Livre blanc AWS

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

Reprise des charges de travail après sinistre sur AWS : reprise dans le cloud: Livre blanc AWS

Table of Contents

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

Date de publication : 12 février 2021 (Révisions du document)

Résumé

La reprise après sinistre est le processus de préparation et de récupération après un sinistre. Un événement qui empêche une charge de travail ou un système d'atteindre ses objectifs opérationnels sur son site de déploiement principal est considéré comme un sinistre. Le présent livre blanc décrit les bonnes pratiques de planification et de test de la reprise après sinistre pour toute charge de travail déployée sur le système AWS. Vous y trouverez également différentes approches pour atténuer les risques et atteindre l'objectif de délai de reprise (RTO) et l'objectif de point de reprise (RPO) pour cette charge de travail.

Introduction

Votre charge de travail doit remplir la fonction prévue de manière correcte et cohérente. Pour y parvenir, vous devez concevoir une architecture de résilience. La résilience est la capacité d'une charge de travail à récupérer après des perturbations de l'infrastructure ou du service, d'acquérir de manière dynamique les ressources de calcul pour satisfaire la demande et d'atténuer les perturbations telles que les erreurs de configuration ou les problèmes réseau temporaires.

La reprise après sinistre est un élément important de votre stratégie de résilience et concerne la façon dont votre charge de travail réagit en cas de sinistre (un sinistre est un événement qui a un impact négatif grave sur votre entreprise). Cette réponse doit être basée sur les objectifs opérationnels 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 reprise (RPO), et réduire les temps d'arrêt lorsque votre charge de travail n'est pas disponible, connue sous le nom d'objectif de délai de reprise (RTO). Vous devez donc implémenter la résilience dans la conception de vos charges de travail dans le cloud afin d'atteindre vos objectifs de reprise (RPO et RTO) pour un sinistre ponctuel donné. Cette approche aide votre organisation à maintenir la continuité des activités dans le cadre de la planification de la continuité de l'activité (BCP).

Ce livre blanc explique comment planifier, concevoir et implémenter des architectures sur AWS qui répondent aux objectifs de reprise après sinistre de votre entreprise. Les informations communiquées ici s'adressent aux personnes qui occupent des postes liés à la technologie, comme les directeurs techniques, les architectes, les développeurs et les membres des équipes opérationnelles.

Reprise après sinistre et disponibilité

La reprise après sinistre peut être comparée à la disponibilité, qui est 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.

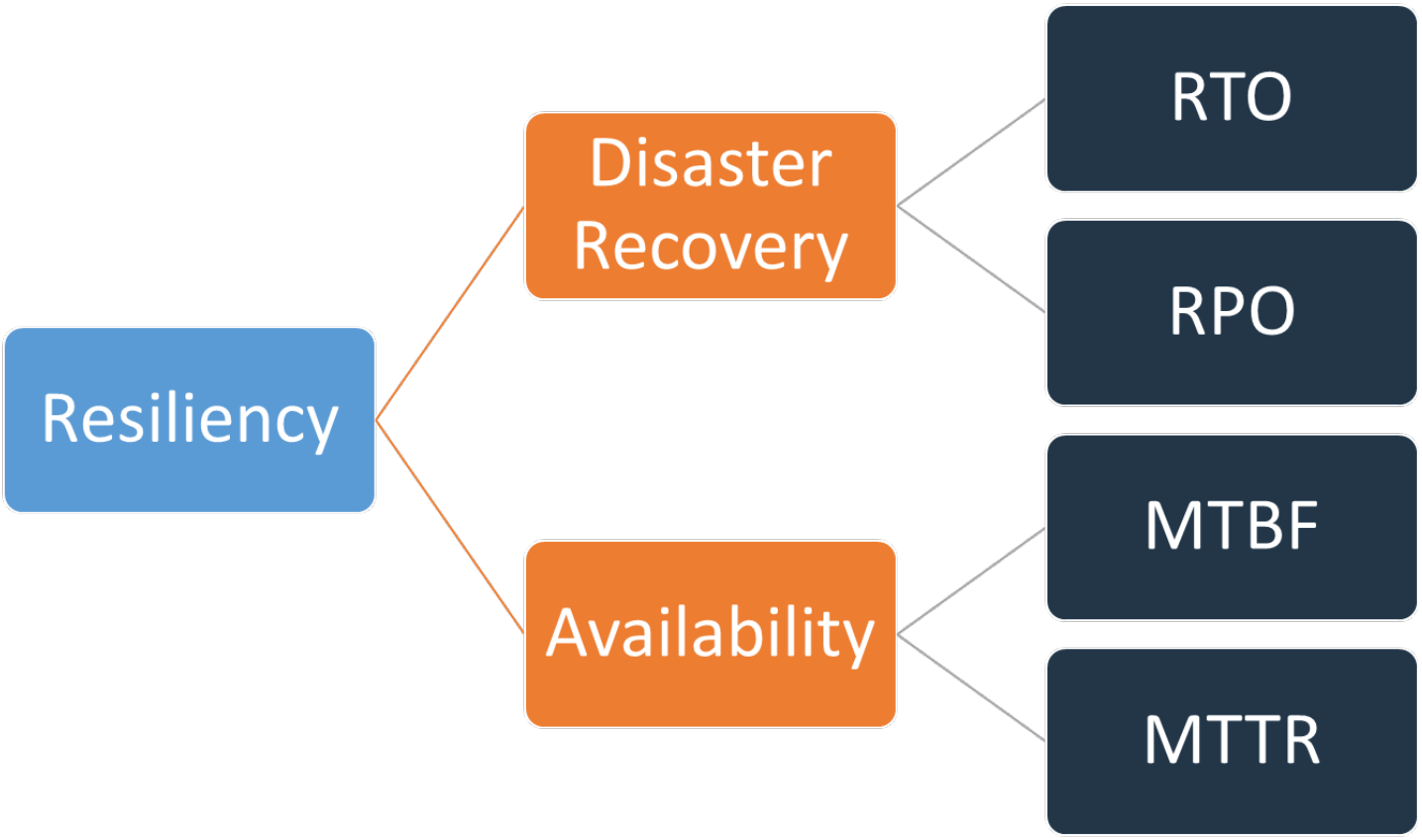

Figure 1 - Objectifs de résilience

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

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

Cette approche est souvent appelée « neuf ». 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 échouées au lieu 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 événements de sinistre, tandis que la disponibilité se concentre sur les perturbations plus courantes à plus petite échelle, telles que les défaillances de composants, les problèmes de réseau et les pics de charge. L'objectif de la reprise après sinistre est la continuité des activités, tandis que la disponibilité concerne la maximisation du temps pendant lequel une charge de travail est disponible pour exécuter les fonctionnalités opérationnelles prévues. Les deux doivent faire partie de votre stratégie de résilience.

Modèle de responsabilité partagée pour la résilience

La résilience est une responsabilité partagée entre 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 exécutant tous les services proposés dans le cloud AWS. Cette infrastructure comprend le matériel, les logiciels, les réseaux et les installations qui exécutent les services AWS Cloud services. AWS déploie les efforts commerciaux raisonnables pour rendre ces services AWS Cloud services disponibles, garantissant ainsi que la disponibilité des services respecte ou dépasse les contrats de niveau de service (SLA) AWS.

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 totalement isolée et se compose de plusieurs zones de disponibilité, qui sont des partitions physiquement isolées de l'infrastructure. Les zones de disponibilité isolent les défaillances susceptibles d'avoir un impact sur la résilience de la charge de travail, les empêchant ainsi d'avoir un impact sur d'autres régions. Mais en même temps, toutes les zones d'une région AWS sont interconnectées avec un réseau à bande passante élevée et à faible temps de latence, sur une fibre métropolitaine dédiée entièrement redondante, fournissant un réseau à haut débit et à faible temps de latence entre les zones. L'ensemble du trafic entre les zones est chiffré. Les performances du réseau sont suffisantes pour réaliser une réplication synchrone entre les zones. Les zones de disponibilité simplifient le processus de partitionnement des applications pour atteindre une haute disponibilité.

Responsabilité du client « Résilience dans le cloud »

Votre responsabilité sera déterminée par les services AWS Cloud services que vous sélectionnez. Votre choix 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) exige que le client effectue toutes les tâches de configuration et de gestion de la résilience nécessaires. Les clients qui déploient des instances Amazon EC2 sont responsables du déploiement des instances EC2 sur plusieurs sites (par exemple des zones de disponibilité AWS), de l'implémentation de la réparation automatique à l'aide de services tels que AWS Auto Scaling, ainsi que de l'utilisation de bonnes pratiques d'architecture de charge de

travail résiliente pour les applications installées sur les instances. Pour les services managés, tels qu'Amazon S3 et Amazon DynamoDB, AWS exploite la couche d'infrastructure, le système d'exploitation et les plateformes, tandis que les clients ont accès aux points de terminaison pour stocker et extraire des données. Vous êtes responsable de la gestion de la résilience de vos données, y compris les 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 utilise la redondance des autres zones de disponibilité pour continuer à traiter les demandes. Une architecture multi-AZ fait également partie d'une stratégie de reprise après sinistre conçue pour mieux isoler et protéger les charges de travail contre des problèmes tels que les pannes de courant, les éclairs, 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 de reprise après sinistre si la région active ne peut plus traiter les demandes.

