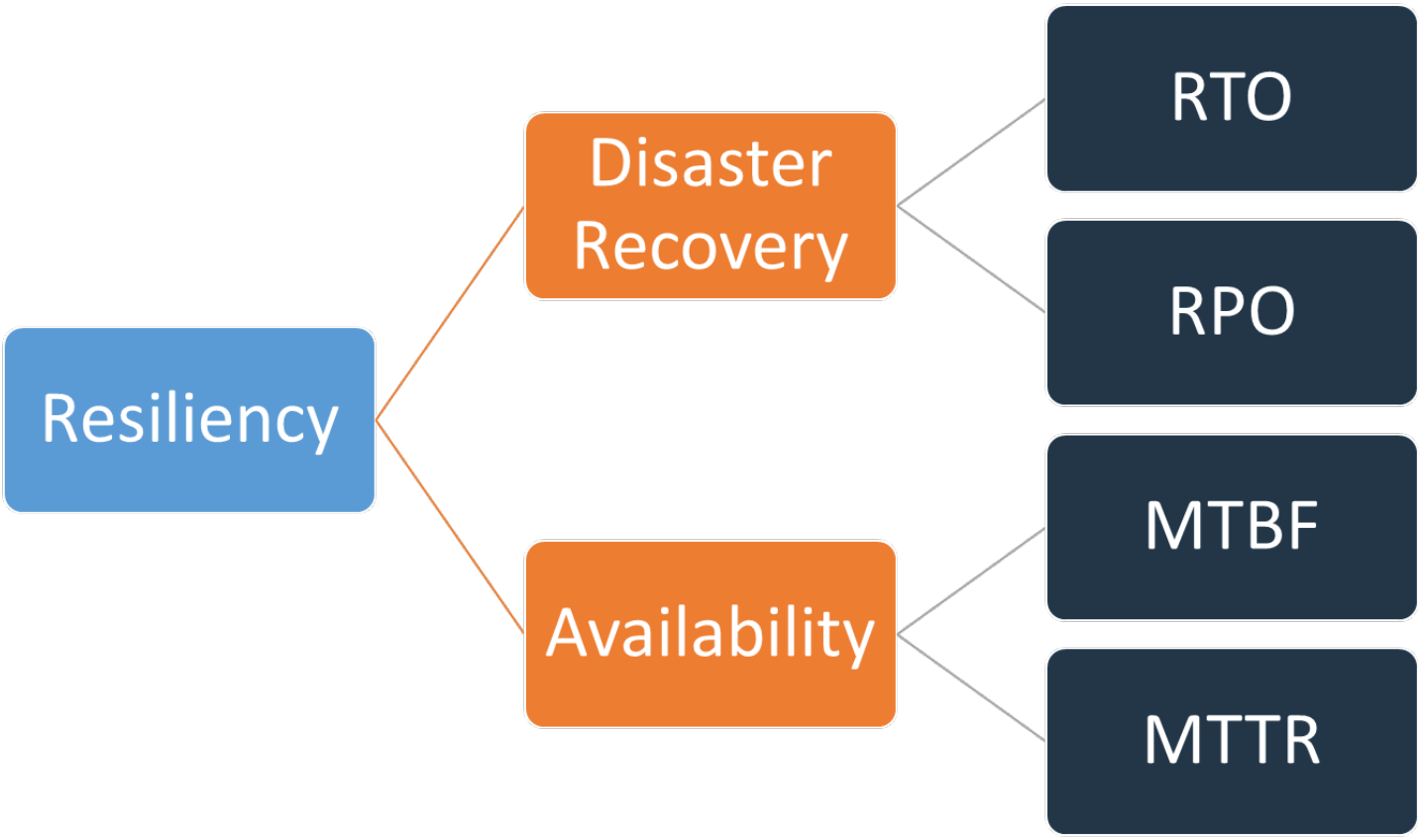

Figure 1 : Objectifs de résilience

La disponibilité est calculée à l'aide du temps moyen entre défaillances (MTBF) et du temps moyen de restauration (MTTR) :

$$Availability = \frac{Available\ for\ Use\ Time}{Total\ Time} = \frac{MTBF}{MTBF + MTTR}$$

Cette approche est souvent appelée « neuf », alors qu'un objectif de disponibilité de 99,9 % est appelé « trois neuf ».

Pour votre charge de travail, il peut être plus facile de compter les demandes réussies et les demandes échouées plutôt que d'utiliser une approche basée sur le temps. Dans ce cas, le calcul suivant peut être utilisé :

$$Availability = \frac{Successful\ Responses}{Valid\ Requests}$$

La reprise après sinistre se concentre sur les catastrophes, tandis que la disponibilité se concentre sur les perturbations les plus courantes de moindre envergure, telles que les défaillances de composants, les problèmes de réseau, les bogues logiciels et les pics de charge. L'objectif de la reprise après sinistre est la continuité des activités, tandis que la disponibilité consiste à optimiser le temps pendant lequel une charge de travail est disponible pour exécuter les fonctionnalités commerciales prévues. Les deux devraient faire partie de votre stratégie de résilience.

Êtes-vous Well-Architected ?

L'AWS Well-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 de l'outil AWS Well-Architected, disponible gratuitement dans l'AWS Management Console, vous pouvez évaluer vos charges de travail par rapport à ces meilleures pratiques en répondant à une série de questions pour chaque pilier.

Les concepts abordés dans ce livre blanc développent les meilleures pratiques contenues dans le livre blanc sur le pilier de fiabilité, en particulier la question REL 13, « Comment planifiez-vous la reprise après sinistre (DR) ? ». Après avoir mis en œuvre les pratiques décrites dans ce livre blanc, assurez-vous de revoir (ou de revoir) votre charge de travail à l'aide de l'outil AWS Well-Architected Tool.

Modèle de responsabilité partagée pour la résilience

La résilience est une responsabilité partagée entre vous AWS et vous, le client. Il est important que vous compreniez comment la reprise après sinistre et la disponibilité, dans le cadre de la résilience, fonctionnent dans le cadre de ce modèle partagé.

Responsabilité d'AWS « Résilience du cloud »

AWS est responsable de la résilience de l'infrastructure qui exécute tous les services proposés dans le cloud AWS. Cette infrastructure comprend le matériel, les logiciels, le réseau et les installations qui exécutent les services cloud AWS. AWS déploie des efforts commercialement raisonnables pour rendre ces services cloud AWS disponibles, en veillant à ce que la disponibilité des services respecte ou dépasse les accords de niveau de service AWS (SLAs).

L'infrastructure cloud mondiale AWS est conçue pour permettre aux clients de créer des architectures de charge de travail hautement résilientes. Chaque région AWS est entièrement isolée et se compose de plusieurs zones de disponibilité, qui sont des partitions d'infrastructure physiquement isolées. Les zones de disponibilité isolent les défaillances susceptibles d'affecter la résilience des charges de travail, en les empêchant d'avoir un impact sur les autres zones de la région. Dans le même temps, toutes les zones d'une région AWS sont interconnectées par un réseau à bande passante élevée et à faible latence, via une fibre métropolitaine dédiée entièrement redondante, fournissant un réseau à haut débit et à faible latence entre les zones. Tout le trafic entre les zones est chiffré. Les performances du réseau sont suffisantes pour réaliser une réplication synchrone entre les zones. Lorsqu'une application est partitionnée AZs, les entreprises sont mieux isolées et protégées contre les problèmes tels que les pannes de courant, les éclairs, les tornades, les ouragans, etc.

Responsabilité du client « Résilience dans le cloud »

Votre responsabilité sera déterminée par les services cloud AWS que vous sélectionnez. Cela détermine la quantité de travail de configuration que vous devez effectuer dans le cadre de vos responsabilités en matière de résilience. Par exemple, un service tel qu'Amazon Elastic Compute Cloud (Amazon EC2) oblige le client à effectuer toutes les tâches de configuration et de gestion de la résilience nécessaires. Les clients qui déploient des EC2 instances Amazon sont chargés de déployer des EC2 instances sur plusieurs sites (tels que les zones de disponibilité AWS), de mettre en œuvre l'autoréparation à l'aide de services tels qu'Amazon EC2 Auto Scaling, ainsi que d'appliquer les meilleures pratiques en matière d'architecture de charge de travail résiliente pour

les applications installées sur les instances. Pour les services gérés, tels qu'Amazon S3 et Amazon DynamoDB, AWS gère la couche d'infrastructure, le système d'exploitation et les plateformes, et les clients accèdent aux points de terminaison pour stocker et récupérer les données. Vous êtes responsable de la gestion de la résilience de vos données, y compris des stratégies de sauvegarde, de gestion des versions et de réplication.

Le déploiement de votre charge de travail sur plusieurs zones de disponibilité d'une région AWS fait partie d'une stratégie de haute disponibilité conçue pour protéger les charges de travail en isolant les problèmes dans une zone de disponibilité et en utilisant la redondance des autres zones de disponibilité pour continuer à traiter les demandes. Une architecture Multi-AZ s'inscrit également dans une stratégie DR conçue pour mieux isoler et protéger les charges de travail contre des problèmes tels que les pannes de courant, la foudre, les tornades, les tremblements de terre, etc. Les stratégies de reprise après sinistre peuvent également utiliser plusieurs régions AWS. Par exemple, dans une configuration active/passive, le service de la charge de travail basculera de sa région active vers sa région DR si la région active ne peut plus traiter les demandes.

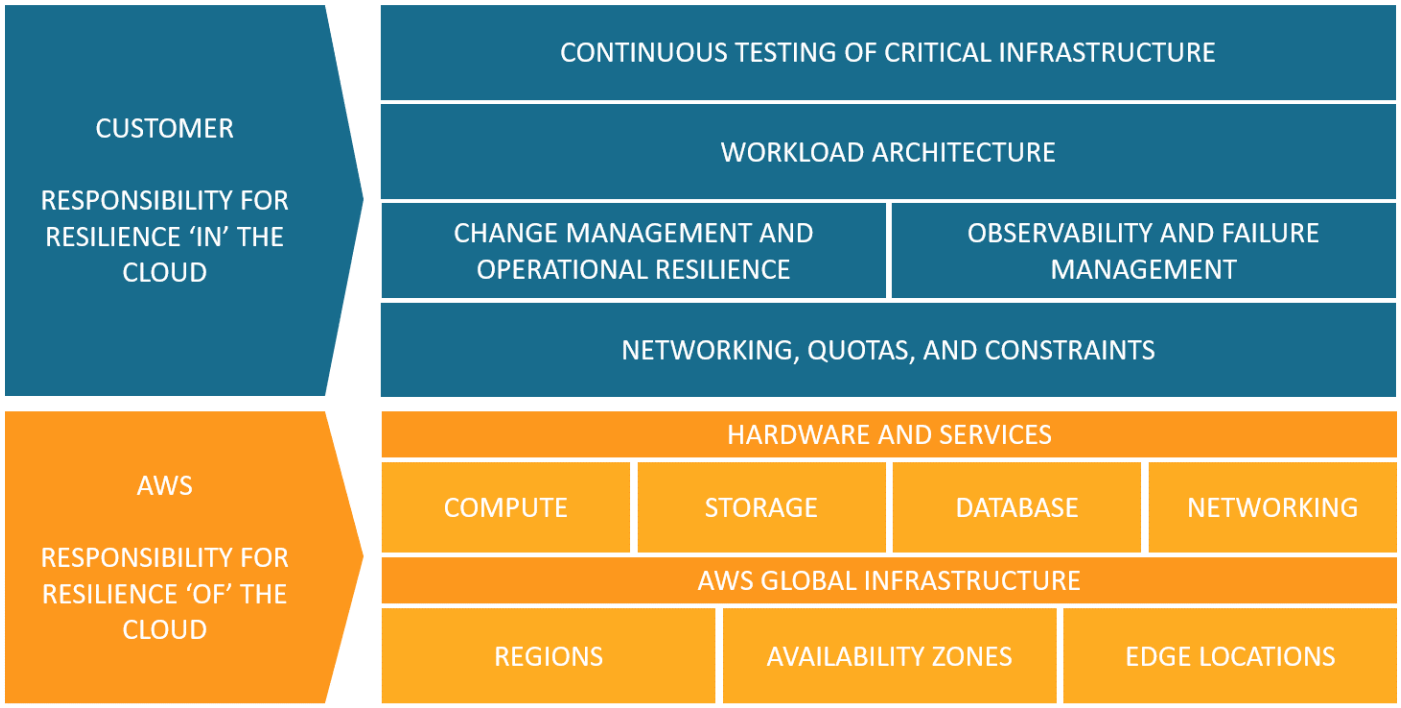

Figure 2 : La résilience est une responsabilité partagée entre AWS et le client

Qu'est-ce qu'une catastrophe ?