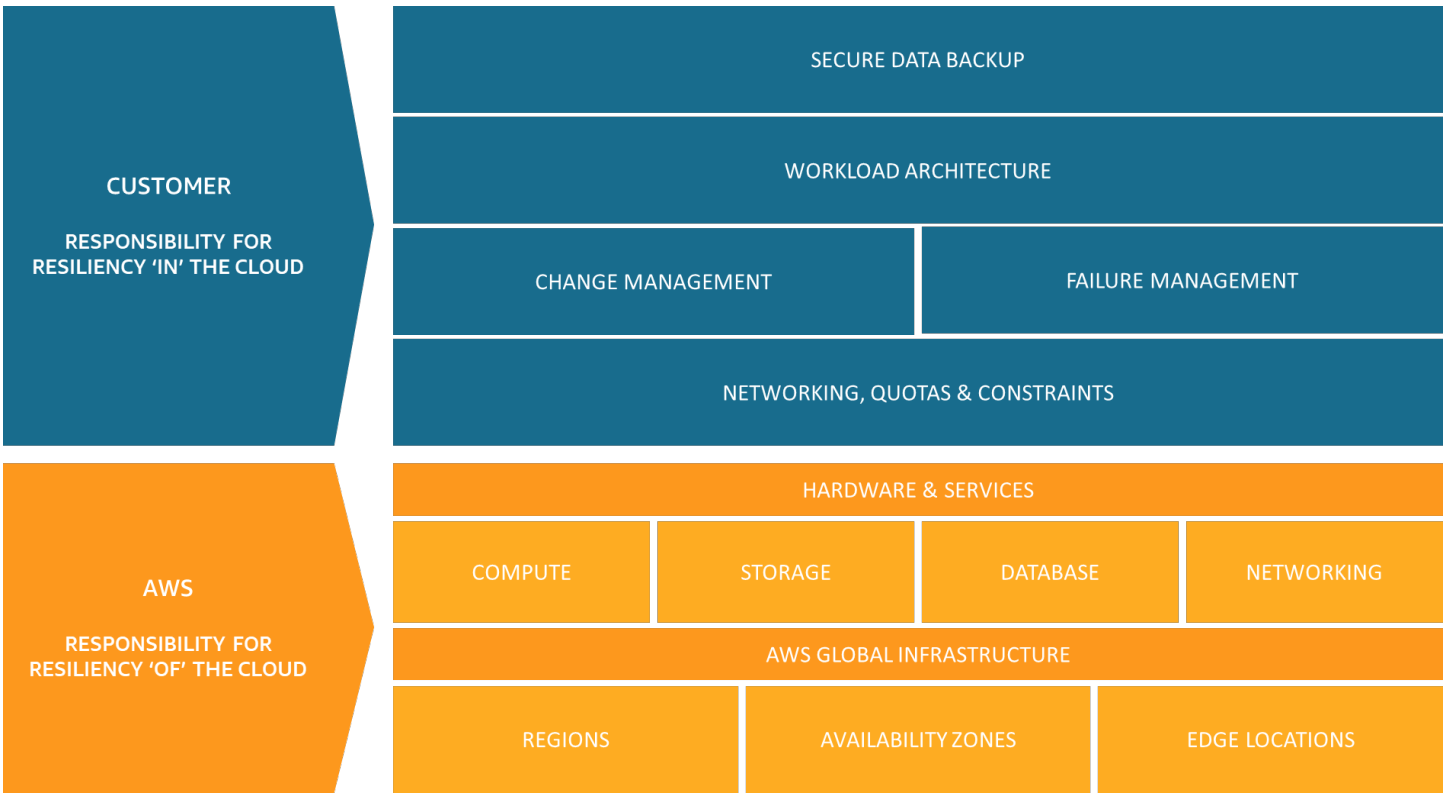

Figure 2 - AWS et le client se partagent la responsabilité d'assurer la résilience

Qu'est-ce qu'un sinistre ?

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

- Catastrophes naturelles, comme les tremblements de terre ou les inondations
- Défaillances techniques, par exemple une panne de courant ou une connectivité réseau
- Actions humaines, par exemple une mauvaise configuration par inadvertance ou un accès ou une modification non autorisé/externe

Chacune de ces catastrophes potentielles aura également un impact géographique qui peut être local, ou à l'échelle d'une région, d'un pays, d'un continent ou du monde. La nature du sinistre et son impact géographique sont tous deux importants lorsque vous élaborez votre stratégie de reprise après sinistre. Par exemple, vous pouvez atténuer un problème d'inondation locale provoquant une panne de centre de données en utilisant une stratégie multi-AZ, car une inondation n'affecterait pas plus d'une zone de disponibilité. En revanche, une attaque contre les données de production nécessiterait que vous invoquiez une stratégie de reprise après sinistre qui basculerait vers des données de sauvegarde dans une autre région AWS.

La haute disponibilité n'équivaut pas à une reprise après sinistre

La disponibilité et la reprise après sinistre reposent toutes deux sur les mêmes bonnes pratiques, telles que la surveillance des pannes, le déploiement sur plusieurs sites et le basculement automatique. Toutefois, la disponibilité se concentre sur les composants de la charge de travail, tandis que la reprise après sinistre se concentre sur des copies distinctes de la charge de travail complète. Les objectifs de la reprise après sinistre sont différents de ceux de la disponibilité, à savoir la mesure du délai de reprise après les événements de plus grande échelle qualifiés de sinistres. 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 un impact sur la disponibilité. Votre stratégie de reprise après sinistre nécessite des approches différentes de celles en matière de disponibilité. Elle doit se concentrer sur le déploiement de systèmes distincts sur plusieurs sites, afin que vous puissiez basculer l'ensemble de la charge de travail si nécessaire.

Vous devez tenir compte de la disponibilité de votre charge de travail dans la planification de la reprise après sinistre, car elle influencera l'approche que vous adoptez. Une charge de travail qui s'exécute sur une seule instance Amazon EC2 dans une zone de disponibilité n'est pas hautement disponible. Si un problème d'inondation locale affecte cette zone de disponibilité, ce scénario nécessite un basculement vers une autre zone de disponibilité pour atteindre les objectifs de reprise après sinistre. Comparez ce scénario à une charge de travail hautement disponible déployée en mode actif/actif multi-site, dans laquelle la charge de travail est déployée sur plusieurs régions actives et toutes les régions acheminent le trafic de production. Dans ce cas, même dans l'éventualité peu probable où un sinistre majeur touche une région entière, la stratégie de reprise après sinistre est réalisée en acheminant l'ensemble du trafic vers les régions restantes.

La façon dont vous abordez les données diffère également entre une stratégie de disponibilité et une stratégie de reprise après sinistre. Prenons l'exemple d'une solution de stockage qui se réplique en continu sur un autre site pour atteindre une haute disponibilité (telle qu'une charge de travail active/active multi-site). 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 sauvegarde à un instant dans le passé 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 (business continuity plan, BCP) de votre organisation. Il ne doit pas s'agir d'un document autonome. Il est inutile de maintenir des objectifs ambitieux de reprise après sinistre pour restaurer une charge de travail si les objectifs opérationnels de cette charge de travail ne peuvent pas être atteints en raison de l'impact du sinistre sur des éléments de votre entreprise autres que votre charge de travail. Par exemple, un tremblement de terre peut vous empêcher de transporter des produits achetés sur votre application d'e-commerce. Même si une reprise après sinistre efficace assure le fonctionnement de votre charge de travail, votre plan de reprise après sinistre doit prendre en compte les besoins de transport. Votre stratégie de reprise après sinistre doit être basée sur les exigences, les priorités et le contexte de l'activité.

Analyse de l'impact sur l'activité et évaluation des risques

Une analyse d'impact sur l'activité doit quantifier l'impact opérationnel d'une perturbation de vos charges de travail. Elle doit identifier, sur les clients internes et externes, l'impact que génère l'impossibilité d'utiliser vos charges de travail et l'effet que cela a sur votre activité. L'analyse doit aider à déterminer la rapidité avec laquelle la charge de travail doit être mise à disposition et la quantité de perte de données pouvant être tolérée. Cependant, il est important de noter que les objectifs de reprise ne doivent pas être définis séparément ; la probabilité d'interruption et le coût de la reprise sont des facteurs clés qui aident à déterminer la valeur opérationnelle de la reprise après sinistre pour une charge de travail.

L'impact sur l'activité peut également varier en fonction du moment où survient le sinistre. Nous vous conseillons 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 si elle survient juste avant la période de paie, mais elle peut avoir un impact faible si elle survient juste après la période de paie.

Une évaluation des risques du type de sinistre et de son impact géographique ainsi qu'un aperçu de l'implémentation technique de votre charge de travail détermineront la probabilité d'une perturbation pour chaque type de sinistre.

Pour les charges de travail hautement critiques, vous pouvez envisager de mettre en place une haute disponibilité dans plusieurs régions avec des sauvegardes continues afin de minimiser l'impact sur

l'activité. Pour les charges de travail moins critiques, ne pas mettre en place de reprise après sinistre peut s'avérer être une stratégie adaptée. Et pour certains scénarios catastrophe, il est également acceptable de ne pas mettre en place de stratégie de reprise après sinistre étant donnée la faible probabilité de survenue du sinistre. N'oubliez pas que la conception des zones de disponibilité au sein d'une région AWS comporte déjà une distance significative entre les zones et une planification minutieuse de l'emplacement. Ainsi, les sinistres les plus courants ne devraient affecter qu'une zone. Par conséquent, une architecture multi-AZ au sein d'une région AWS peut déjà répondre à 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 adapté de valeur opérationnelle en fonction de l'impact et du risque opérationnels.

Ces informations vous permettent de documenter la menace, le risque, l'impact et le coût des différents scénarios de sinistre et des options de reprise associées. Ces informations doivent être utilisées pour déterminer vos objectifs de reprise 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 délai de reprise (RTO) et l'objectif de point de reprise (RPO).

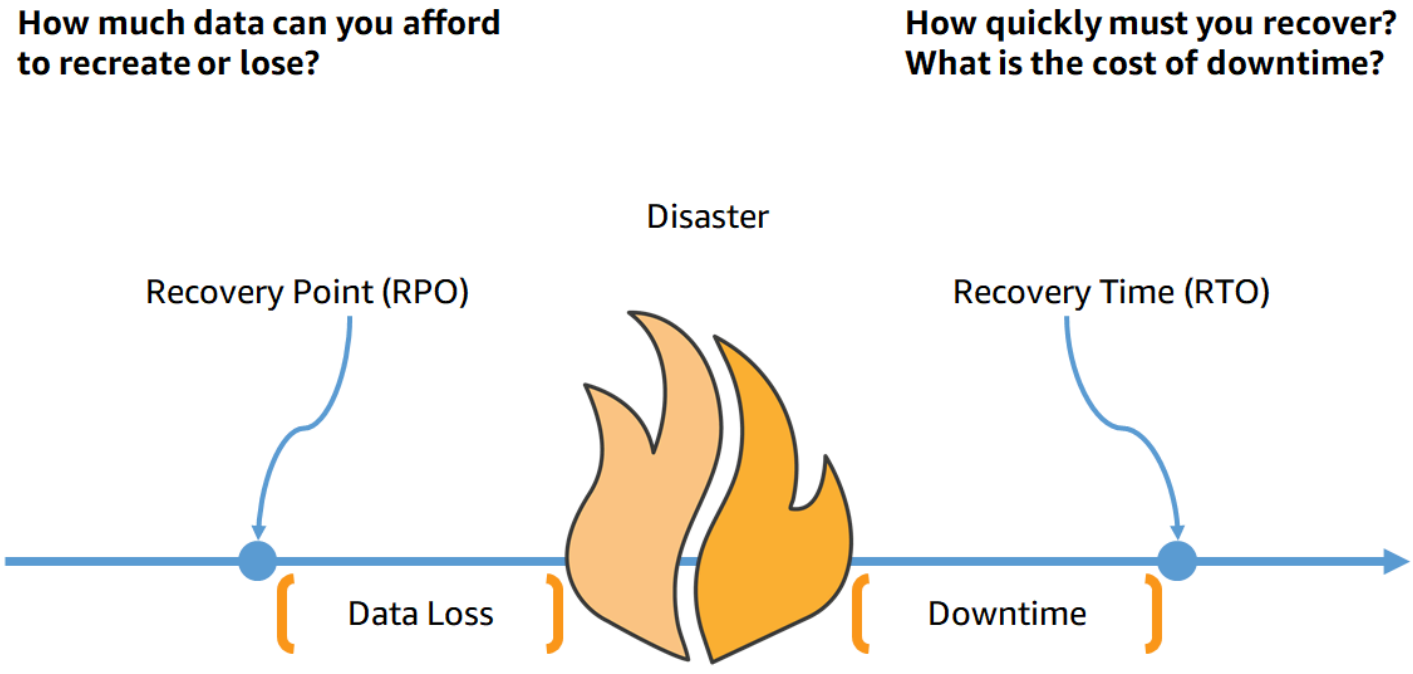

Figure 3 - Objectifs de reprise

L'objectif de délai de reprise (RTO) correspond au délai maximum acceptable entre l'interruption du service et la restauration du service. Cet objectif détermine ce qui est considéré comme un créneau acceptable d'indisponibilité du service. Il est défini par l'organisation.