Lorsque vous planifiez une reprise après sinistre, évaluez votre plan pour ces trois principales catégories de catastrophes :

- Catastrophes naturelles, telles que les tremblements de terre ou les inondations
- Défaillances techniques, telles qu'une panne de courant ou une connexion réseau
- Actions humaines, telles qu'une mauvaise configuration ou un accès ou une modification par un unauthorized/outside tiers

Chacune de ces catastrophes potentielles aura également un impact géographique qui peut être local, régional, national, continental ou mondial. La nature du sinistre et son impact géographique sont tous deux importants lors de l'élaboration de votre stratégie de reprise après sinistre. Par exemple, vous pouvez atténuer un problème d'inondation local à l'origine d'une panne de centre de données en utilisant une stratégie multi-AZ, car cela n'affecterait qu'une seule zone de disponibilité. Cependant, une attaque contre les données de production vous obligerait à invoquer une stratégie de reprise après sinistre qui bascule pour sauvegarder les données dans une autre région AWS.

La haute disponibilité n'est pas synonyme de reprise après sinistre

La disponibilité et la reprise après sinistre reposent toutes deux sur les mêmes bonnes pratiques, telles que la surveillance des défaillances, le déploiement sur plusieurs sites et le basculement automatique sur incident. Cependant, la disponibilité se concentre sur les composants de la charge de travail, tandis que la reprise après sinistre se concentre sur des copies discrètes de l'ensemble de la charge de travail. La reprise après sinistre a des objectifs différents de ceux de la disponibilité, qui consiste à mesurer le délai de reprise après des événements de grande envergure qualifiés de catastrophes. Vous devez d'abord vous assurer que votre charge de travail répond à vos objectifs de disponibilité, car une architecture hautement disponible vous permettra de répondre aux besoins des clients en cas d'événements ayant une incidence sur la disponibilité. Votre stratégie de reprise après sinistre nécessite des approches différentes de celles relatives à la disponibilité, en se concentrant sur le déploiement de systèmes distincts sur plusieurs sites, afin de pouvoir prendre en charge l'intégralité de la charge de travail si nécessaire.

Vous devez tenir compte de la disponibilité de votre charge de travail lors de la planification de la reprise après sinistre, car elle influencera l'approche que vous adopterez. Une charge de travail exécutée sur une seule EC2 instance Amazon dans une zone de disponibilité ne présente pas de haute disponibilité. Si un problème d'inondation local affecte cette zone de disponibilité, ce scénario nécessite le basculement vers une autre zone de disponibilité pour atteindre les objectifs de reprise après sinistre. Comparez ce scénario à une charge de travail à haute disponibilité déployée sur plusieurs sites actif/actif, dans laquelle la charge de travail est déployée sur plusieurs régions actives et où toutes les régions desservent le trafic de production. Dans ce cas, même dans le cas peu probable où une catastrophe massive rendrait une région inutilisable, la stratégie de reprise après sinistre est mise en œuvre en acheminant tout le trafic vers les régions restantes.

La façon dont vous abordez les données est également différente entre la disponibilité et la reprise après sinistre. Envisagez une solution de stockage qui se réplique en continu sur un autre site pour garantir une haute disponibilité (par exemple, une active/active charge de travail multisite). Si un ou plusieurs fichiers sont supprimés ou endommagés sur le périphérique de stockage principal, ces modifications destructrices peuvent être répliquées sur le périphérique de stockage secondaire. Dans ce scénario, malgré la haute disponibilité, la capacité de basculement en cas de suppression ou de corruption des données sera compromise. Au lieu de cela, une point-in-time sauvegarde est également requise dans le cadre d'une stratégie de reprise après sinistre.

Plan de continuité des activités (BCP)

Votre plan de reprise après sinistre doit être un sous-ensemble du plan de continuité des activités (BCP) de votre organisation, il ne doit pas s'agir d'un document autonome. Il est inutile de maintenir des objectifs de reprise après sinistre ambitieux pour restaurer une charge de travail si les objectifs commerciaux de cette charge de travail ne peuvent pas être atteints en raison de l'impact du sinistre sur des éléments de votre activité autres que votre charge de travail. Par exemple, un tremblement de terre peut vous empêcher de transporter les produits achetés sur votre application de commerce électronique. Même si une reprise après sinistre efficace permet à votre charge de travail de fonctionner, votre BCP doit répondre aux besoins de transport. Votre stratégie de reprise après sinistre doit être basée sur les exigences, les priorités et le contexte de l'entreprise.

Analyse de l'impact commercial et évaluation des risques

Une analyse d'impact commercial doit quantifier l'impact commercial d'une interruption de vos charges de travail. Il doit identifier l'impact sur les clients internes et externes de l'impossibilité d'utiliser vos charges de travail et l'effet que cela a sur votre entreprise. L'analyse devrait permettre de déterminer la rapidité avec laquelle la charge de travail doit être mise à disposition et le degré de perte de données qui peut être toléré. Cependant, il est important de noter que les objectifs de reprise ne doivent pas être définis de manière isolée ; la probabilité d'une interruption et le coût de la reprise sont des facteurs clés qui contribuent à déterminer la valeur commerciale de la reprise après sinistre pour une charge de travail.

L'impact commercial peut dépendre du temps. Vous pouvez envisager d'en tenir compte dans votre planification de reprise après sinistre. Par exemple, une perturbation de votre système de paie est susceptible d'avoir un impact très important sur l'entreprise juste avant que tout le monde ne soit payé, mais elle peut avoir un impact faible juste après que tout le monde a déjà été payé.

Une évaluation des risques liés au type de sinistre et à son impact géographique ainsi qu'une vue d'ensemble de la mise en œuvre technique de votre charge de travail détermineront la probabilité d'une perturbation pour chaque type de sinistre.

Pour les charges de travail très critiques, vous pouvez envisager de déployer une infrastructure dans plusieurs régions avec la réplication des données et des sauvegardes continues en place afin de minimiser l'impact commercial. Pour les charges de travail moins critiques, une stratégie valable peut consister à ne pas mettre en place de reprise après sinistre. Et pour certains scénarios de catastrophe, il est également valable de ne pas avoir de stratégie de reprise après sinistre en tant

que décision éclairée basée sur une faible probabilité que la catastrophe se produise. N'oubliez pas que les zones de disponibilité d'une région AWS sont déjà conçues avec une distance significative entre elles et une planification minutieuse de leur localisation, de sorte que les catastrophes les plus courantes ne devraient affecter qu'une zone et pas les autres. Par conséquent, une architecture multi-AZ au sein d'une région AWS peut déjà répondre à la plupart de vos besoins en matière d'atténuation des risques.

Le coût des options de reprise après sinistre doit être évalué afin de garantir que la stratégie de reprise après sinistre fournit le niveau de valeur commerciale approprié compte tenu de l'impact et des risques commerciaux.

Grâce à toutes ces informations, vous pouvez documenter la menace, le risque, l'impact et le coût des différents scénarios de sinistre ainsi que les options de restauration associées. Ces informations doivent être utilisées pour déterminer vos objectifs de restauration pour chacune de vos charges de travail.

Objectifs de reprise (RTO et RPO)

Lors de la création d'une stratégie de reprise après sinistre (DR), les entreprises planifient le plus souvent l'objectif de temps de reprise (RTO) et l'objectif de point de reprise (RPO).

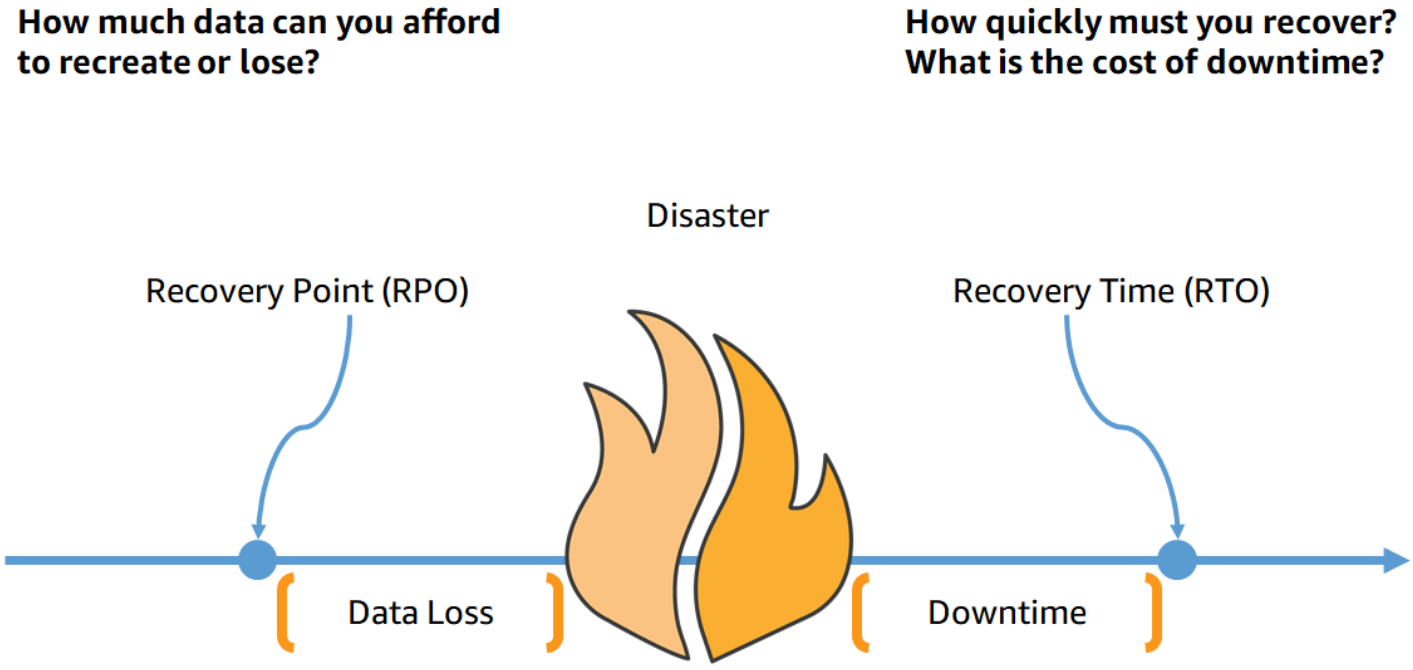

Figure 3 - Objectifs de rétablissement

L'objectif de temps de rétablissement (RTO) est le délai maximum acceptable entre l'interruption du service et le rétablissement du service. Cet objectif détermine ce qui est considéré comme une fenêtre temporelle acceptable lorsque le service n'est pas disponible et est défini par l'organisation.

Quatre stratégies de reprise après sinistre sont abordées dans ce paper : sauvegarde et restauration, pilote, veille chaude et multisite active/active (voir Disaster Recovery Options in the Cloud). Dans le schéma suivant, l'entreprise a déterminé son RTO maximal autorisé ainsi que la limite de ce qu'elle peut dépenser pour sa stratégie de restauration du service. Compte tenu des objectifs de l'entreprise, les stratégies DR Pilot Light ou Warm Standby satisferont à la fois au RTO et aux critères de coût.

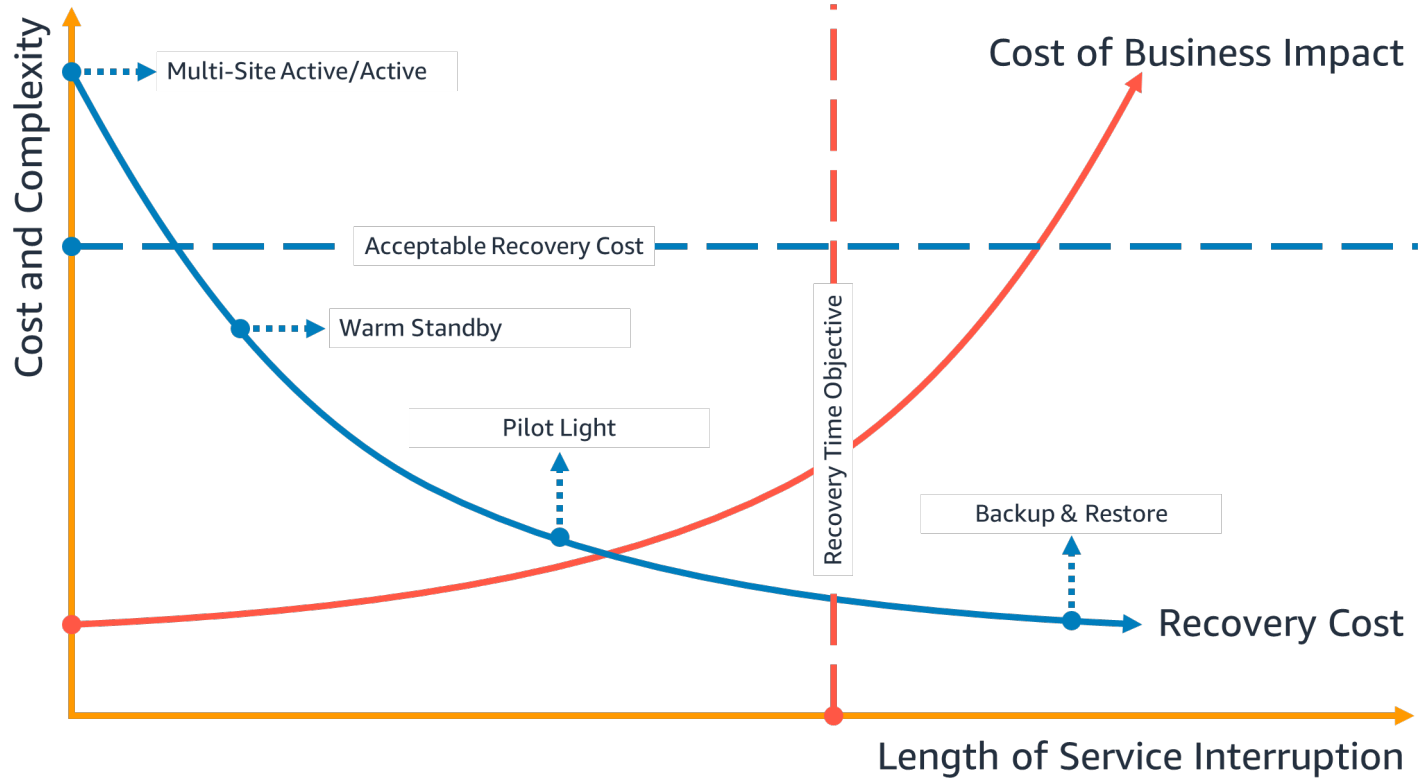

Figure 4 : objectif en matière de temps de rétablissement

L'objectif de point de restauration (RPO) est le délai maximal acceptable depuis le dernier point de récupération des données. Cet objectif détermine ce qui est considéré comme une perte de données acceptable entre le dernier point de reprise et l'interruption de service et est défini par l'organisation.

Dans le schéma suivant, l'entreprise a déterminé son RPO maximal autorisé ainsi que la limite de ce qu'elle peut dépenser pour sa stratégie de récupération de données. Parmi les quatre stratégies DR, la stratégie Pilot Light ou Warm Standby DR répond à la fois aux critères de RPO et de coût.

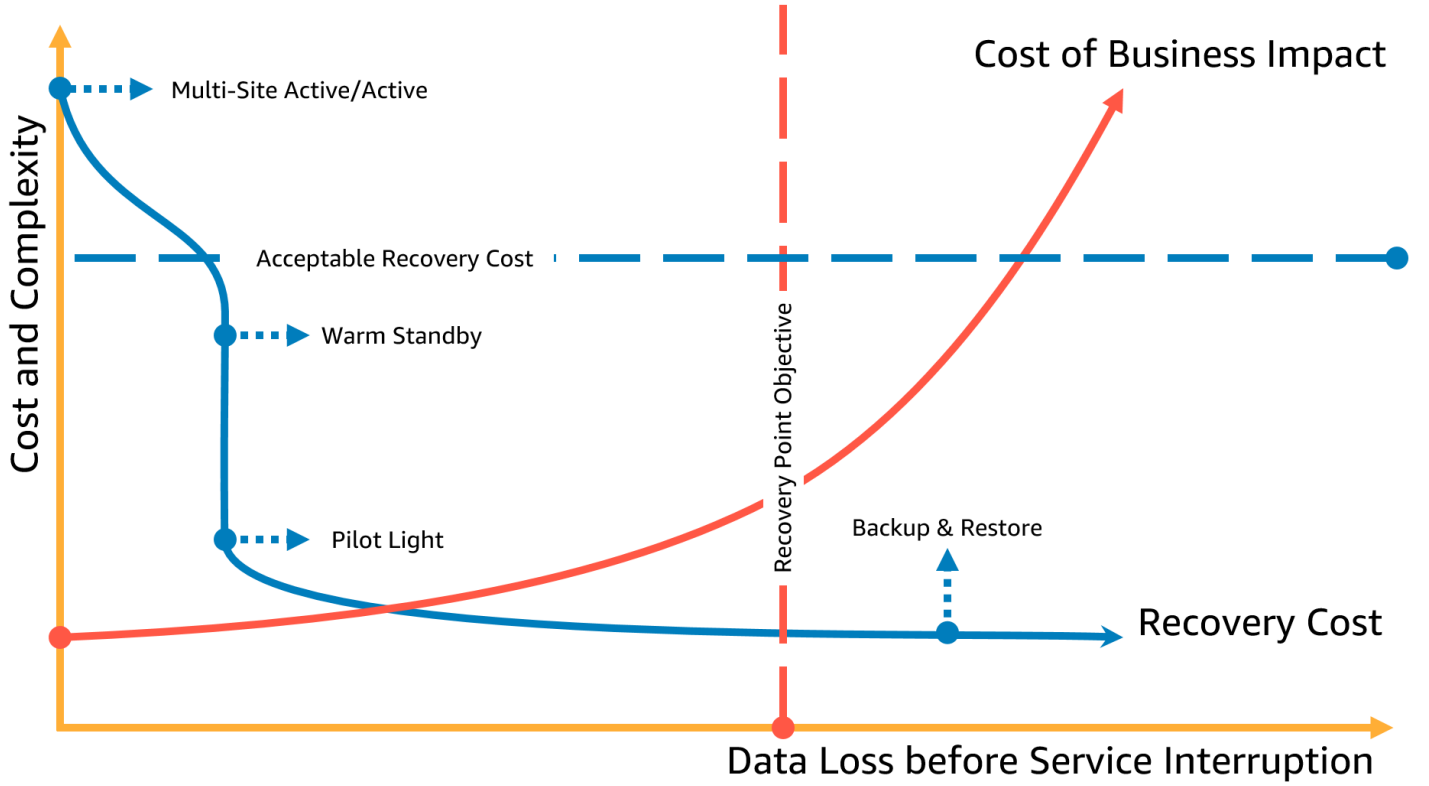

Figure 5 : objectif du point de rétablissement

> **Note**
>
> Si le coût de la stratégie de reprise est supérieur au coût de la panne ou de la perte, l'option de reprise ne doit pas être mise en place à moins d'un facteur secondaire tel que les exigences réglementaires. Envisagez des stratégies de recouvrement dont les coûts varient lors de cette évaluation.

La reprise après sinistre est différente dans le cloud

Les stratégies de reprise après sinistre évoluent avec l'innovation technique. Un plan de reprise après sinistre sur site peut impliquer le transport physique de bandes ou la réplication de données vers un autre site. Votre organisation doit réévaluer l'impact commercial, les risques et le coût de ses précédentes stratégies de reprise après sinistre afin d'atteindre ses objectifs de reprise après sinistre sur AWS. La reprise après sinistre dans le cloud AWS présente les avantages suivants par rapport aux environnements traditionnels :

- Restaurez rapidement après un sinistre avec une complexité réduite
- Des tests simples et reproductibles vous permettent de tester plus facilement et plus fréquemment
- La réduction des frais de gestion réduit la charge opérationnelle
- Les possibilités d'automatisation, de réduction des risques d'erreur et d'amélioration du temps de restauration

AWS vous permet d'échanger les dépenses d'investissement fixes d'un centre de données de sauvegarde physique contre les dépenses d'exploitation variables d'un environnement cloud de taille adaptée, ce qui peut réduire les coûts de manière significative.

Pour de nombreuses entreprises, la reprise après sinistre sur site reposait sur le risque d'interruption d'une ou de plusieurs charges de travail dans un centre de données et sur la restauration de données sauvegardées ou répliquées dans un centre de données secondaire. Lorsque les organisations déploient des charges de travail sur AWS, elles peuvent mettre en œuvre une charge de travail bien conçue et s'appuyer sur la conception de l'infrastructure cloud mondiale AWS pour atténuer les effets de telles perturbations. Consultez le livre blanc AWS Well-Architected Framework - Reliability Pillar pour plus d'informations sur les meilleures pratiques architecturales en matière de conception et d'exploitation de charges de travail fiables, sécurisées, efficaces et économiques dans le cloud. Utilisez-le AWS Well-Architected Toolpour revoir régulièrement vos charges de travail afin de vous assurer qu'elles respectent les meilleures pratiques et les directives du Well-Architected Framework. L'outil est disponible gratuitement dans le AWS Management Console.

Si vos charges de travail sont sur AWS, vous n'avez pas à vous soucier de la connectivité du centre de données (à l'exception de votre capacité à y accéder), de l'alimentation, de la climatisation, de l'extinction des incendies et du matériel. Tout cela est géré pour vous et vous avez accès à plusieurs zones de disponibilité isolées en cas de panne (chacune composée d'un ou de plusieurs centres de données distincts).

Région AWS unique

En cas de sinistre dû à une interruption ou à la perte d'un centre de données physique, la mise en œuvre d'une charge de travail hautement disponible dans plusieurs zones de disponibilité au sein d'une même région AWS permet d'atténuer les risques naturels et techniques. La sauvegarde continue des données dans cette région unique peut réduire le risque de menaces humaines, telles qu'une erreur ou une activité non autorisée susceptible d'entraîner une perte de données. Chaque région AWS est composée de plusieurs zones de disponibilité, chacune isolée des défaillances des autres zones. Chaque zone de disponibilité comprend à son tour un ou plusieurs centres de données physiques distincts. Pour mieux isoler les problèmes importants et atteindre une haute disponibilité, vous pouvez partitionner les charges de travail entre plusieurs zones d'une même région. Les zones de disponibilité sont conçues pour assurer la redondance physique et assurer la résilience, permettant des performances ininterrompues, même en cas de panne de courant, d'interruption d'Internet, d'inondation ou d'autres catastrophes naturelles. Consultez AWS Global Cloud Infrastructure pour découvrir comment AWS s'y prend.

En déployant sur plusieurs zones de disponibilité au sein d'une même région AWS, votre charge de travail est mieux protégée contre la défaillance d'un seul (voire de plusieurs) centres de données. Pour plus de sécurité lors de votre déploiement dans une seule région, vous pouvez sauvegarder les données et la configuration (y compris la définition de l'infrastructure) dans une autre région. Cette stratégie réduit la portée de votre plan de reprise après sinistre pour inclure uniquement la sauvegarde et la restauration des données. Tirer parti de la résilience multirégionale en effectuant une sauvegarde dans une autre région AWS est simple et peu coûteux par rapport aux autres options multirégionales décrites dans la section suivante. Par exemple, la sauvegarde sur Amazon Simple Storage Service (Amazon S3) vous permet de récupérer immédiatement vos données. Toutefois, si votre stratégie de reprise après sinistre pour certaines parties de vos données impose des exigences plus souples en termes de temps de récupération (de quelques minutes à quelques heures), l'utilisation d'Amazon Glacier ou d'Amazon Glacier Deep Archive réduira considérablement les coûts de votre stratégie de sauvegarde et de restauration.

Certaines charges de travail peuvent être soumises à des exigences réglementaires en matière de résidence des données. Si cela s'applique à votre charge de travail dans une localité qui ne compte actuellement qu'une seule région AWS, en plus de concevoir des charges de travail multi-AZ pour une haute disponibilité, comme indiqué ci-dessus, vous pouvez également les utiliser AZs au sein de cette région comme emplacements distincts, ce qui peut être utile pour répondre aux exigences de résidence des données applicables à votre charge de travail au sein de cette région. Les stratégies

de reprise après sinistre décrites dans les sections suivantes utilisent plusieurs régions AWS, mais peuvent également être mises en œuvre à l'aide de zones de disponibilité plutôt que de régions.