Quatre stratégies de reprise après sinistre sont abordées dans ce document : sauvegarde et restauration, environnement de veille, secours à chaud et mode actif/actif multi-site (voir Options de reprise après sinistre dans le cloud). Dans le diagramme suivant, l'entreprise a déterminé son RTO maximal autorisé, ainsi que le montant limite qu'elle peut dépenser pour sa stratégie de restauration de service. Compte tenu des objectifs de l'entreprise, la stratégie de reprise après sinistre Environnement de veille ou la stratégie Secours à chaud répond à la fois aux critères de RTO et de coût.

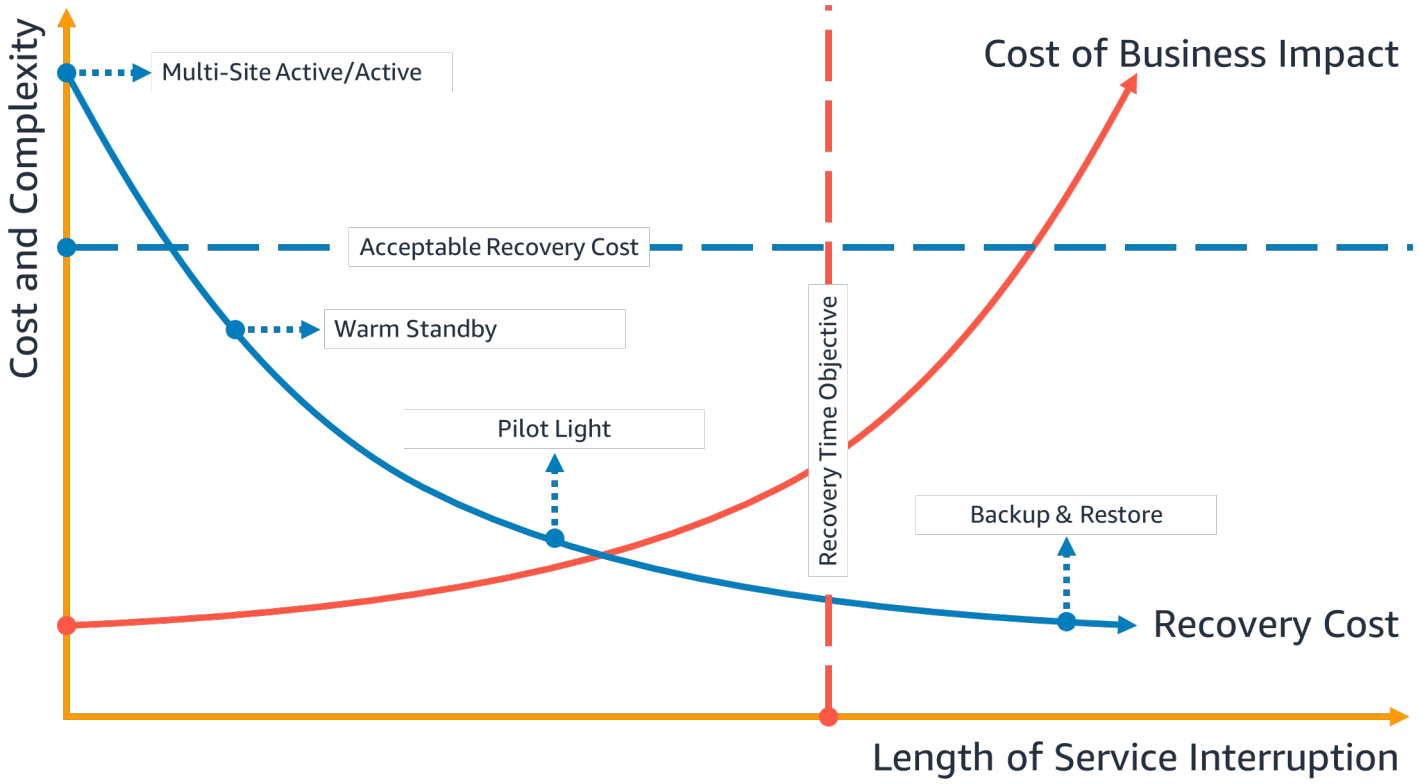

Figure 4 : objectif de délai de reprise

L'objectif de point de reprise (RPO) correspond au temps maximal acceptable depuis le dernier point de reprise des données. Cet objectif détermine ce qui est considéré comme étant une perte de données acceptable entre le dernier point de reprise et l'interruption du service. Il est défini par l'organisation.

Dans le diagramme suivant, l'entreprise a déterminé son RPO maximal autorisé, ainsi que le montant limite qu'elle peut dépenser pour sa stratégie de récupération des données. Parmi les quatre

stratégies de reprise après sinistre, la stratégie Environnement de veille ou la stratégie Secours à chaud répond à la fois aux critères de RPO et de coût.

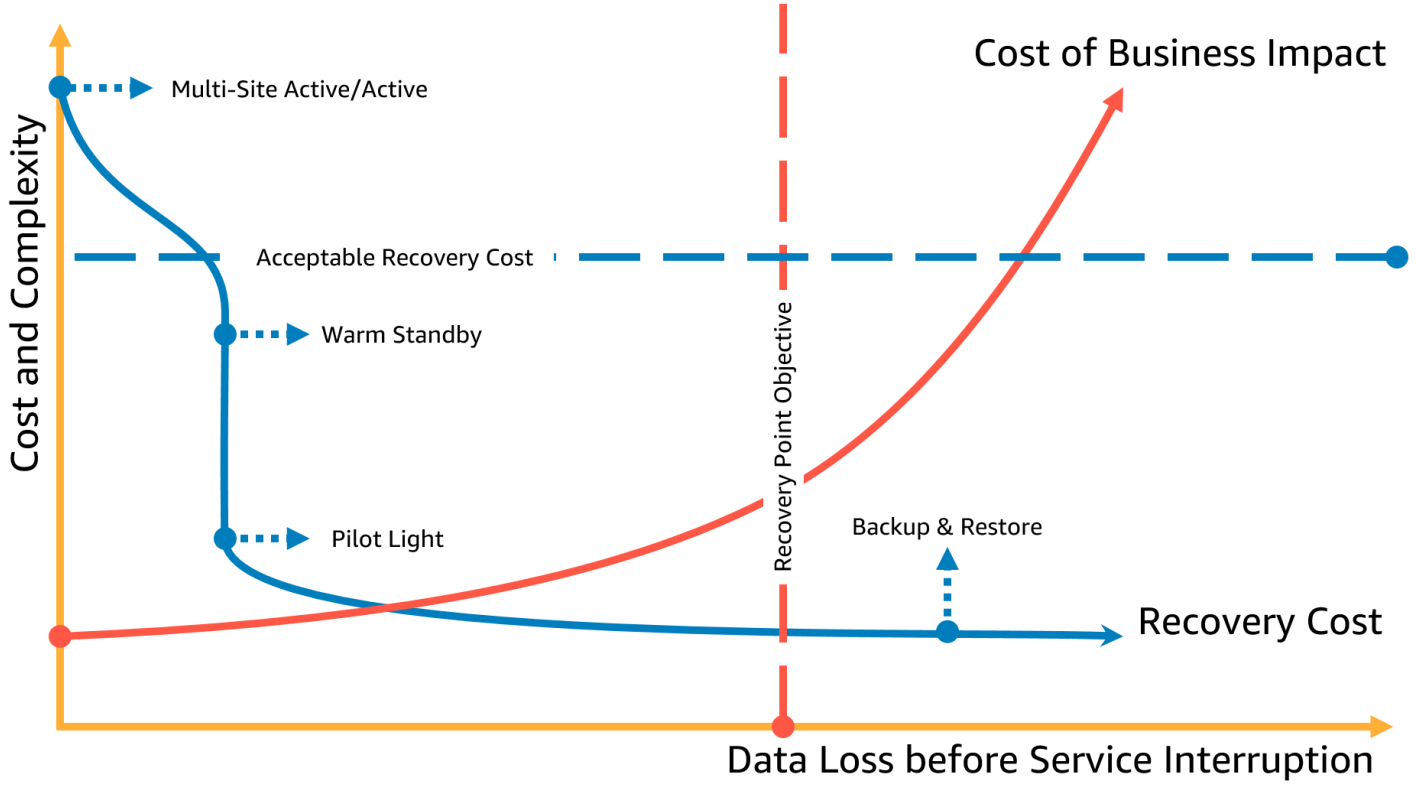

Figure 5 - Objectif de point de reprise

> **Note**
>
> Si le coût de la reprise est supérieur au coût de la défaillance ou de la perte, l'option de reprise ne doit pas être mise en place, sauf en cas de facteur secondaire tel que des exigences réglementaires.

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 opérationnel, 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 :

- Reprise rapide après un sinistre tout en profitant d'une complexité réduite
- Des tests simples et répétables vous permettent de tester plus facilement et plus fréquemment
- Des frais généraux de gestion réduits limitent la charge opérationnelle
- Les possibilités d'automatisation limitent les risques d'erreur et améliorent le délai de reprise

AWS vous permet de remplacer les dépenses en capital fixe d'un centre de données de sauvegarde physique par les dépenses d'exploitation variables d'un environnement correctement dimensionné dans le cloud, ce qui permet de réduire considérablement les coûts.