Plusieurs régions AWS

En cas de sinistre impliquant le risque de perdre plusieurs centres de données situés à une distance significative les uns des autres, vous devez envisager des options de reprise après sinistre afin de pallier les catastrophes naturelles et techniques affectant une région entière au sein d'AWS. Toutes les options décrites dans les sections suivantes peuvent être mises en œuvre sous forme d'architectures multirégionales afin de se protéger contre de telles catastrophes.

Options de reprise après sinistre dans le cloud

Les stratégies de reprise après sinistre mises à votre disposition au sein d'AWS peuvent être classées en quatre grandes catégories, allant du faible coût et de la faible complexité des sauvegardes aux stratégies plus complexes utilisant plusieurs régions actives. Active/passive les stratégies utilisent un site actif (tel qu'une région AWS) pour héberger la charge de travail et gérer le trafic. Le site passif (tel qu'une autre région AWS) est utilisé pour la restauration. Le site passif ne dessert pas activement le trafic tant qu'un événement de basculement n'est pas déclenché.

Il est essentiel d'évaluer et de tester régulièrement votre stratégie de reprise après sinistre afin de pouvoir l'invoquer en toute confiance, le cas échéant. Utilisez AWS Resilience Hub pour valider et suivre en permanence la résilience de vos AWS charges de travail, notamment pour déterminer si vous êtes susceptible d'atteindre vos objectifs de RTO et de RPO.

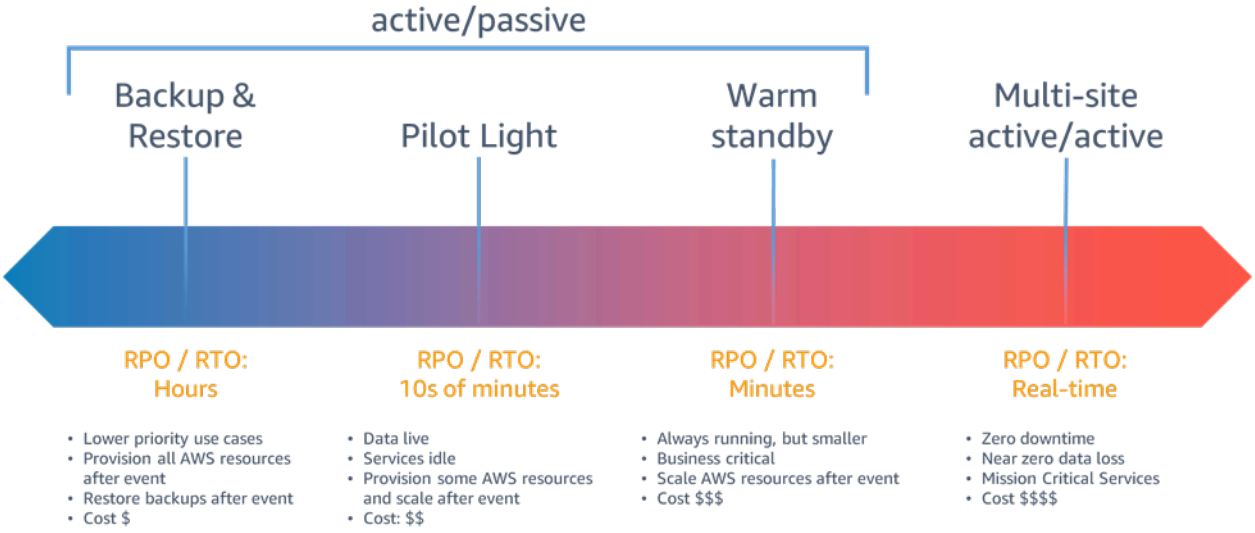

Figure 6 - Stratégies de reprise après sinistre

En cas de sinistre dû à une interruption ou à la perte d'un centre de données physique pour une charge de travail bien conçue et hautement disponible, vous n'aurez peut-être besoin que d'une approche de sauvegarde et de restauration pour la reprise après sinistre. Si votre définition d'un sinistre va au-delà de la perturbation ou de la perte d'un centre de données physique au profit de celui d'une région ou si vous êtes soumis à des exigences réglementaires qui l'exigent, vous devriez envisager une lampe pilote, une mise en veille chaude ou une solution multisite actif/active.

Lorsque vous choisissez votre stratégie et les ressources AWS pour la mettre en œuvre, gardez à l'esprit qu'au sein d'AWS, nous divisons généralement les services entre le plan de données et le

plan de contrôle. Le plan de données vise à fournir un service en temps réel, tandis que les plans de contrôle servent à configurer l'environnement. Pour une résilience maximale, vous devez utiliser uniquement les opérations du plan de données dans le cadre de votre opération de basculement. Cela est dû au fait que les plans de données ont généralement des objectifs de conception de disponibilité plus élevés que les plans de contrôle.

Sauvegarde et restauration

La sauvegarde et la restauration constituent une approche appropriée pour prévenir la perte ou la corruption des données. Cette approche peut également être utilisée pour faire face à un sinistre régional en répliquant les données vers d'autres régions AWS, ou pour pallier le manque de redondance des charges de travail déployées dans une seule zone de disponibilité. Outre les données, vous devez redéployer l'infrastructure, la configuration et le code de l'application dans la région de restauration. Pour permettre le redéploiement rapide de l'infrastructure sans erreur, vous devez toujours effectuer le déploiement en utilisant l'infrastructure en tant que code (IaC) à l'aide de services tels que AWS CloudFormationou le AWS Cloud Development Kit (AWS CDK). Sans IaC, la restauration des charges de travail dans la région de restauration peut s'avérer complexe, ce qui augmentera les temps de restauration et pourrait même dépasser votre RTO. Outre les données utilisateur, veillez à sauvegarder le code et la configuration, y compris les Amazon Machine Images (AMIs) que vous utilisez pour créer des EC2 instances Amazon. Vous pouvez l'utiliser AWS CodePipelinepour automatiser le redéploiement du code et de la configuration de l'application.

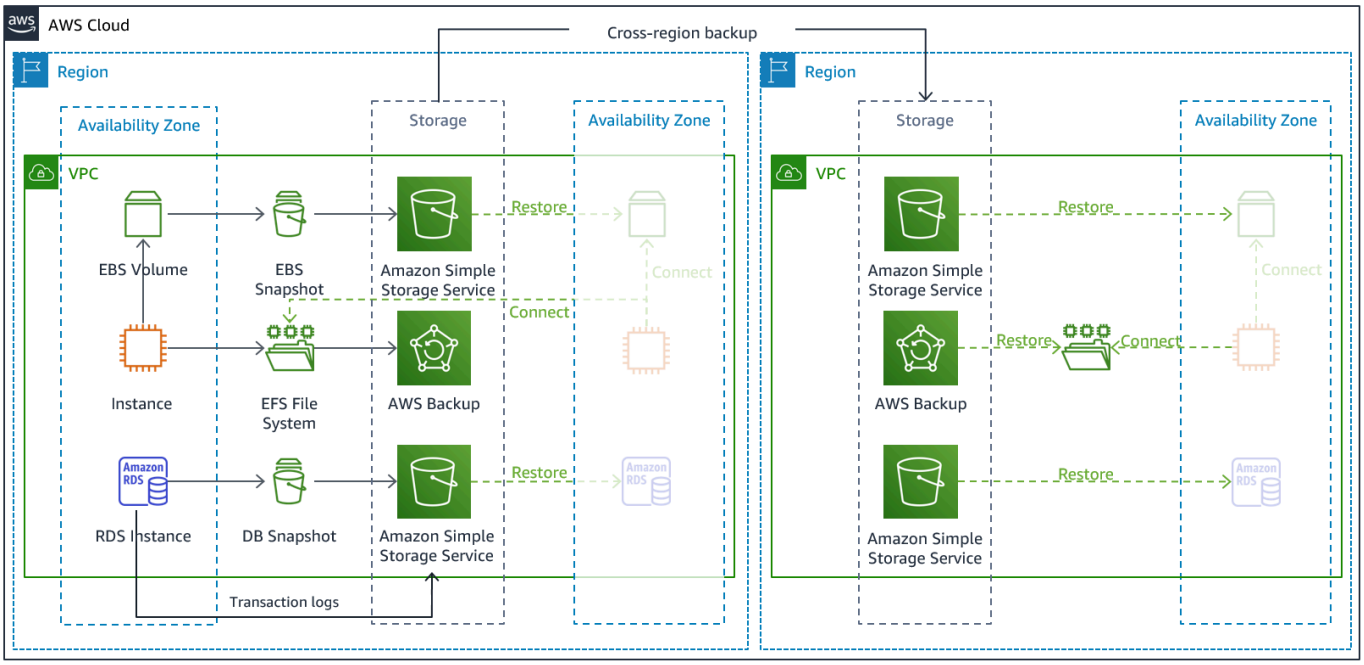

Figure 7 : architecture de sauvegarde et de restauration

Services AWS

Les données de votre charge de travail nécessiteront une stratégie de sauvegarde exécutée périodiquement ou en continu. La fréquence à laquelle vous exécutez votre sauvegarde déterminera le point de restauration que vous pouvez atteindre (qui doit correspondre à votre RPO). La sauvegarde doit également offrir un moyen de la restaurer à l'endroit où elle a été effectuée. La sauvegarde avec point-in-time restauration est disponible via les services et ressources suivants :

- Instantané d'Amazon Elastic Block Store (Amazon EBS)
- Sauvegarde Amazon DynamoDB
- Instantané Amazon RDS
- Instantané de base de données Amazon Aurora
- Sauvegarde Amazon EFS (lors de l'utilisation AWS Backup)
- Instantané Amazon Redshift
- Instantané Amazon Neptune
- Amazon DocumentDB
- Amazon FSx pour Windows File Server, Amazon FSx pour Lustre, Amazon FSx pour NetApp ONTAP et Amazon FSx pour OpenZFS

Pour Amazon Simple Storage Service (Amazon S3), vous pouvez utiliser Amazon S3 Cross-Region Replication (CRR) pour copier des objets de manière asynchrone dans un compartiment S3 dans la région DR en continu, tout en fournissant un contrôle de version pour les objets stockés afin que vous puissiez choisir votre point de restauration. La réplication continue des données présente l'avantage d'être le délai le plus court (proche de zéro) pour sauvegarder vos données, mais elle peut ne pas vous protéger contre les catastrophes telles que la corruption des données ou les attaques malveillantes (telles que la suppression non autorisée de données) ni contre les point-in-time sauvegardes. La réplication continue est abordée dans la section Services AWS pour Pilot Light.

AWS Backupfournit un emplacement centralisé pour configurer, planifier et surveiller les fonctionnalités de sauvegarde AWS pour les services et ressources suivants :

- Volumes d'Amazon Elastic Block Store (Amazon EBS)
- EC2Instances Amazon

- Bases de données Amazon Relational Database Service (Amazon RDS) (y compris les bases de données Amazon Aurora)
- Tableaux Amazon DynamoDB
- Systèmes de fichiers Amazon Elastic File System (Amazon EFS)
- AWS Storage GatewayVolumes
- Amazon FSx pour Windows File Server, Amazon FSx pour Lustre, Amazon FSx pour NetApp ONTAP et Amazon FSx pour OpenZFS

AWS Backup prend en charge la copie de sauvegardes entre régions, par exemple vers une région de reprise après sinistre.

En tant que stratégie de reprise après sinistre supplémentaire pour vos données Amazon S3, activez la gestion des versions des objets S3. Le versionnement des objets protège vos données dans S3 des conséquences des actions de suppression ou de modification en conservant la version d'origine avant l'action. Le versionnement des objets peut être une solution utile pour atténuer les catastrophes liées à des erreurs humaines. Si vous utilisez la réplication S3 pour sauvegarder des données dans votre région DR, Amazon S3 ajoute par défaut un marqueur de suppression uniquement dans le compartiment source lorsqu'un objet est supprimé dans le compartiment source. Cette approche protège les données de la région DR contre les suppressions malveillantes dans la région source.

Outre les données, vous devez également sauvegarder la configuration et l'infrastructure nécessaires pour redéployer votre charge de travail et atteindre votre objectif de temps de restauration (RTO). AWS CloudFormationfournit une infrastructure sous forme de code (IaC) et vous permet de définir toutes les ressources AWS de votre charge de travail afin de pouvoir les déployer et les redéployer de manière fiable sur plusieurs comptes AWS et régions AWS. Vous pouvez sauvegarder les EC2 instances Amazon utilisées par votre charge de travail sous la forme d'Amazon Machine Images (AMIs). L'AMI est créée à partir d'instantanés du volume racine de votre instance et de tout autre volume EBS attaché à votre instance. Vous pouvez utiliser cette AMI pour lancer une version restaurée de l' EC2 instance. Une AMI peut être copiée dans ou entre les régions. Vous pouvez également les utiliser AWS Backuppour copier des sauvegardes entre comptes et vers d'autres régions AWS. La fonctionnalité de sauvegarde entre comptes permet de se protéger contre les catastrophes telles que les menaces internes ou la compromission des comptes. AWS Backup ajoute également des fonctionnalités de EC2 sauvegarde supplémentaires : outre les volumes EBS individuels de l'instance, stocke et suit AWS Backup également les métadonnées suivantes : type d'instance, cloud privé virtuel (VPC) configuré, groupe de sécurité, rôle IAM, configuration de

surveillance et balises. Toutefois, ces métadonnées supplémentaires ne sont utilisées que lors de la restauration de la EC2 sauvegarde dans la même région AWS.

Toutes les données stockées dans la région de reprise après sinistre sous forme de sauvegarde doivent être restaurées au moment du basculement. AWS Backup offre une fonctionnalité de restauration, mais n'active pas actuellement la restauration planifiée ou automatique. Vous pouvez implémenter la restauration automatique dans la région de reprise après sinistre à l'aide du SDK AWS. APIs AWS Backup Vous pouvez configurer cette tâche comme une tâche récurrente régulière ou déclencher une restauration chaque fois qu'une sauvegarde est terminée. La figure suivante montre un exemple de restauration automatique à l'aide d'Amazon Simple Notification Service (Amazon SNS) et. AWS Lambda La mise en œuvre d'une restauration périodique planifiée des données est une bonne idée, car la restauration des données à partir d'une sauvegarde est une opération du plan de contrôle. Si cette opération n'était pas disponible lors d'un sinistre, vous auriez toujours des banques de données opérationnelles créées à partir d'une sauvegarde récente.

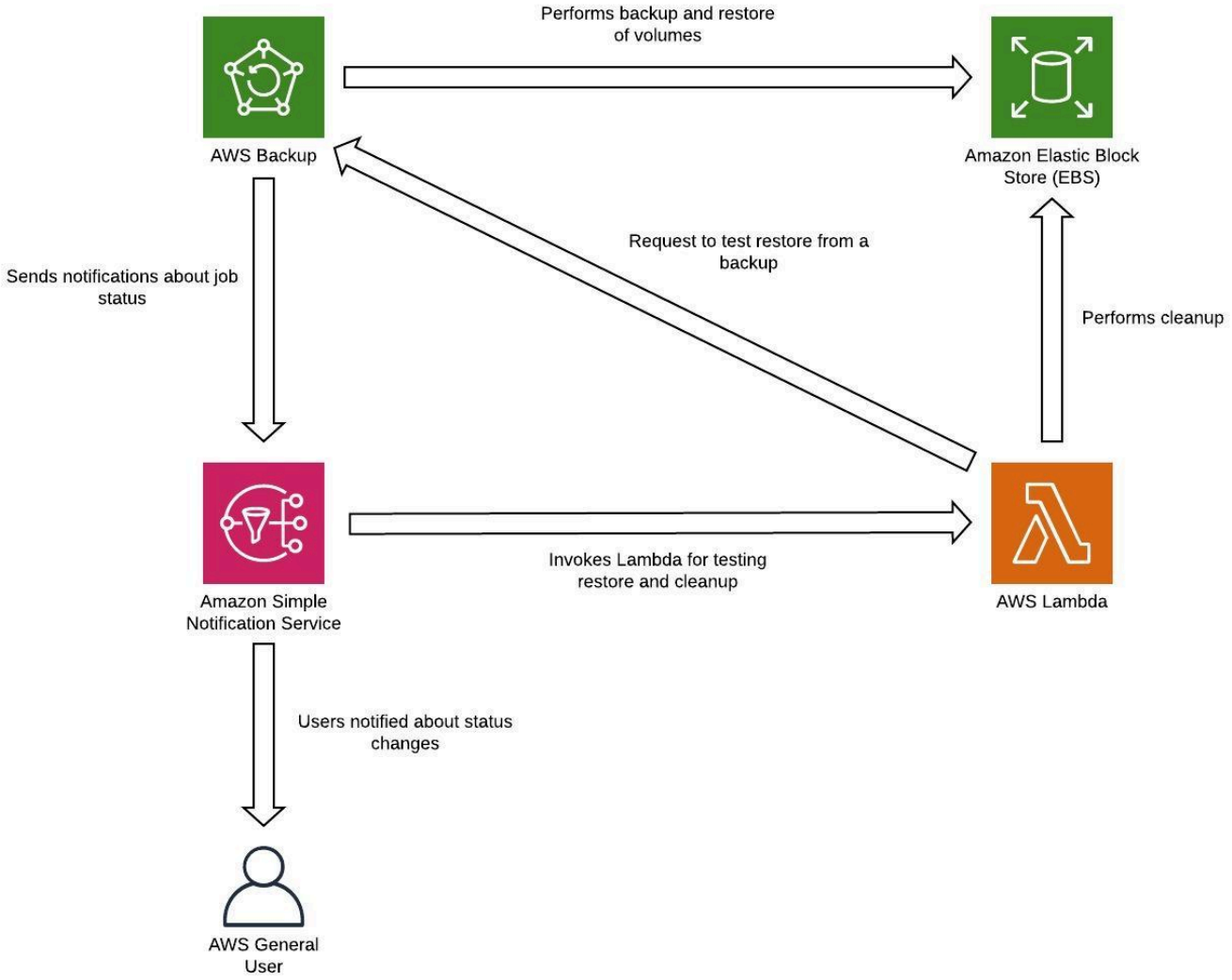

Figure 8 : restauration et test des sauvegardes

> **Note**
>
> Votre stratégie de sauvegarde doit inclure le test de vos sauvegardes. Consultez la section Tests de reprise après sinistre pour plus d'informations. Reportez-vous au document AWS Well-Architected Lab : Testing Backup and Restore of Data pour une démonstration pratique de la mise en œuvre.

Veilleuse

Avec l'approche pilote, vous répliquez vos données d'une région à l'autre et vous fournissez une copie de votre infrastructure de charge de travail principale. Les ressources requises pour prendre en charge la réplication et la sauvegarde des données, telles que les bases de données et le stockage d'objets, sont toujours actives. D'autres éléments, tels que les serveurs d'applications, sont chargés avec le code et les configurations de l'application, mais sont « désactivés » et ne sont utilisés que pendant les tests ou lorsque le basculement après sinistre est invoqué. Dans le cloud, vous avez la flexibilité de déprovisionner les ressources lorsque vous n'en avez pas besoin et de les provisionner lorsque vous en avez besoin. Une bonne pratique en cas de « désactivation » consiste à ne pas déployer la ressource, puis à créer la configuration et les fonctionnalités nécessaires pour la déployer (« activation ») en cas de besoin. Contrairement à l'approche de sauvegarde et de restauration, votre infrastructure principale est toujours disponible et vous avez toujours la possibilité de fournir rapidement un environnement de production à grande échelle en activant et en développant vos serveurs d'applications.

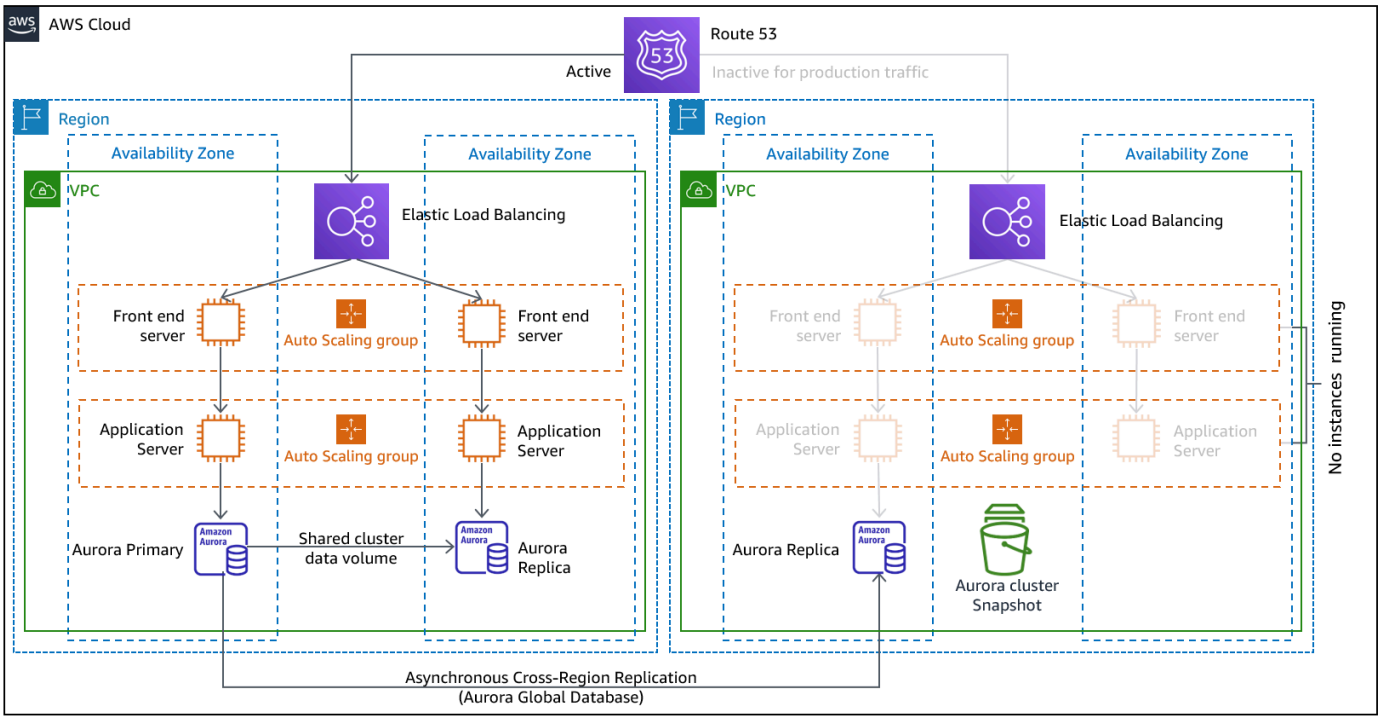

Figure 9 : architecture de la veilleuse

Une approche pilote légère minimise le coût permanent de la reprise après sinistre en minimisant les ressources actives, et simplifie la reprise au moment d'un sinistre, car les exigences d'infrastructure de base sont toutes réunies. Cette option de restauration vous oblige à modifier votre approche

de déploiement. Vous devez apporter des modifications d'infrastructure de base à chaque région et déployer les modifications de charge de travail (configuration, code) simultanément dans chaque région. Cette étape peut être simplifiée en automatisant vos déploiements et en utilisant l'infrastructure en tant que code (IaC) pour déployer l'infrastructure sur plusieurs comptes et régions (déploiement complet de l'infrastructure dans la région principale et déploiement de l'infrastructure réduite/désactivée dans les régions DR). Il est recommandé d'utiliser un compte différent par région afin de garantir le plus haut niveau d'isolation des ressources et de sécurité (dans le cas où des informations d'identification compromises font également partie de vos plans de reprise après sinistre).

Avec cette approche, vous devez également vous prémunir contre un sinistre lié aux données. La réplication continue des données vous protège contre certains types de catastrophes, mais elle peut ne pas vous protéger contre la corruption ou la destruction des données, sauf si votre stratégie inclut également le versionnement des données stockées ou des options de point-in-time restauration. Vous pouvez sauvegarder les données répliquées dans la région sinistrée pour créer des point-in-time sauvegardes dans cette même région.

Services AWS

Outre l'utilisation des services AWS décrits dans la section Backup and Restore pour créer point-in-time des sauvegardes, considérez également les services suivants pour votre stratégie pilote.