Pour de nombreuses organisations, la reprise après sinistre sur site était basée sur le risque d'interruption d'une charge de travail ou de charges de travail dans un centre de données et sur la récupération de données sauvegardées ou répliquées vers un centre de données secondaire. Lorsque les organisations déploient des charges de travail sur AWS, elles peuvent implémenter une charge de travail bien structurée et s'appuyer sur la conception de l'infrastructure cloud mondiale AWS pour atténuer l'effet de telles perturbations. Consultez le livre blanc AWS Well-Architected Framework - pilier de fiabilité pour obtenir de plus amples informations sur les bonnes pratiques architecturales pour la conception et l'exploitation de charges de travail fiables, sécurisées, efficaces et rentables dans le cloud.

Si vos charges de travail se trouvent 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 des pannes (chacune composée d'un ou de plusieurs centres de données distincts).

Région AWS unique

En cas de sinistre lié à la perturbation ou à la perte d'un centre de données physique, l'implémentation d'une charge de travail hautement disponible dans plusieurs zones de disponibilité au sein d'une même région AWS permet d'atténuer les risques de catastrophes naturelles et techniques. Cela permet également de réduire le risque de menaces humaines telles que les erreurs ou les activités non autorisées pouvant entraîner une perte de données. Chaque région AWS est composée de plusieurs zones de disponibilité, chacune étant isolée des défaillances des autres zones. Chaque zone de disponibilité se compose à son tour de plusieurs centres de données physiques. Vous pouvez donc partitionner des charges de travail sur plusieurs zones de disponibilité dans la même région afin de mieux isoler les problèmes éventuels et d'atteindre une meilleure disponibilité. Les zones de disponibilité offrent une redondance physique et fournissent de la résilience. Cela permet d'obtenir des performances sans interruptions, même dans le cas de pannes de courant, de coupures Internet, d'inondations et d'autres catastrophes naturelles. Consultez Infrastructure cloud mondiale AWS pour en savoir plus.

En déployant sur plusieurs zones de disponibilité d'une même région AWS, votre charge de travail est mieux protégée contre les défaillances d'un seul (voire de plusieurs) centres de données. Dans le cas d'un déploiement dans une seule région, vous pouvez sauvegarder les données et la configuration (y compris la définition de l'infrastructure) dans une autre région, afin d'obtenir des garanties supplémentaires. Cette stratégie réduit la portée de votre plan de reprise après sinistre afin de n'inclure que la sauvegarde et la restauration des données. Profiter de la résilience multi-régions en effectuant une sauvegarde dans une autre région AWS est une solution simple et peu coûteuse, par rapport aux autres options multi-régions décrites dans la section suivante. Par exemple, la sauvegarde sur Amazon Simple Storage Service (Amazon S3) vous donne accès à la récupération immédiate de vos données. Toutefois, si les exigences de votre stratégie de reprise après sinistre pour certaines de vos données sont plus souples concernant le délai de reprise (quelques heures plutôt que quelques minutes), l'utilisation d'Amazon S3 Glacier ou d'Amazon S3 Glacier Deep Archive réduira considérablement les coûts de votre stratégie de sauvegarde et de reprise.

Certaines charges de travail peuvent avoir des exigences réglementaires concernant la résidence des données. Si cela concerne votre charge de travail dans une localité qui ne possède 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 utiliser les zones de disponibilité au sein de cette région en tant qu'emplacements distincts. Cela peut s'avérer 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 implémentées à l'aide de zones de disponibilité au lieu de régions.

Régions AWS multiples

En cas de sinistre qui inclut le risque de perte de plusieurs centres de données situés à une distance significative les uns des autres, vous devez envisager des options de reprise après sinistre pour atténuer les risques naturels et techniques qui affectent une région entière au sein d'AWS. Toutes les options décrites dans les sections suivantes peuvent être implémentées en tant qu'architectures multi-régions afin de se protéger contre de tels désastres.

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 approches, allant du faible coût et de la faible complexité des sauvegardes à des stratégies plus complexes utilisant plusieurs régions actives. Il est essentiel de tester régulièrement votre stratégie de reprise après sinistre afin de pouvoir l'invoquer en toute confiance, le cas échéant.

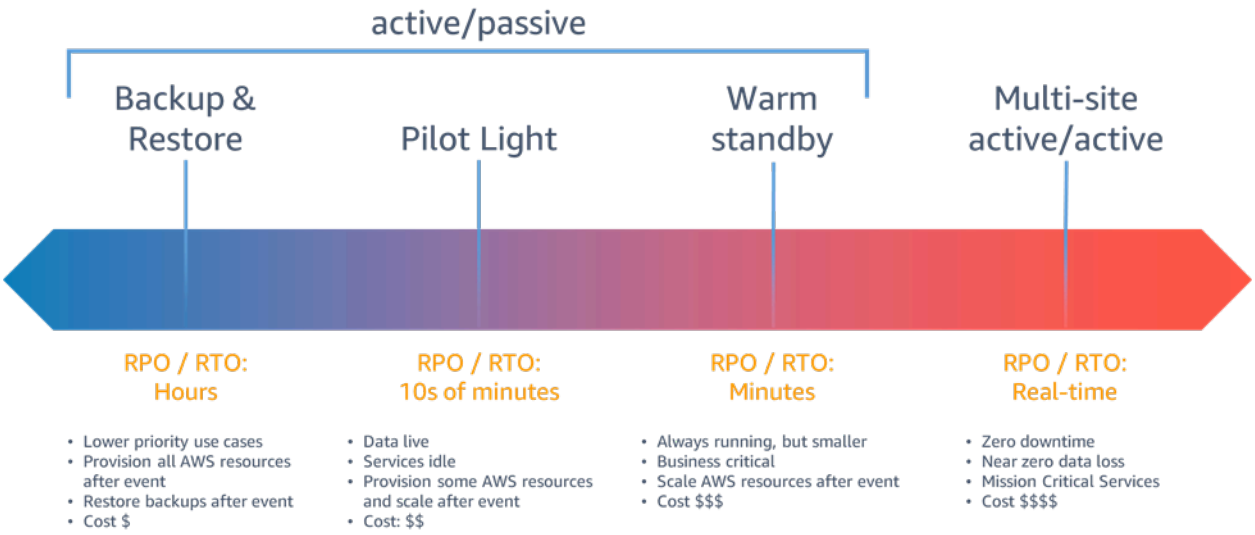

Figure 6 - Stratégies de reprise après sinistre

En cas de sinistre lié à la perturbation ou à la perte d'un centre de données physique pour une charge de travail hautement disponible et bien structurée, il se peut que vous ayez uniquement besoin d'une approche de sauvegarde et de restauration pour la reprise après sinistre. Si votre définition d'un sinistre implique plus que la perturbation ou la perte d'un centre de données physique, par exemple à l'échelle d'une région, ou si vous êtes soumis à des exigences réglementaires qui l'exigent, vous devriez envisager uns stratégie Environnement de veille, Secours à chaud ou Actif/actif multi-site.

Sauvegarde et restauration

La sauvegarde et la restauration constituent une approche appropriée pour atténuer les risques de perte ou de corruption des données. Cette approche peut également être utilisée pour atténuer un sinistre régional en répliquant les données vers d'autres régions AWS, ou pour atténuer 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 d'application dans la région de reprise. Pour permettre le redéploiement rapide et sans erreur de l'infrastructure, vous

devez toujours déployer à l'aide d'Infrastructure as code (IaC) par le biais de services tels que AWS CloudFormation ou AWS Cloud Development Kit (AWS CDK). Sans l'IaC, restaurer les charges de travail dans la région de reprise peut s'avérer complexe. Cela entraînera une augmentation des délais de reprise et éventuellement un dépassement de votre RTO. Outre les données utilisateur, veillez à sauvegarder le code et la configuration, y compris les images Amazon Machine Image (AMI) que vous utilisez pour créer des instances Amazon EC2. Vous pouvez utiliser AWS CodePipeline pour automatiser le redéploiement du code et de la configuration de l'application.

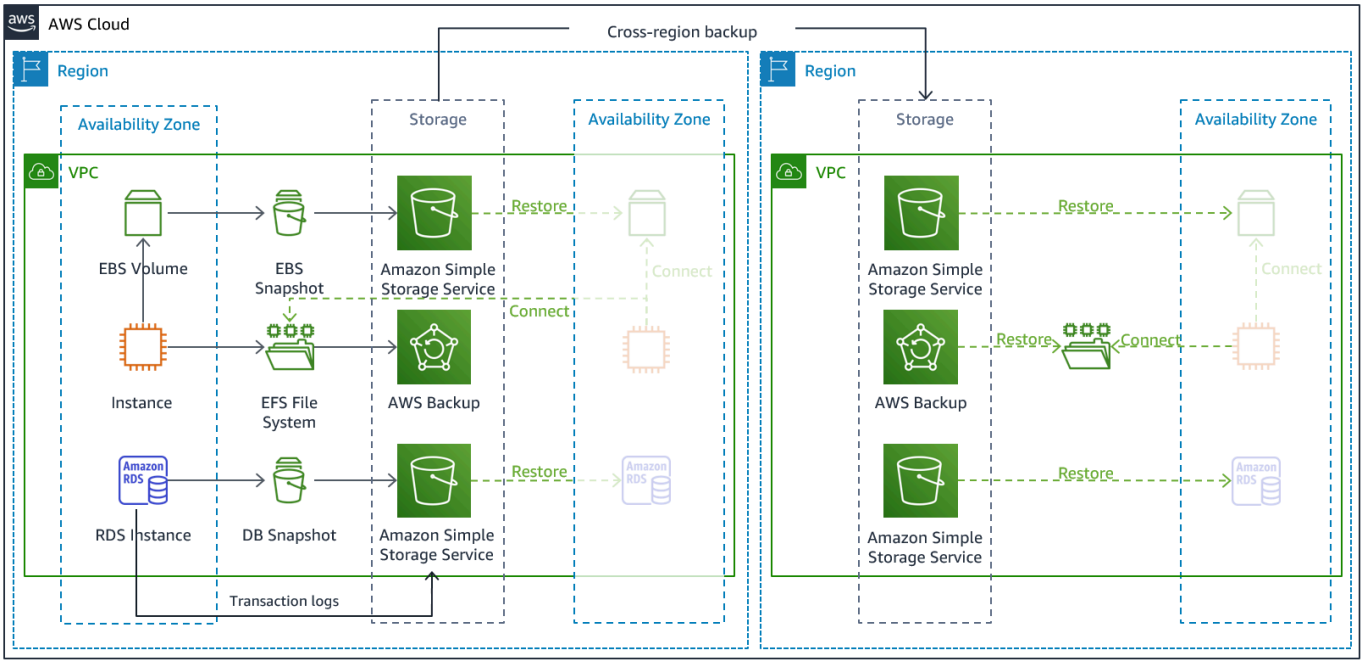

Figure 7 - Architecture de sauvegarde et restauration

Services AWS

Les données de votre charge de travail nécessiteront une stratégie de sauvegarde qui s'exécute périodiquement ou qui est continue. La fréquence à laquelle vous exécutez votre sauvegarde déterminera votre point de reprise réaliste (qui doit correspondre à votre RPO). La sauvegarde doit également permettre de restaurer au moment de la récupération. La sauvegarde avec restauration à un instant dans le passé est disponible via les services et ressources suivants :

- Instantané Amazon Elastic Block Store (Amazon EBS)
- Sauvegarde Amazon DynamoDB
- Instantané Amazon RDS
- instantané de bases de données Amazon Aurora

- Sauvegarde Amazon EFS (lors de l'utilisation d'AWS Backup)
- Instantané Amazon Redshift
- Instantané Amazon Neptune

Pour Amazon Simple Storage Service (Amazon S3), vous pouvez utiliser la réplication entre régions Amazon S3 (CRR) pour copier de manière asynchrone des objets vers un compartiment S3 dans la région de reprise après sinistre en continu, tout en fournissant une gestion des versions pour les objets stockés afin que vous puissiez choisir votre point de restauration. La réplication continue des données présente l'avantage d'offrir le délai le plus court (proche de zéro) pour sauvegarder vos données, mais peut ne pas protéger contre les catastrophes telles que la corruption de données ou les attaques malveillantes (telles que la suppression non autorisée de données) ainsi que les sauvegardes à un instant donné. La réplication continue est abordée dans la section Services AWS pour la stratégie Environnement de veille.

AWS Backup fournit un emplacement centralisé pour configurer, planifier et surveiller les capacités de sauvegarde AWS pour les services et ressources suivants :

- Volumes Amazon Elastic Block Store (Amazon EBS)
- Instances Amazon EC2
- Bases de données Amazon Relational Database Service (Amazon RDS) (y compris les bases de données Amazon Aurora)
- Tables Amazon DynamoDB
- Systèmes de fichiers Amazon Elastic File System (Amazon EFS)
- Volumes AWS Storage Gateway
- Amazon FSx for Windows File Server et Amazon FSx for Lustre

AWS Backup prend en charge la copie des sauvegardes entre les 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. La gestion des versions des objets protège vos données dans S3 contre les conséquences des actions de suppression ou de modification en conservant la version d'origine avant l'action. La gestion des versions d'objets peut être utile pour atténuer les catastrophes provenant d'erreurs humaines. Si vous utilisez la réplication S3 pour sauvegarder des données dans votre région de reprise après sinistre, alors, lorsqu'un objet est supprimé dans le compartiment

source, Amazon S3 ajoute par défaut un marqueur de suppression dans le compartiment source uniquement. Cette approche protège les données de la région de reprise après sinistre 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 délai de reprise (RTO). AWS CloudFormation fournit l'Infrastructure as Code (IaC) et vous permet de définir toutes les ressources AWS de votre charge de travail afin que vous puissiez déployer et redéployer de manière fiable vers plusieurs comptes AWS et régions AWS. Vous pouvez sauvegarder les instances Amazon EC2 utilisées par votre charge de travail en tant qu'images Amazon Machine Image (AMI). L'AMI est créée à partir d'instantanés du volume racine de votre instance et de tout autre volume EBS associé à votre instance. Vous pouvez utiliser cette AMI pour lancer une version restaurée de l'instance EC2. Une AMI peut être copiée au sein ou entre plusieurs régions. Vous pouvez également utiliser AWS Backup pour copier des sauvegardes entre des comptes et vers d'autres régions AWS. La fonctionnalité de sauvegarde entre comptes contribue à la protection contre les sinistres tels que les menaces internes ou la compromission de compte. AWS Backup ajoute également des fonctionnalités supplémentaires pour la sauvegarde EC2, en plus des volumes EBS individuels de l'instance, AWS Backup stocke et suit également les métadonnées suivantes : type d'instance, cloud privé virtuel (VPC) configuré, groupe de sécurité, rôle IAM, configuration de surveillance et étiquettes. Toutefois, ces métadonnées supplémentaires ne sont utilisées que lors de la restauration de la sauvegarde EC2 dans la même région AWS.

Toutes les données stockées dans la région de reprise après sinistre en tant que sauvegardes doivent être restaurées au moment du basculement. AWS Backup offre une fonction de restauration, mais ne permet pas actuellement la restauration planifiée ou automatique. Vous pouvez implémenter une restauration automatique dans la région de reprise après sinistre à l'aide du kit SDK AWS afin d'appeler des API pour AWS Backup. Vous pouvez configurer la restauration en tant que tâche récurrente 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 de AWS Lambda.

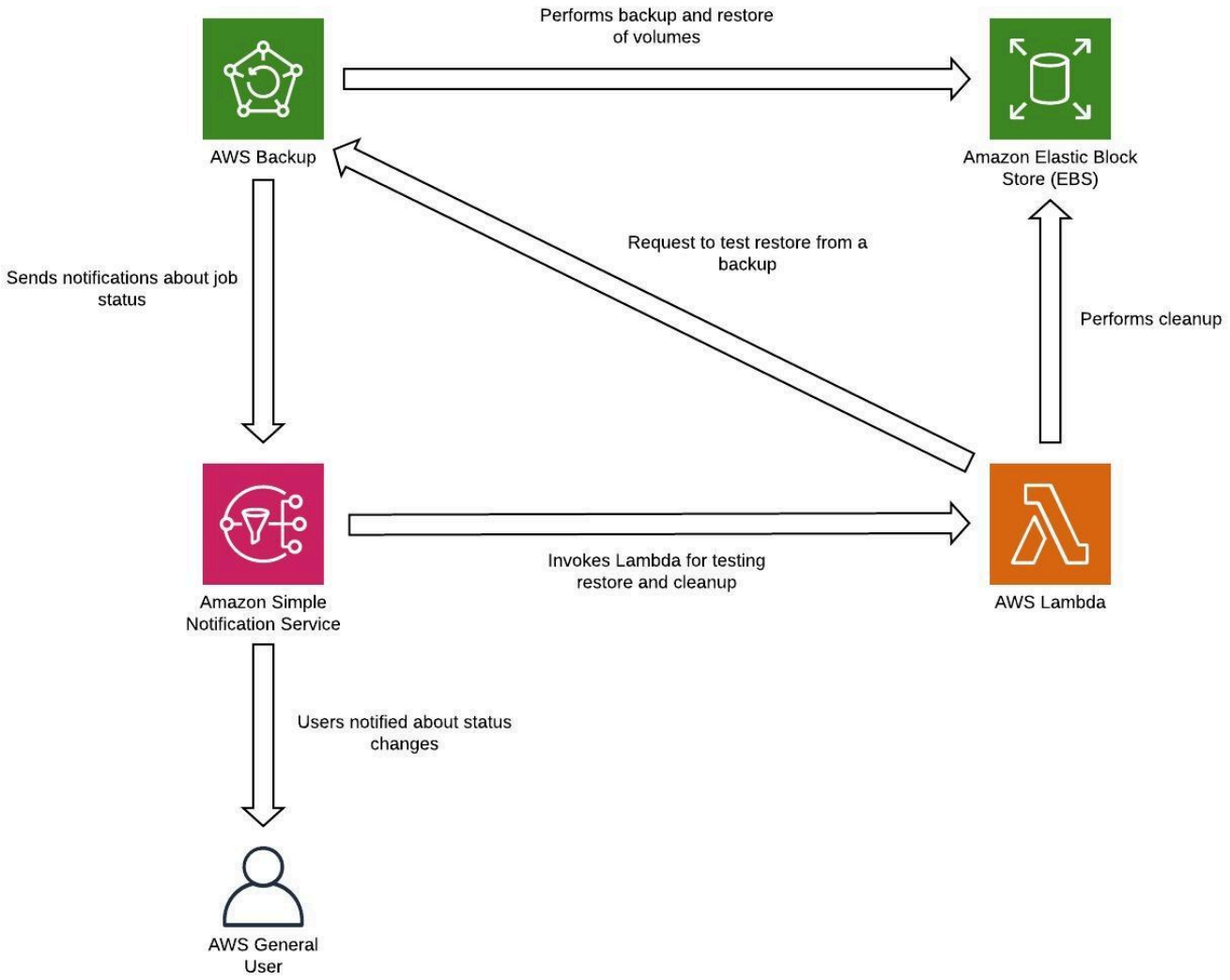

Figure 8 - restauration et test des sauvegardes

> **Note**
>
> Votre stratégie de sauvegarde doit inclure le test de vos sauvegardes. Pour de plus amples informations, veuillez consulter la section Test de reprise après sinistre. Pour obtenir une démonstration pratique de l'implémentation, veuillez consulter Ateliers AWS Well-Architected : Test de la sauvegarde et de la restauration des données.

Environnement de veille

Avec l'approche Environnement de veille, vous répliquez vos données d'une région à une autre et mettez en service une copie de votre infrastructure de charge de travail principale. Les ressources nécessaires pour prendre en charge la réplication et la sauvegarde des données, telles que les bases de données et le stockage d'objets, sont toujours disponibles. D'autres éléments, tels que les serveurs d'applications, sont chargés avec le code d'application et les configurations, mais sont désactivés, et ne sont utilisés que pendant les tests ou lorsque le basculement de reprise après sinistre est invoqué. Contrairement à l'approche de sauvegarde et de restauration, votre infrastructure principale est toujours disponible et vous avez toujours la possibilité de provisionner rapidement un environnement de production à grande échelle en utilisant et en procédant à une évolutivité horizontale de vos serveurs d'applications.