Dans un premier temps, la réplication continue des données vers des bases de données actives et des magasins de données dans la région de reprise après sinistre est la meilleure approche pour un faible RPO (lorsqu'elle est utilisée en plus des point-in-time sauvegardes évoquées précédemment). AWS fournit une réplication continue, asynchrone et interrégionale des données à l'aide des services et ressources suivants :

- Réplication d'Amazon Simple Storage Service (Amazon S3)
- Amazon RDS lit les répliques
- Bases de données mondiales Amazon Aurora
- Tableaux globaux Amazon DynamoDB
- Clusters globaux Amazon DocumentDB
- Banque de données mondiale pour Amazon ElastiCache (Redis OSS)

Grâce à la réplication continue, les versions de vos données sont disponibles presque immédiatement dans votre région DR. Les temps de réplication réels peuvent être surveillés à l'aide

de fonctionnalités de service telles que le contrôle du temps de réplication S3 (S3 RTC) pour les objets S3 et les fonctionnalités de gestion des bases de données mondiales Amazon Aurora.

Lorsque vous basculez pour exécuter votre read/write charge de travail depuis la région de reprise après sinistre, vous devez promouvoir une réplique en lecture RDS pour en faire l'instance principale. Pour les instances de base de données autres qu'Aurora, le processus prend quelques minutes et le redémarrage fait partie du processus. Pour la réplication entre régions (CRR) et le basculement avec RDS, l'utilisation de la base de données globale Amazon Aurora présente plusieurs avantages. La base de données globale utilise une infrastructure dédiée qui laisse vos bases de données entièrement disponibles pour servir votre application et peut être répliquée vers la région secondaire avec une latence généralement inférieure à une seconde (et bien inférieure à 100 millisecondes dans une région AWS). Avec la base de données mondiale Amazon Aurora, si votre région principale subit une dégradation des performances ou une panne, vous pouvez confier la responsabilité de lecture/écriture à l'une des régions secondaires en moins d'une minute, même en cas de panne régionale complète. Vous pouvez également configurer Aurora pour surveiller le temps de latence du RPO de tous les clusters secondaires afin de vous assurer qu'au moins un cluster secondaire reste dans votre fenêtre de RPO cible.

Une version réduite de votre infrastructure de charge de travail principale avec moins ou moins de ressources doit être déployée dans votre région de reprise après sinistre. Vous pouvez ainsi définir votre infrastructure et la déployer de manière cohérente sur les comptes AWS et les régions AWS. AWS CloudFormation AWS CloudFormation utilise des pseudo-paramètres prédéfinis pour identifier le compte AWS et la région AWS dans lesquels il est déployé. Par conséquent, vous pouvez implémenter une logique conditionnelle dans vos CloudFormation modèles afin de déployer uniquement la version réduite de votre infrastructure dans la région DR. Pour les déploiements d' EC2 instances, une Amazon Machine Image (AMI) fournit des informations telles que la configuration matérielle et les logiciels installés. Vous pouvez implémenter un pipeline Image Builder qui crée les éléments dont AMIs vous avez besoin et les copier à la fois dans votre région principale et dans votre région de sauvegarde. Cela permet de s'assurer que ces golden AMIs disposent de tout ce dont vous avez besoin pour redéployer ou augmenter votre charge de travail dans une nouvelle région, en cas de sinistre. Les EC2 instances Amazon sont déployées dans une configuration réduite (moins d'instances que dans votre région principale). Pour étendre l'infrastructure afin de prendre en charge le trafic de production, consultez Amazon EC2 Auto Scaling dans la section Warm Standby.

Pour une active/passive configuration telle que la veilleuse, tout le trafic est initialement dirigé vers la région principale et passe à la région de reprise après sinistre si la région principale n'est plus disponible. Cette opération de basculement peut être lancée automatiquement ou manuellement. Le basculement automatique basé sur des contrôles de santé ou des alarmes doit être utilisé avec

prudence. Même en utilisant les meilleures pratiques décrites ici, le temps et le point de restauration seront supérieurs à zéro, ce qui entraînera une certaine perte de disponibilité et de données. Si vous tombez alors que vous n'en avez pas besoin (fausse alerte), vous subissez ces pertes. Le basculement manuel est donc souvent utilisé. Dans ce cas, nous vous conseillons tout de même d'automatiser les étapes de basculement, de sorte que vous n'ayez à appuyer que sur un bouton pour lancer le basculement.

Il existe plusieurs options de gestion du trafic à prendre en compte lors de l'utilisation AWS des services.

L'une des options consiste à utiliser Amazon Route 53. À l'aide d'Amazon Route 53, vous pouvez associer plusieurs points de terminaison IP dans une ou plusieurs régions AWS à un nom de domaine Route 53. Vous pouvez ensuite acheminer le trafic vers le point de terminaison approprié sous ce nom de domaine. En cas de basculement, vous devez transférer le trafic vers le point de terminaison de restauration, et non vers le point de terminaison principal. Les bilans de santé d'Amazon Route 53 surveillent ces points de terminaison. À l'aide de ces contrôles de santé, vous pouvez configurer le basculement DNS initié automatiquement pour garantir que le trafic est envoyé uniquement vers des points de terminaison sains, ce qui constitue une opération extrêmement fiable effectuée sur le plan de données. Pour implémenter cela à l'aide d'un basculement initié manuellement, vous pouvez utiliser Amazon Application Recovery Controller (ARC). Avec ARC, vous pouvez créer des bilans de santé de Route 53 qui ne vérifient pas réellement l'état de santé, mais agissent plutôt comme des commutateurs marche/arrêt sur lesquels vous avez un contrôle total. À l'aide de l'interface de ligne de commande AWS ou du kit SDK AWS, vous pouvez créer un script de basculement à l'aide de cette API de plan de données hautement disponible. Votre script active ces commutateurs (les contrôles de santé de la Route 53) en indiquant à la Route 53 d'envoyer le trafic vers la région de restauration plutôt que vers la région principale. Une autre option utilisée par certains pour le basculement manuel consiste à utiliser une politique de routage pondérée et à modifier les pondérations des régions principale et de restauration afin que tout le trafic soit dirigé vers la région de restauration. Sachez toutefois qu'il s'agit d'une opération de plan de contrôle et qu'elle n'est donc pas aussi résiliente que l'approche du plan de données utilisant Amazon Application Recovery Controller (ARC).

Une autre option consiste à utiliser AWS Global Accelerator. À l'aide de l' AnyCast adresse IP, vous pouvez associer plusieurs points de terminaison dans une ou plusieurs régions AWS à la même adresse IP publique statique. AWS Global Accelerator achemine ensuite le trafic vers le point de terminaison approprié associé à cette adresse. Les contrôles de santé de Global Accelerator surveillent les terminaux. À l'aide de ces contrôles de santé, AWS Global Accelerator vous contrôlez l'état de vos applications et achemine automatiquement le trafic utilisateur vers

le point de terminaison de l'application sain. Dans le cas d'un basculement initié manuellement, vous pouvez régler le point de terminaison qui reçoit le trafic à l'aide des numéros de signalisation, mais notez qu'il s'agit d'une opération de plan de contrôle. Global Accelerator réduit les temps de latence du point de terminaison de l'application, car il utilise le vaste réseau périphérique d'AWS pour acheminer le trafic vers le backbone du réseau AWS dès que possible. Global Accelerator évite également les problèmes de mise en cache qui peuvent survenir avec les systèmes DNS (tels que Route 53).

Amazon CloudFront propose le basculement d'origine, selon lequel, si une demande donnée vers le point de terminaison principal échoue, CloudFront achemine la demande vers le point de terminaison secondaire. Contrairement aux opérations de basculement décrites précédemment, toutes les demandes suivantes sont toujours envoyées au point de terminaison principal et le basculement est effectué pour chaque demande.

AWS Reprise après sinistre élastique

AWS Elastic Disaster Recovery (DRS) réplique en continu les applications hébergées sur le serveur et les bases de données hébergées sur le serveur à partir de n'importe quelle source en AWS utilisant la réplication au niveau des blocs du serveur sous-jacent. Elastic Disaster Recovery vous permet d'utiliser une région AWS Cloud comme cible de reprise après sinistre pour une charge de travail hébergée sur site ou chez un autre fournisseur de cloud, ainsi que pour son environnement. Il peut également être utilisé pour la reprise après sinistre des charges de travail AWS hébergées si celles-ci se composent uniquement d'applications et de bases de données hébergées sur EC2 (c'est-à-dire pas RDS). Elastic Disaster Recovery utilise la stratégie Pilot Light, qui consiste à conserver une copie des données et des ressources « désactivées » dans un Amazon Virtual Private Cloud (Amazon VPC) utilisé comme zone intermédiaire. Lorsqu'un événement de basculement est déclenché, les ressources intermédiaires sont utilisées pour créer automatiquement un déploiement à pleine capacité dans le VPC Amazon cible utilisé comme lieu de restauration.

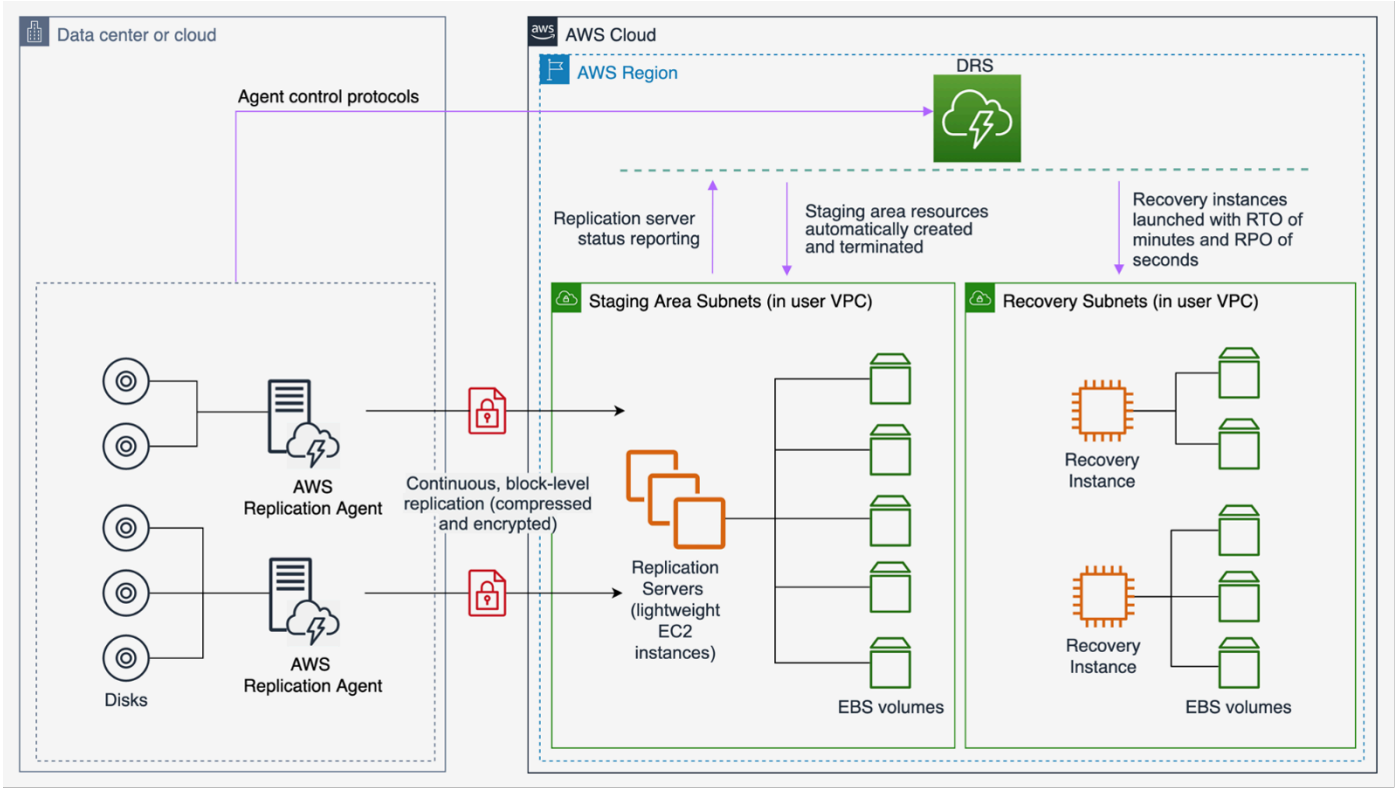

Figure 10 : architecture AWS Elastic Disaster Recovery

Secours semi-automatique

L'approche du secours semi-automatique consiste à s'assurer qu'il existe une copie réduite verticalement, mais entièrement fonctionnelle, de votre environnement de production dans une autre région. Cette approche étend le concept d'environnement en veille et réduit le temps de récupération, car votre charge de travail reste active dans une autre région. Cette approche vous permet également d'effectuer plus facilement des tests ou de mettre en œuvre des tests continus afin de renforcer la confiance dans votre capacité à vous remettre après un sinistre.

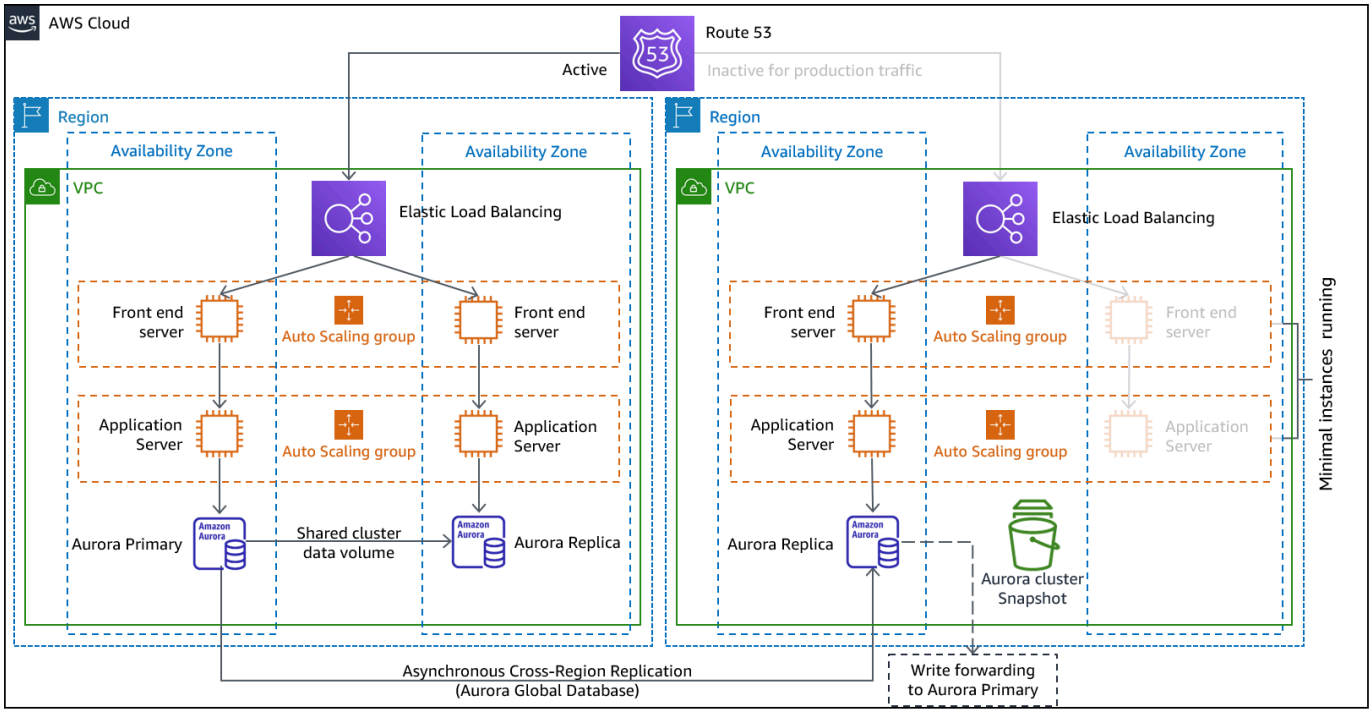

Figure 11 : architecture Warm Standby

Remarque : La différence entre veilleuse et veille chaude peut parfois être difficile à comprendre. Les deux incluent un environnement dans votre région DR avec des copies des actifs principaux de votre région. La différence est que la lampe pilote ne peut pas traiter les demandes sans que des mesures supplémentaires ne soient prises au préalable, tandis que le mode veille peut gérer le trafic (à des niveaux de capacité réduits) immédiatement. L'approche pilote vous oblige à « activer » les serveurs, éventuellement à déployer une infrastructure supplémentaire (non essentielle) et à passer à l'échelle supérieure, tandis que le mode veille chaude vous oblige uniquement à le faire évoluer (tout est déjà déployé et fonctionne). Utilisez vos besoins en matière de RTO et de RPO pour vous aider à choisir entre ces approches.

Services AWS

Tous les services AWS couverts par les rubriques sauvegarde, restauration et pilote sont également utilisés en mode veille pour la sauvegarde des données, la réplication des données, le routage active/passive du trafic et le déploiement de l'infrastructure, y compris EC2 les instances.

Amazon EC2 Auto Scaling est utilisé pour dimensionner les ressources, notamment EC2 les instances Amazon, les tâches Amazon ECS, le débit Amazon DynamoDB et les répliques Amazon Aurora au sein d'une région AWS. Amazon EC2 Auto Scaling adapte le déploiement de l' EC2

instance dans les zones de disponibilité d'une région AWS, garantissant ainsi la résilience au sein de cette région. Utilisez Auto Scaling pour étendre votre région DR à une capacité de production maximale, dans le cadre de stratégies de mise en veille ou de veille chaude. Par exemple EC2, pour augmenter le paramètre de capacité souhaité sur le groupe Auto Scaling. Vous pouvez ajuster ce paramètre manuellement via le SDK AWS AWS Management Console, automatiquement, ou en redéployant votre AWS CloudFormation modèle à l'aide de la nouvelle valeur de capacité souhaitée. Vous pouvez utiliser AWS CloudFormation des paramètres pour faciliter le redéploiement du CloudFormation modèle. Assurez-vous que les quotas de service dans votre région DR sont suffisamment élevés pour ne pas vous empêcher de passer à la capacité de production.

Auto Scaling étant une activité relevant du plan de contrôle, le fait d'en dépendre diminuera la résilience de votre stratégie de restauration globale. Il s'agit d'un compromis. Vous pouvez choisir de fournir une capacité suffisante pour que la région de restauration puisse gérer l'intégralité de la charge de production telle que déployée. Cette configuration statiquement stable est appelée hot standby (voir la section suivante). Vous pouvez également choisir de fournir moins de ressources, ce qui vous coûtera moins cher, tout en vous appuyant sur Auto Scaling. Certaines implémentations de DR déploieront suffisamment de ressources pour gérer le trafic initial, garantissant ainsi un faible RTO, puis s'appuieront sur Auto Scaling pour accélérer le trafic suivant.

Multisite actif/actif

Vous pouvez exécuter votre charge de travail simultanément dans plusieurs régions dans le cadre d'une stratégie actif/active ou actif/passive multisite. Le multisite active/active gère le trafic provenant de toutes les régions dans lesquelles il est déployé, tandis que le mode de veille active ne traite que le trafic provenant d'une seule région, et les autres régions ne sont utilisées que pour la reprise après sinistre. Grâce à une active/active approche multisite, les utilisateurs peuvent accéder à votre charge de travail dans toutes les régions dans lesquelles elle est déployée. Cette approche est l'approche la plus complexe et la plus coûteuse en matière de reprise après sinistre, mais elle peut réduire le temps de reprise à un niveau proche de zéro dans la plupart des cas de sinistre si les choix technologiques et la mise en œuvre sont appropriés (toutefois, la corruption des données peut nécessiter des sauvegardes, ce qui se traduit généralement par un point de reprise différent de zéro). La mise en veille prolongée utilise une active/passive configuration dans laquelle les utilisateurs ne sont dirigés que vers une seule région et les régions DR n'absorbent pas de trafic. La plupart des clients trouvent que s'ils veulent créer un environnement complet dans la deuxième région, il est logique de l'utiliser actif/actif. Sinon, si vous ne souhaitez pas utiliser les deux régions pour gérer le trafic utilisateur, Warm Standby propose une approche plus économique et moins complexe sur le plan opérationnel.

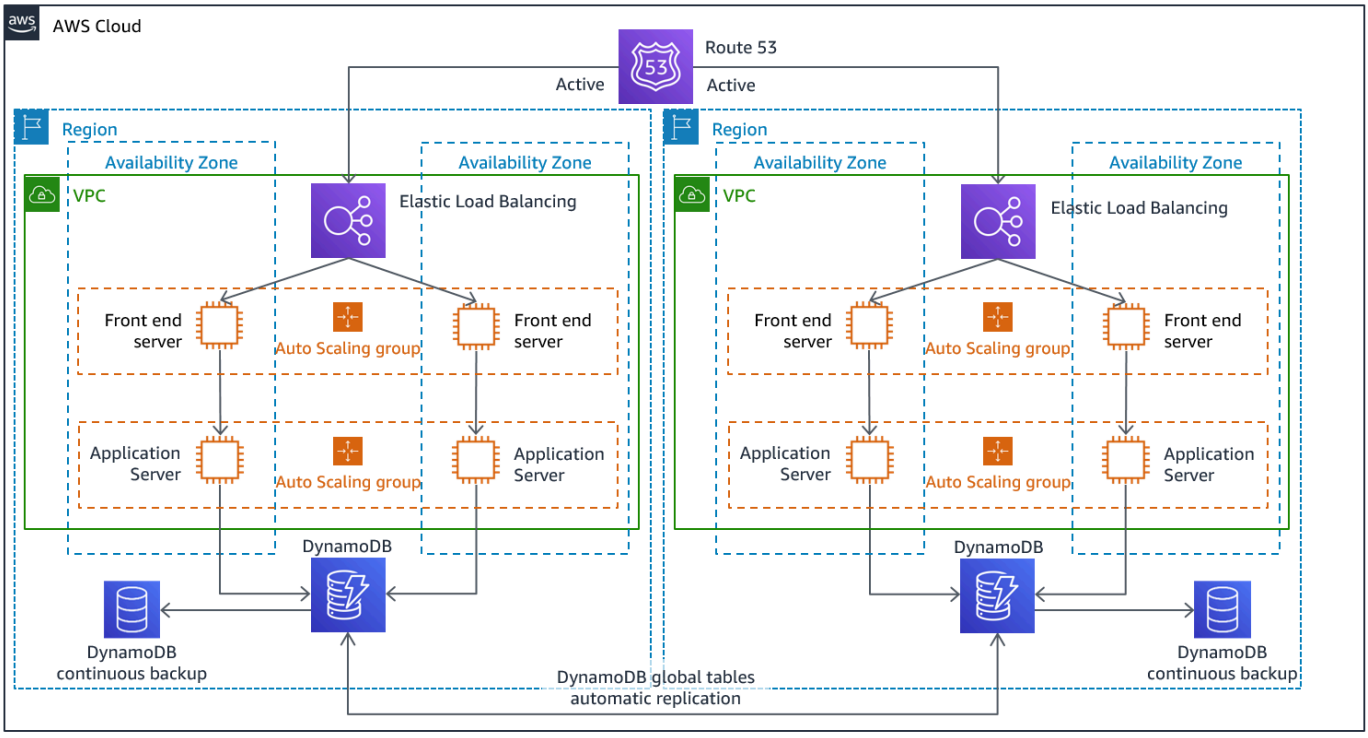

Figure 12 : active/active architecture multisite (remplacez un chemin actif par Inactif pour le mode veille à chaud)

Une approche multisite active/active, because the workload is running in more than one Region, there is no such thing as failover in this scenario. Disaster recovery testing in this case would focus on how the workload reacts to loss of a Region: Is traffic routed away from the failed Region? Can the other Region(s) handle all the traffic? Testing for a data disaster is also required. Backup and recovery are still required and should be tested regularly. It should also be noted that recovery times for a data disaster involving data corruption, deletion, or obfuscation will always be greater than zero and the recovery point will always be at some point before the disaster was discovered. If the additional complexity and cost of a multi-site active/active (ou « hot standby ») étant nécessaire pour maintenir des temps de restauration proches de zéro, des efforts supplémentaires doivent être déployés pour maintenir la sécurité et prévenir les erreurs humaines afin d'atténuer les risques de catastrophes humaines.

Services AWS

Tous les services AWS couverts par les rubriques sauvegarde et restauration, Pilot Light et Warm Standby sont également utilisés ici pour la sauvegarde des données, la réplication point-in-time

des données, le routage active/active du trafic, ainsi que le déploiement et le dimensionnement de l'infrastructure, y compris EC2 les instances.