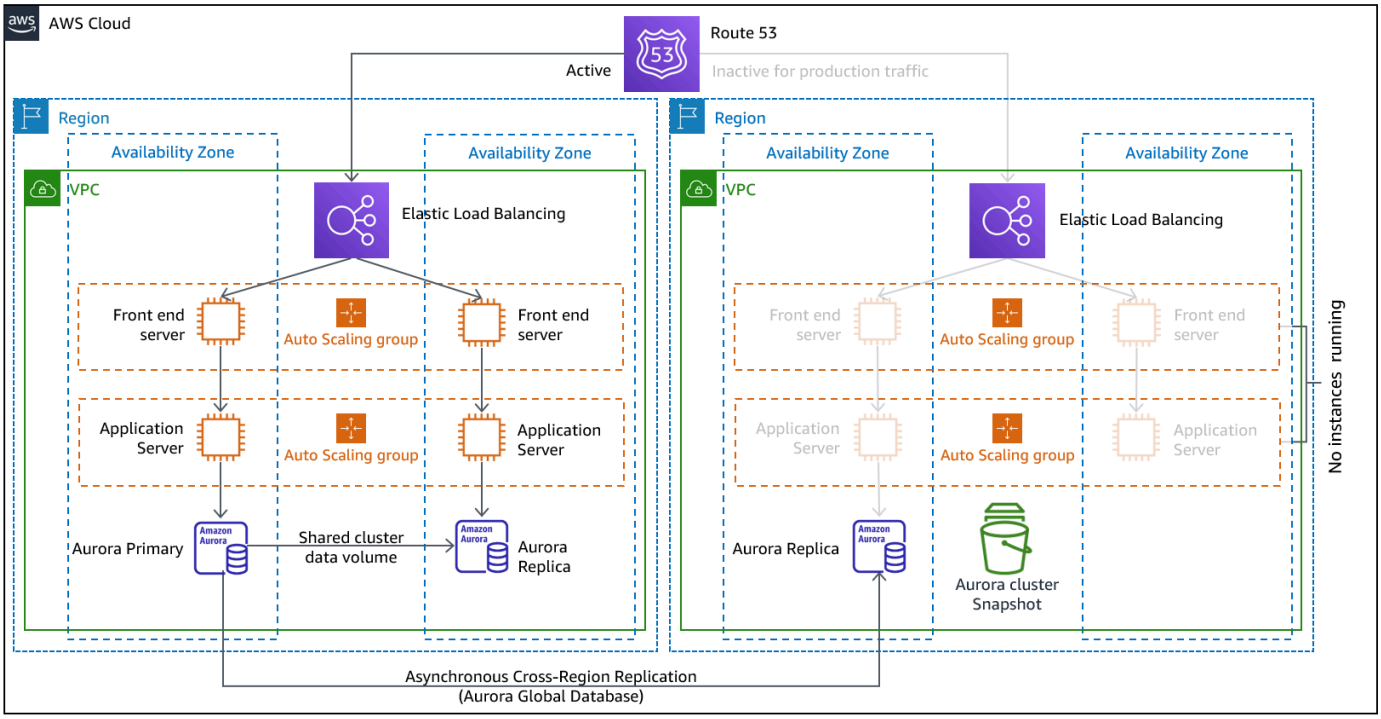

Figure 9 - Environnement de veille

Une approche Environnement de veille réduit le coût permanent de la reprise après sinistre en réduisant les ressources actives, et simplifie la reprise en cas de sinistre, car les exigences de l'infrastructure de base sont toutes en place. Cette option de récupération vous oblige à modifier votre approche du déploiement. Vous devez apporter des modifications à l'infrastructure principale de chaque région et déployer les modifications de la charge de travail (configuration, code) simultanément dans chaque région. Cette étape peut être simplifiée en automatisant vos

déploiements et en utilisant l'Infrastructure as code (IaC) pour déployer l'infrastructure sur plusieurs comptes et régions (déploiement complet de l'infrastructure dans la région principale et déploiement d'infrastructure réduit/désactivé vers les régions de reprise après sinistre). Il est recommandé d'utiliser un compte différent par région pour fournir le plus haut niveau d'isolation des ressources et de sécurité (dans le cas où les informations d'identification compromises font également partie de vos plans de reprise après sinistre).

Avec cette approche, vous devez également atténuer les risques liés aux données en cas de sinistre. La réplication continue des données vous protège contre certains types de sinistre, mais elle peut ne pas vous protéger contre la corruption ou la destruction des données, sauf si votre stratégie inclut également la gestion des versions des données stockées ou des options de restauration à un instant dans le passé. Vous pouvez sauvegarder les données répliquées dans la région sinistrée pour créer des sauvegardes à un instant donné dans cette même région.

Services AWS

En plus d'utiliser les services AWS abordés dans la section Sauvegarde et restauration pour créer des sauvegardes à un instant donné, pensez également aux services suivants pour votre stratégie Environnement de veille.

Pour la stratégie Environnement de veille, 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 sauvegardes à un instant donné évoquées précédemment). AWS fournit une réplication entre régions continue et asynchrone pour les données à l'aide des services et des ressources suivants :

- Réplication Amazon Simple Storage Service (Amazon S3)
- Réplicas en lecture Amazon RDS
- Bases de données Amazon Aurora Global Database
- Tables globales Amazon DynamoDB

Avec la réplication continue, les versions de vos données sont disponibles presque immédiatement dans votre région de reprise après sinistre. Les temps de réplication réels peuvent être surveillés à l'aide de fonctionnalités de service telles que le contrôle du délai de réplication S3 (S3 RTC) pour les objets S3 et les fonctionnalités de gestion des bases de données Amazon Aurora Global Database.

Lors du basculement destiné à exécuter votre application en lecture/écriture à partir de la région de reprise après sinistre, vous devez promouvoir un réplica en lecture RDS pour qu'il devienne l'instance

principale. Pour les instances de bases de données autres qu'Aurora, le processus dure quelques minutes et le redémarrage fait partie du processus. Pour la réplication entre régions (CRR) et le basculement avec RDS, l'utilisation d'Amazon Aurora Global Database présente plusieurs avantages. La base de données globale utilise une infrastructure dédiée qui laisse vos bases entièrement disponibles pour servir votre application et peut être répliquée dans une région secondaire avec une latence inférieure à une seconde (et beaucoup moins que 100 millisecondes au sein d'une région AWS). Avec Amazon Aurora Global Database, si votre région principale subit une dégradation des performances ou une panne, vous pouvez promouvoir l'une des régions secondaires afin qu'elle prenne des responsabilités en lecture/écriture en moins d'une minute, même en cas de panne régionale complète. La promotion peut être automatique et il n'y a pas de redémarrage.

Une version réduite de votre infrastructure de charge de travail principale avec moins de ressources ou des ressources plus petites doit être déployée dans votre région de reprise après sinistre. À l'aide d'AWS CloudFormation, vous pouvez définir votre infrastructure et la déployer de manière cohérente sur les comptes AWS et les régions AWS. AWS CloudFormation utilise des pseudo-paramètres prédéfinis pour identifier le compte AWS et la région AWS dans laquelle il est déployé. Par conséquent, vous pouvez implémenter une logique de condition dans vos modèles CloudFormation pour déployer uniquement la version réduite de votre infrastructure dans la région de reprise après sinistre. Pour les déploiements d'instance EC2, une image Amazon Machine Image (AMI) fournit des informations concernant par exemple la configuration matérielle et les logiciels installés. Vous pouvez implémenter un pipeline Image Builder qui crée les AMI dont vous avez besoin et les copier dans votre région principale et dans votre région de sauvegarde. Cela permet de garantir que ces AMI finales disposent de tout ce dont vous avez besoin, en cas de sinistre, pour redéployer ou faire monter en puissance votre charge de travail dans une nouvelle région. Les instances Amazon EC2 sont déployées dans une configuration réduite (qui comporte moins d'instances que dans votre région principale). Vous pouvez utiliser la mise en veille prolongée pour placer les instances EC2 dans un état d'arrêt, ce qui vous permet de ne pas payer de frais EC2, mais uniquement le stockage utilisé. Pour démarrer des instances EC2, vous pouvez créer des scripts à l'aide de l'interface de ligne de commande (CLI) AWS ou du kit SDK AWS. Pour augmenter la capacité de l'infrastructure afin de prendre en charge le trafic de production, veuillez consulter AWS Auto Scaling dans la section Secours à chaud.

Pour une configuration Actif/Veille telle que l'Environnement de veille, l'ensemble du 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. Vous avez le choix entre deux options de gestion du trafic concernant les services AWS. La première option 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 avec un nom de domaine Route 53. Vous pouvez ensuite acheminer le trafic vers le point de terminaison approprié sous ce nom de domaine. Les surveillances de l'état Amazon Route 53 surveillent ces points de terminaison. À l'aide de ces surveillances de l'état, vous pouvez configurer le basculement DNS pour vous assurer que le trafic est envoyé vers des points de terminaison en bon état.

La deuxième option consiste à utiliser AWS Global Accelerator. À l'aide d'AnyCast IP, vous pouvez associer plusieurs points de terminaison dans une ou plusieurs régions AWS avec la ou les mêmes adresses IP statiques. AWS Global Accelerator achemine ensuite le trafic vers le point de terminaison approprié associé à cette adresse. Les surveillances de l'état Global Accelerator surveillent les points de terminaison. À l'aide de ces surveillances de l'état, AWS Global Accelerator vérifie automatiquement l'état de vos applications et achemine le trafic utilisateur uniquement vers un point de terminaison d'application en bon état. Global Accelerator offre des latences plus faibles au point de terminaison de l'application, car il utilise le vaste réseau périphérique AWS pour placer le trafic sur le réseau principal AWS dès que possible. Global Accelerator évite également les problèmes de mise en cache qui peuvent survenir avec les systèmes DNS (comme Route 53).

CloudEndure Disaster Recovery

CloudEndure Disaster Recovery, disponible sur AWS Marketplace, réplique en continu les applications hébergées sur le serveur et les bases de données hébergées sur le serveur depuis n'importe quelle source vers AWS en utilisant la réplication au niveau bloc du serveur sous-jacent. CloudEndure Disaster Recovery vous permet d'utiliser le Cloud AWS en tant que région de reprise après sinistre pour une charge de travail sur site et son environnement. Il peut également être utilisé pour la reprise après sinistre de charges de travail hébergées par AWS si elles se composent uniquement d'applications et de bases de données hébergées sur EC2 (et non sur RDS). CloudEndure Disaster Recovery utilise la stratégie Environnement de veille, qui gère une copie des données et des ressources désactivées dans un Amazon Virtual Private Cloud (Amazon VPC) utilisé comme zone de transit. Lorsqu'un événement de basculement est déclenché, les ressources mises en transit sont utilisées pour créer automatiquement un déploiement pleine capacité dans l'Amazon VPC cible utilisé comme emplacement de reprise.