Pour les active/passive scénarios décrits précédemment (Pilot Light et Warm Standby), Amazon Route 53 AWS Global Accelerator peut être utilisé pour acheminer le trafic réseau vers la région active. Dans le active/active cadre de cette stratégie, ces deux services permettent également de définir des politiques qui déterminent quels utilisateurs accèdent à quel point de terminaison régional actif. AWS Global Accelerator Vous définissez alors un numéro de trafic pour contrôler le pourcentage de trafic dirigé vers chaque point de terminaison de l'application. Amazon Route 53 prend en charge cette approche basée sur le pourcentage, ainsi que de nombreuses autres politiques disponibles, notamment celles basées sur la géoproximité et la latence. Global Accelerator exploite automatiquement le vaste réseau de serveurs périphériques AWS pour intégrer le trafic au backbone du réseau AWS dès que possible, réduisant ainsi les latences des demandes.

La réplication asynchrone des données avec cette stratégie permet un RPO proche de zéro. Les services AWS tels que la base de données mondiale Amazon Aurora utilisent une infrastructure dédiée qui laisse vos bases de données entièrement disponibles pour servir votre application, et peuvent être répliquées dans un maximum de cinq régions secondaires avec une latence typique inférieure à une seconde. With conçoit active/passive strategies, writes occur only to the primary Region. The difference with active/active la manière dont la cohérence des données avec les écritures dans chaque région active est gérée. Il est courant de concevoir les lectures des utilisateurs pour qu'elles soient diffusées depuis la région la plus proche de chez eux, connue sous le nom de lecture locale. Avec les écritures, plusieurs options s'offrent à vous :

- Une stratégie globale d'écriture achemine toutes les écritures vers une seule région. En cas d'échec de cette région, une autre région serait encouragée à accepter les écrits. La base de données globale Aurora convient parfaitement à Write Global, car elle prend en charge la synchronisation avec les répliques en lecture dans toutes les régions, et vous pouvez promouvoir l'une des régions secondaires pour qu'elle prenne des read/write responsabilités en moins d'une minute. Aurora prend également en charge le transfert d'écriture, qui permet aux clusters secondaires d'une base de données globale Aurora de transférer des instructions SQL qui effectuent des opérations d'écriture vers le cluster principal.
- Une stratégie locale d'écriture achemine les écritures vers la région la plus proche (tout comme les lectures). Les tables globales Amazon DynamoDB permettent une telle stratégie, en autorisant la lecture et l'écriture depuis toutes les régions dans lesquelles votre table globale est déployée. Les tables globales Amazon DynamoDB utilisent un dernier rédacteur pour gagner la réconciliation entre les mises à jour simultanées.

- Une stratégie partitionnée en écriture attribue les écritures à une région spécifique en fonction d'une clé de partition (comme l'ID utilisateur) afin d'éviter les conflits d'écriture. La réplication Amazon S3 configurée de manière bidirectionnelle peut être utilisée dans ce cas et prend actuellement en charge la réplication entre deux régions. Lorsque vous mettez en œuvre cette approche, veillez à activer la synchronisation des modifications des répliques sur les compartiments A et B afin de répliquer les modifications des métadonnées des répliques, telles que les listes de contrôle d'accès aux objets (ACLs), les balises d'objets ou les verrous d'objets sur les objets répliqués. Vous pouvez également configurer s'il faut ou non répliquer les marqueurs de suppression entre les compartiments de vos régions actives. Outre la réplication, votre stratégie doit également inclure point-in-time des sauvegardes afin de vous protéger contre les événements de corruption ou de destruction des données.

AWS CloudFormation est un outil puissant permettant d'appliquer une infrastructure déployée de manière cohérente entre les comptes AWS dans plusieurs régions AWS. AWS CloudFormation StackSetsétend cette fonctionnalité en vous permettant de créer, de mettre à jour ou de supprimer des CloudFormation piles sur plusieurs comptes et régions en une seule opération. Bien qu'il AWS CloudFormation utilise YAML ou JSON pour définir l'infrastructure en tant que code, il vous AWS Cloud Development Kit (AWS CDK)permet de définir l'infrastructure en tant que code à l'aide de langages de programmation familiers. Votre code est converti et est CloudFormation ensuite utilisé pour déployer des ressources dans AWS.

Détection

Il est important de savoir dès que possible que vos charges de travail ne produisent pas les résultats commerciaux escomptés. Ainsi, vous pouvez rapidement déclarer un sinistre et vous remettre d'un incident. Pour atteindre des objectifs de rétablissement ambitieux, ce temps de réponse associé à des informations appropriées est essentiel pour atteindre les objectifs de rétablissement. Si votre objectif de temps de reprise est d'une heure, vous devez détecter l'incident, informer le personnel concerné, engager vos processus d'escalade, évaluer les informations (si vous en avez) sur le délai de reprise prévu (sans exécuter le plan de reprise après sinistre), déclarer un sinistre et récupérer dans l'heure qui suit.

> **Note**
>
> Si les parties prenantes décident de ne pas invoquer la DR même si le RTO est en danger, réévaluez les plans et objectifs de DR. La décision de ne pas invoquer de plans de reprise après sinistre peut être due à des plans inadéquats ou à un manque de confiance dans leur exécution.

Il est essentiel de tenir compte de la détection, de la notification, de l'escalade, de la découverte et de la déclaration des incidents dans votre planification et vos objectifs afin de fournir des objectifs réalistes et réalisables qui apportent une valeur commerciale.

AWS publie la plupart des up-to-the-minute informations relatives à la disponibilité des services sur le Service Health Dashboard. Renseignez-vous à tout moment pour obtenir des informations sur le statut actuel ou abonnez-vous à un fil RSS pour être informé des interruptions de chaque service individuel. Si vous rencontrez un problème opérationnel en temps réel avec l'un de nos services qui n'apparaît pas sur le Service Health Dashboard, vous pouvez créer une demande de Support.

Tableau de bord AWS HealthFournit des informations sur AWS Health les événements susceptibles d'affecter votre compte. Les informations sont présentées de deux manières : un tableau de bord qui montre les événements récents et à venir organisés par catégorie, et un journal des événements complet qui contient tous les événements des 90 derniers jours.

Pour répondre aux exigences RTO les plus strictes, vous pouvez implémenter un basculement automatique basé sur des contrôles de santé. Concevez des bilans de santé représentatifs de l'expérience utilisateur et basés sur des indicateurs de performance clés. Des bilans de santé

approfondis mettent en œuvre les fonctionnalités clés de votre charge de travail et vont au-delà de simples contrôles du rythme cardiaque. Utilisez des contrôles de santé approfondis basés sur plusieurs signaux. Faites preuve de prudence en adoptant cette approche afin de ne pas déclencher de fausses alarmes, car le fait de basculer lorsque cela n'est pas nécessaire peut en soi présenter des risques de disponibilité.

Tester la reprise après sinistre

Testez la mise en œuvre de la reprise après sinistre pour valider la mise en œuvre et testez régulièrement le basculement vers la région DR de votre charge de travail afin de vous assurer que le RTO et le RPO sont respectés.

Une tendance à éviter consiste à développer des chemins de restauration rarement exécutés. Par exemple, vous pouvez avoir un magasin de données secondaire qui est utilisé pour les requêtes en lecture seule. Lorsque vous écrivez dans un magasin de données et que l'instance principale connaît une défaillance, vous pouvez basculer vers le magasin de données secondaire. Si vous ne testez pas fréquemment ce basculement, vous constaterez peut-être que vos hypothèses sur les capacités du magasin de données secondaire sont incorrectes. La capacité du secondaire, qui était peut-être suffisante lors du dernier test, peut ne plus être en mesure de tolérer la charge dans ce scénario, ou les quotas de service dans la région secondaire peuvent ne pas être suffisants.

Notre expérience a montré que le seul chemin de récupération après erreur qui fonctionne est celui que vous testez fréquemment. C'est pourquoi il est préférable de disposer d'un petit nombre de chemins de restauration.

Vous pouvez établir des modèles de reprise et tester ceux-ci régulièrement. Si votre chemin de restauration est complexe ou critique, vous devez tout de même exécuter régulièrement cet échec en production pour vérifier que le chemin de restauration fonctionne.

Gérez la dérive de configuration dans la région DR. Assurez-vous que votre infrastructure, vos données et votre configuration répondent aux besoins de la région DR. Par exemple, vérifiez cela AMIs et les quotas de service le sont up-to-date.

Vous pouvez l'utiliser AWS Configpour surveiller et enregistrer en permanence les configurations de vos ressources AWS. AWS Config peut détecter la dérive et déclencher AWS Systems Manager Automation pour corriger la dérive et déclencher des alarmes. AWS CloudFormationpeut également détecter la dérive des piles que vous avez déployées.

Conclusion

Les clients sont responsables de la disponibilité de leurs applications dans le cloud. Il est important de définir ce qu'est un sinistre et de disposer d'un plan de reprise après sinistre qui reflète cette définition et l'impact que cela peut avoir sur les résultats commerciaux. Créez un objectif de temps de restauration (RTO) et un objectif de point de reprise (RPO) sur la base d'une analyse d'impact et d'une évaluation des risques, puis choisissez l'architecture appropriée pour atténuer les risques liés aux catastrophes. Assurez-vous que la détection des catastrophes est possible et opportune : il est essentiel de savoir quand les objectifs sont menacés. Assurez-vous d'avoir un plan et validez-le en le testant. Les plans de reprise après sinistre qui n'ont pas été validés risquent de ne pas être mis en œuvre en raison d'un manque de confiance ou de l'incapacité à atteindre les objectifs de reprise après sinistre.

Collaborateurs

Les personnes qui ont contribué à ce document incluent :

- Alex Livingstone, responsable des opérations cloud, AWS Enterprise Support
- Seth Eliot, architecte principal des solutions de fiabilité, Amazon Web Services

Suggestions de lecture

Pour en savoir plus, voir :

- AWS Centre d'architecture
- Pilier de fiabilité, AWS Well-Architected Framework
- Liste de contrôle du plan de reprise après sinistre
- Mettre en œuvre des bilans de santé
- Architecture de reprise après sinistre (DR) sur AWS, partie I : stratégies de reprise dans le cloud
- Architecture de reprise après sinistre (DR) sur AWS, partie II : Backup et restauration avec restauration rapide
- Architecture de reprise après sinistre (DR) sur AWS, partie III : Pilot Light et Warm Standby
- Architecture de reprise après sinistre (DR) sur AWS, partie IV : actif/actif multisite
- Création de mécanismes de reprise après sinistre à l’aide d’Amazon Route 53
- Minimiser les dépendances dans un plan de reprise après sinistre
- Des laboratoires pratiques de AWS reprise après sinistre Well-Architected
- AWS Implémentations de solutions : Architecture d'applications multi-régions
- AWS re:Invent 2018 : Modèles d'architecture pour les applications active-active multirégionales (09-R2) ARC2

Historique du document

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

Modification	Description	Date
Mises à jour mineures	Corrections de bugs et nombreuses modifications mineures.	1er avril 2022
Livre blanc mis à jour	Mises à jour éditoriales mineures.	21 mars 2022
Livre blanc mis à jour	Ajout d'informations sur le plan de données et le plan de contrôle. Ajout de détails supplémentaires sur la façon d'implémenter le active/passive basculement. Remplacement de CloudEndure la reprise après sinistre par AWS Elastic Disaster Recovery.	17 février 2022
Mise à jour mineure	AWS Well-Architected Tool informations ajoutées.	11 février 2022
Publication initiale	Livre blanc publié pour la première fois.	12 février 2021

Avis

Il incombe aux clients de procéder à une évaluation indépendante des informations contenues dans le présent document. Ce document : (a) est fourni à titre informatif uniquement, (b) représente les offres de produits et les pratiques actuelles d'AWS, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ni aucune garantie de la part d'AWS et de ses filiales, fournisseurs ou concédants de licence. Les produits ou services AWS sont fournis « tels quels » sans garanties, déclarations ou conditions d'aucune sorte, qu'elles soient explicites ou implicites. Les responsabilités et obligations d'AWS vis-à-vis de ses clients sont régies par les contrats AWS. Le présent document ne fait partie d'aucun, et ne modifie aucun, contrat entre AWS et ses clients.

AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le AWS glossaire dans la Glossaire AWS référence.

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.