Figure 10 - Architecture CloudEndure Disaster Recovery

Secours à chaud

L'approche Secours à chaud implique de s'assurer qu'il existe une copie réduite, mais entièrement fonctionnelle, de votre environnement de production dans une autre région. Cette approche étend le concept d'Environnement de veille et réduit le délai de reprise, car votre charge de travail est toujours active dans une autre région. Cette approche vous permet également d'effectuer plus facilement des tests ou d'implémenter des tests continus afin d'augmenter la confiance dans votre capacité de reprise après sinistre.

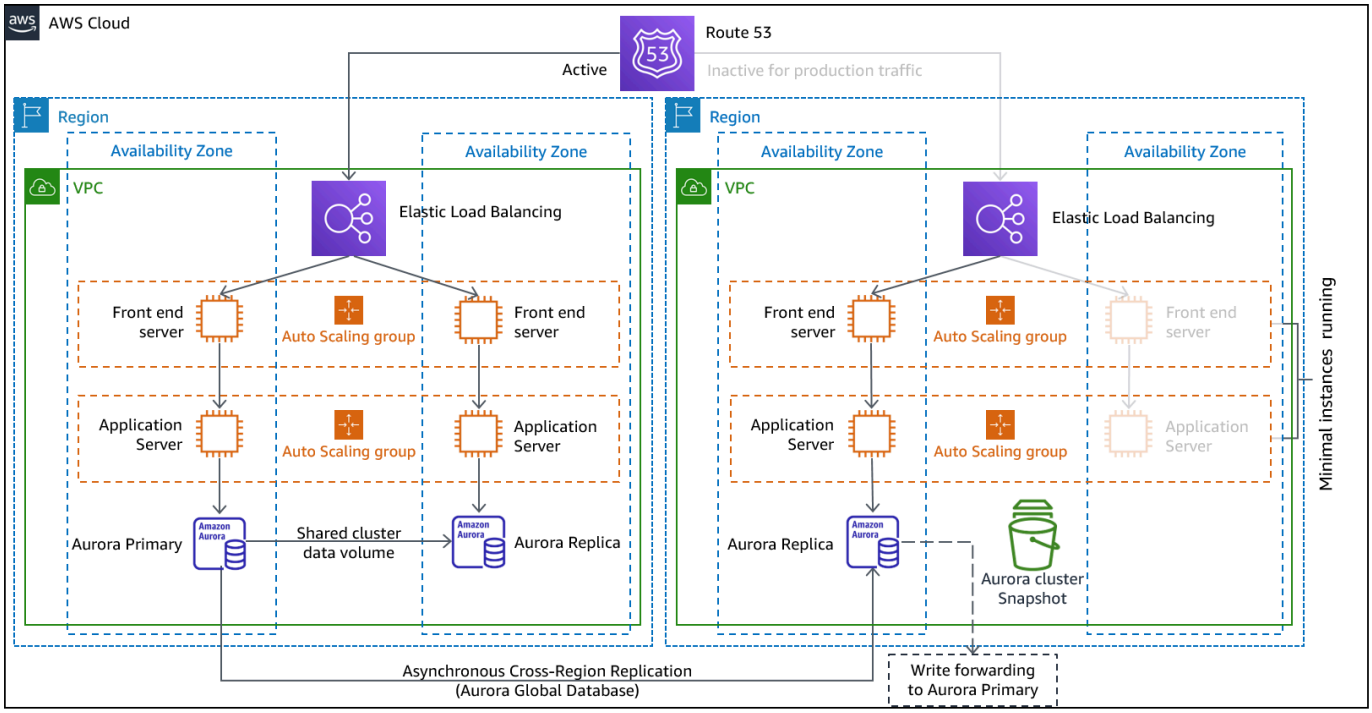

Figure 11 - Architecture Secours à chaud

Remarque : La différence entre l'Environnement de veille et le Secours à chaud peut parfois être difficile à comprendre. Les deux architectures incluent un environnement dans votre région de reprise après sinistre avec des copies des actifs de votre région principale. La différence réside dans le fait que l'Environnement de veille ne peut pas traiter les demandes sans que des mesures supplémentaires ne soient prises au préalable, tandis que le Secours à chaud peut traiter le trafic (à des niveaux de capacité réduits) immédiatement. L'approche Environnement de veille nécessite de « démarrer » les serveurs, déployer éventuellement une infrastructure (non principale) supplémentaire et d'augmenter la charge, tandis que l'approche Secours à chaud ne nécessite que d'augmenter la charge (tout est déjà déployé et en cours d'exécution). Choisissez l'une ou l'autre approche en fonction de vos besoins en termes de RTO et de RPO.

Services AWS

Tous les services AWS couverts par la sauvegarde et la restauration et l'Environnement de veille sont également utilisés dans l'approche Secours à chaud pour la sauvegarde des données, la réplication des données, le routage du trafic actif/en veille et le déploiement de l'infrastructure, y compris les instances EC2.

AWS Auto Scaling est utilisé pour mettre à l'échelle les ressources, notamment les instances Amazon EC2, les tâches Amazon ECS, le débit Amazon DynamoDB et les réplicas Amazon Aurora au sein d'une région AWS. Amazon EC2 Auto Scaling met à l'échelle le déploiement de l'instance EC2 dans les zones de disponibilité d'une région AWS, offrant ainsi une résilience au sein de cette région. Utilisez Auto Scaling pour faire monter en puissance votre région de reprise après sinistre à sa pleine capacité de production, dans le cadre d'une stratégie d'Environnement de veille ou de Secours à chaud. Par exemple, pour EC2, augmentez le paramètre de capacité souhaitée dans le groupe Auto Scaling. Vous pouvez ajuster ce paramètre manuellement via AWS Management Console, automatiquement via le kit SDK AWS ou en redéployant votre modèle AWS CloudFormation à l'aide de la nouvelle valeur de capacité souhaitée. Vous pouvez utiliser des paramètres AWS CloudFormation pour faciliter le redéploiement du modèle CloudFormation. Assurez-vous que les quotas de service dans votre région de reprise après sinistre sont suffisamment élevés pour ne pas vous empêcher de procéder à une augmentation d'échelle de la capacité de production.

Mode actif/actif multi-site

Vous pouvez exécuter votre charge de travail simultanément dans plusieurs régions dans le cadre d'une stratégie actif/actif multi-site ou actif/passif de secours. La stratégie actif/actif multi-site dirige le trafic de toutes les régions vers lesquelles il est déployé, tandis que la stratégie de Secours à chaud ne traite que le trafic d'une seule région. Et les autres régions ne sont utilisées que pour la reprise après sinistre. Avec une approche actif/actif multi-site, les utilisateurs peuvent accéder à votre charge de travail dans toutes les régions dans lesquelles elle est déployée. Cette approche est la plus complexe et la plus coûteuse en matière de reprise après sinistre, mais elle peut réduire votre délai de reprise à près de zéro pour la plupart des sinistres lorsque les choix technologiques et l'implémentation appropriés sont effectués (toutefois, la corruption des données peut nécessiter des sauvegardes, ce qui aboutit généralement à un point de reprise différent de zéro). Le mode Secours à chaud utilise une configuration active/passive dans laquelle les utilisateurs ne sont dirigés que vers une seule région et les régions de reprise après sinistre ne prennent pas de trafic. La plupart des clients considèrent que s'ils veulent créer un environnement complet dans la deuxième région, il est logique de l'utiliser en mode actif/actif. Si vous ne souhaitez pas utiliser les deux régions pour gérer

le trafic utilisateur, le secours à chaud offre une approche plus économique et moins complexe sur le plan opérationnel.

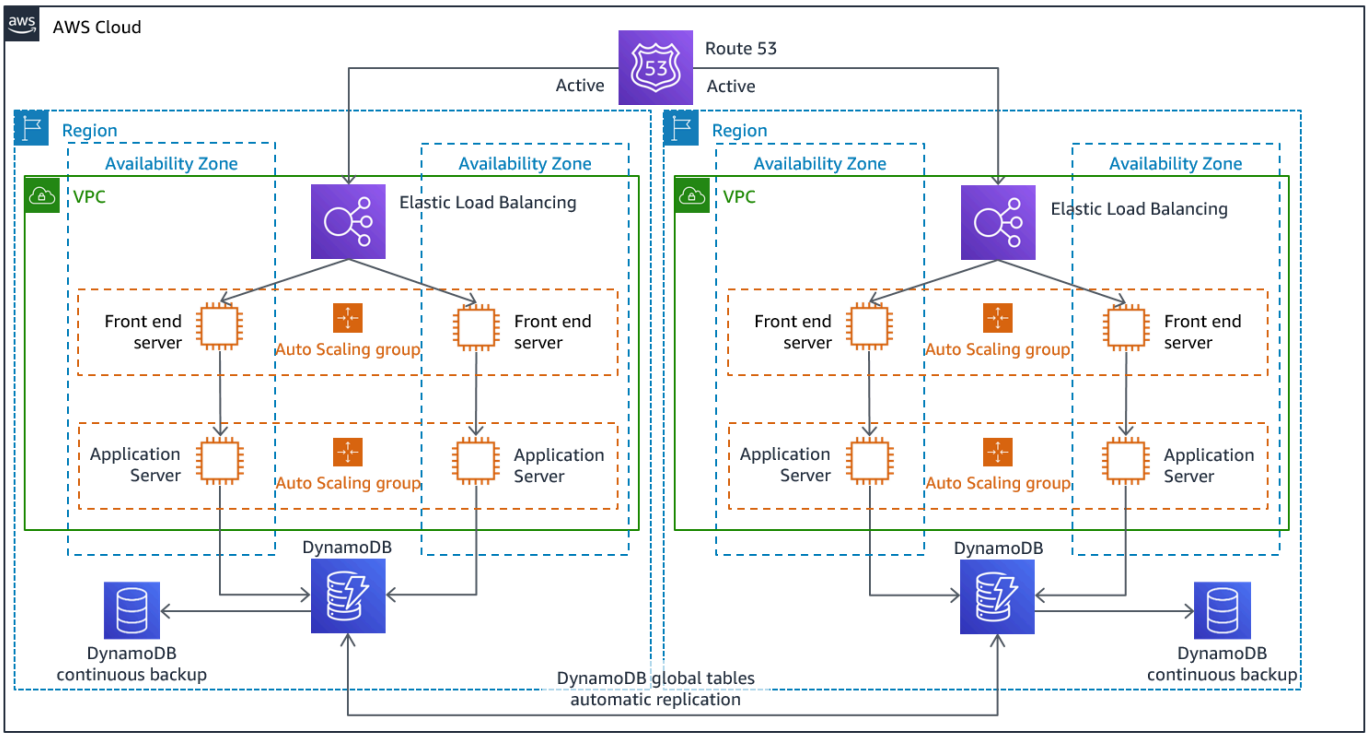

Figure 12 : architecture active/active multi-site (changer un chemin d'accès actif en chemin d'accès inactif pour le secours à chaud)

En mode actif/actif multi-site, étant donné que la charge de travail s'exécute dans plusieurs régions, il n'existe pas de basculement dans ce scénario. Dans ce cas, les tests de reprise après sinistre se concentrent sur la façon dont la charge de travail réagit à la perte d'une région : le trafic est-il acheminé hors de la région défaillante ? Les autres régions peuvent-elles gérer l'ensemble du trafic ? Il est également nécessaire de tester les données en cas de sinistre. La sauvegarde et la restauration sont toujours nécessaires et doivent être testées régulièrement. Il convient également de noter que les délais de reprise pour les données en cas de sinistre impliquant une corruption, une suppression ou une obfuscation des données seront toujours supérieurs à zéro et que le point de reprise sera toujours situé à un certain point avant la découverte du sinistre. Si la complexité et le coût supplémentaires d'une approche actif/actif multi-site (ou de secours à chaud) sont nécessaires pour maintenir des délais de reprise proches de zéro, des efforts supplémentaires doivent être faits pour maintenir la sécurité et prévenir les erreurs humaines afin d'atténuer les risques de catastrophe humaine.

Services AWS

Tous les services AWS couverts par les modes Sauvegarde et la restauration, Environnement de veille et Secours à chaud sont également utilisés ici pour la sauvegarde des données à un instant dans le passé, la réplication des données, le routage du trafic actif/actif, ainsi que le déploiement et la mise à l'échelle de l'infrastructure, y compris les instances EC2.

Pour les scénarios actifs/passifs évoqués précédemment (Environnement de veille et Secours à chaud), Amazon Route 53 et AWS Global Accelerator peuvent tous deux être utilisés pour acheminer le trafic réseau vers la région active. Pour la stratégie active/active exposée ici, ces deux services permettent également de définir des politiques qui déterminent quels utilisateurs se dirigent vers quel point de terminaison régional actif. Avec AWS Global Accelerator vous définissez un variateur de trafic pour contrôler le pourcentage de trafic dirigé vers chaque point de terminaison d'application. Amazon Route 53 prend en charge cette approche en pourcentage, ainsi que plusieurs autres politiques disponibles, notamment celles basées sur la proximité géographique et la latence. Global Accelerator exploite automatiquement le vaste réseau de serveurs Edge AWS pour intégrer le trafic vers le réseau principal AWS dès que possible, ce qui réduit les latences des demandes.

La réplication des données associée à cette stratégie permet d'atteindre un objectif de point de reprise proche de zéro. Les services AWS comme Amazon Aurora Global Database, utilisent une infrastructure dédiée qui laisse vos bases de données entièrement disponibles pour servir votre application et peuvent être répliqués dans une région secondaire avec, généralement, une latence inférieure à une seconde. Avec les stratégies en mode actif/passif, les écritures se produisent uniquement dans la région principale. La différence par rapport à la stratégie en mode actif/actif porte sur la conception de la gestion des écritures dans chaque région active. Il est courant de concevoir des lectures utilisateur qui seront acheminées depuis la région la plus proche, système appelé système de lecture local. Avec les écritures, plusieurs options s'offrent à vous :

- Une stratégie d'écriture globale achemine toutes les écritures vers une seule région. En cas d'échec de cette région, une autre région est promue pour accepter les écritures. La base de données globale Aurora convient parfaitement à l'écriture globale, car elle prend en charge la synchronisation avec des réplicas en lecture entre les régions, et vous pouvez promouvoir l'une des régions secondaires pour qu'elle prenne des responsabilités en lecture/écriture en moins d'une minute.
- Une stratégie d'écriture locale achemine les écritures vers la région la plus proche (tout comme les lectures). Les tables globales Amazon DynamoDB permettent ce type de stratégie, autorisant la lecture et l'écriture à partir de chaque région dans laquelle votre table globale est déployée. Les

tables globales Amazon DynamoDB utilisent un rapprochement last writer wins (dernière version valide) entre des mises à jour concomitantes.

- Une stratégie partitionnée en écriture attribue des écritures à une région spécifique en fonction d'une clé de partition (par exemple un ID utilisateur) pour 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. Lors de l'implémentation de cette approche, veillez à activer la synchronisation des modifications de réplica sur les compartiments A et B pour répliquer les modifications des métadonnées de réplica, telles que les listes de contrôle d'accès aux objets, les balises d'objet ou les verrous d'objet sur les objets répliqués. Vous pouvez également configurer la réplication ou la non réplication des marqueurs de suppression entre les compartiments de vos régions actives. Outre la réplication, votre stratégie doit inclure des sauvegardes à un instant dans le passé pour vous protéger contre les événements de corruption ou de destruction des données.

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

Détection

Il est important de savoir dès que possible que vos charges de travail ne génèrent pas les résultats opérationnels qu'elles devraient. Vous pouvez alors rapidement déclarer un sinistre et reprendre les activités suite à un incident. Pour fixer des objectifs de reprise efficaces, il est essentiel d'associer ce temps de réponse à des informations appropriées pour atteindre les objectifs de reprise. Si votre objectif de point de reprise est d'une heure, vous devez détecter l'incident, en informer le personnel approprié, engager vos processus d'alerte, évaluer les informations (le cas échéant) concernant le temps de reprise prévu (sans exécuter le plan de reprise après sinistre), déclarer un sinistre et récupérer dans un délai d'une heure.

> **Note**
>
> Si les parties prenantes décident de ne pas invoquer la reprise après sinistre même si le RTO risque de ne pas être respecté, vous devez réévaluer les plans et les objectifs de reprise après sinistre. La décision de ne pas invoquer les plans de reprise après sinistre peut être due à des plans inadaptés ou à un manque de confiance dans leur exécution.

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

AWS publie les dernières informations concernant la disponibilité de ses services sur le tableau de bord Service Health Dashboard. Consultez-le à tout moment pour obtenir des informations sur l'état actuel ou abonnez-vous à un flux RSS pour être informé des interruptions de chaque service. Si vous rencontrez un problème opérationnel en temps réel avec l'un de nos services qui n'apparaît pas sur le tableau de bord Service Health Dashboard, vous pouvez créer une demande de support.

Le AWS Health Dashboard fournit des informations sur les événements AWS Health pouvant 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 les exigences de RTO les plus strictes, vous pouvez implémenter un basculement automatique basé sur des surveillances de l'état. Concevez des surveillances de l'état représentatives de l'expérience utilisateur et basées sur des indicateurs de rendement clés. Les surveillances de

l'état approfondies exercent les fonctionnalités clés de votre charge de travail et vont au-delà des surveillances superficielles. Utilisez des surveillances de l'état approfondies basées sur plusieurs signaux. Vous devez être prudent avec cette approche, afin de ne pas déclencher de fausses alarmes. En effet, déclencher un basculement alors qu'il n'est pas nécessaire de le faire peut entraîner des risques de disponibilité.

Test de reprise après sinistre

Testez votre procédure de reprise après sinistre pour la valider et testez régulièrement le basculement vers la région de reprise après sinistre de votre charge de travail pour vous assurer que les objectifs RTO et RPO sont bien atteints.

Il convient d'éviter de développer des chemins de récupération 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 de l'instance secondaire, qui peut avoir été suffisante lors de votre dernier test, peut ne plus être en mesure de tolérer la charge dans le cadre de ce scénario, ou les quotas de service de la région secondaire peuvent ne pas être suffisants.

Notre expérience nous a montré que seul un chemin de reprise après erreur testé fréquemment fonctionne réellement. C'est pourquoi l'idéal est de n'avoir qu'un petit nombre de chemins de reprise.

Vous pouvez établir des modèles de reprise et tester ceux-ci régulièrement. Si vous avez un chemin de reprise complexe ou critique, vous devez toujours exécuter régulièrement cette panne en production pour vous assurer du bon fonctionnement de ce chemin de reprise.

Gestion de l'écart de configuration au niveau de la région de reprise après sinistre Assurez-vous que votre infrastructure, vos données et votre configuration sont conformes aux besoins de la région de reprise après sinistre. Par exemple, vérifiez que les AMI et les quotas de service sont à jour.

Vous pouvez utiliser AWS Config pour surveiller et enregistrer en continu vos configurations de ressources AWS. AWS Config peut détecter la dérive et déclencher AWS Systems Manager Automation pour corriger la dérive et déclencher des alarmes. AWS CloudFormation peut 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 d'avoir un plan de reprise après sinistre qui reflète cette définition et l'impact qu'elle peut avoir sur les résultats opérationnels. Créez un objectif de délai de reprise (RTO) et un objectif de point de reprise (RPO) en fonction d'une analyse d'impact et d'une évaluation des risques, puis choisissez l'architecture adaptée pour atténuer les risques. Assurez-vous que la détection des sinistres est possible et rapide ; il est essentiel de savoir quand les objectifs sont menacés. Assurez-vous d'avoir un plan et validez-le par des tests. Les plans de reprise après sinistre qui n'ont pas été validés risquent de ne pas être implémentés en raison d'un manque de confiance ou d'un échec dans la réalisation des objectifs de reprise après sinistre.

Participants

Ont contribué à la préparation du présent document :

- Alex Livingstone, Practice Lead Cloud Operations, AWS Enterprise Support
- Seth Eliot, architecte principal de solutions de fiabilité, Amazon Web Services

Autres lectures

Pour en savoir plus, voir :

- Pilier Fiabilité - AWS Well-Architected Framework
- Liste de contrôle du plan de reprise après sinistre
- Implémentation des surveillances de l'état
- Implémentations de solutions AWS : Multi-Region Application Architecture
- AWS re:Invent 2018 : modèles d'architecture des applications actives multi-régions (ARC209-R2)

Historique du document

Modification	Description	Date
Publication initiale	Première publication.	12 février 2021

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

Mentions légales

Les clients sont responsables de leur propre évaluation indépendante des informations contenues dans ce document. Le présent document : (a) est fourni à titre informatif uniquement, (b) représente les offres et pratiques actuelles de produits AWS, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ou assurance de la part d'AWS et de ses affiliés, fournisseurs ou concédants de licences. Les produits ou services AWS sont fournis « en l'état » sans garantie, représentation ou condition, de quelque nature que ce soit, explicite ou implicite. Les responsabilités et obligations d'AWS envers ses clients sont déterminées par les contrats AWS, et le présent document ne fait pas partie d'un contrat entre AWS et ses clients, ni le modifie.