Whitepaper zu AWS

Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud

Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud: Whitepaper zu AWS

Table of Contents

Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud

Erscheinungsdatum: 12. Februar 2021 (Dokumentversionen)

Überblick

Die Notfallwiederherstellung ist der Prozess der Vorbereitung auf eine Katastrophe und die entsprechende Wiederherstellung. Ein Ereignis, das einen Workload oder ein System daran hindert, seine Geschäftsziele an seinem primär bereitgestellten Standort zu erfüllen, wird als Katastrophe bezeichnet. Dieses Dokument beschreibt die bewährten Methoden zur Planung und zum Testen der Notfallwiederherstellung für jeden in AWS bereitgestellten Workload und bietet verschiedene Ansätze zur Risikominderung und zur Gewährleistung der Recovery Time Objective (RTO)- und Recovery Point Objective (RPO)-Ziele für den Workload.

Einführung

Ihr Workload muss seine vorgesehenen Funktionen korrekt und konsistent ausführen. Um dies zu erreichen, müssen Sie eine Resilienz-Architektur entwickeln. Resilienz ist die Fähigkeit eines Workloads, sich von Infrastruktur- oder Service-Störungen zu erholen, dynamisch Datenverarbeitungsressourcen zu nutzen, um die Nachfrage zu bewältigen, und Störungen wie Fehlkonfigurationen oder vorübergehende Netzwerkprobleme abzumildern.

Die Notfallwiederherstellung (Disaster Recovery, DR) ist ein wichtiger Bestandteil Ihrer Resilienz-Strategie und betrifft die Reaktion Ihres Workloads auf eine Katastrophe (eine Katastrophe ist ein Ereignis, das schwerwiegende negative Auswirkungen auf Ihr Geschäft hat). Diese Reaktion muss auf den Geschäftszielen Ihres Unternehmens basieren, die die Strategie Ihres Workloads zur Vermeidung von Datenverlusten (Recovery Point Objective (RPO)) und zur Reduzierung der Ausfallzeiten, in denen Ihr Workload nicht zur Verfügung steht (Recovery Time Objective (RTO)) definieren. Sie müssen daher bei der Entwicklung Ihrer Workloads in der Cloud eine Ausfallsicherheit implementieren, um Ihre Wiederherstellungsziele (RPO und RTO) für ein bestimmtes einmaliges Katastrophenereignis zu erreichen. Dieser Ansatz hilft Ihrem Unternehmen bei der Aufrechterhaltung der Geschäftskontinuität im Rahmen des Betriebskontinuitätsplans (BCP).

In diesem Dokument geht es darum, wie Sie Architekturen auf AWS planen, entwerfen und implementieren, die die Ziele der Notfallwiederherstellung für Ihr Unternehmen erfüllen. Die hier vermittelten Informationen richten sich an Personen in technischen Funktionen, wie Chief Technology Officers (CTOs), Architekten, Entwickler und Mitglieder des Betriebsteams.

Notfallwiederherstellung und Verfügbarkeit

Die Notfallwiederherstellung kann mit der Verfügbarkeit verglichen werden, die eine weitere wichtige Komponente Ihrer Resilienz-Strategie ist. Während bei der Notfallwiederherstellung Ziele für einmalige Ereignisse erfasst werden, messen Verfügbarkeitsziele Durchschnittswerte über einen bestimmten Zeitraum.

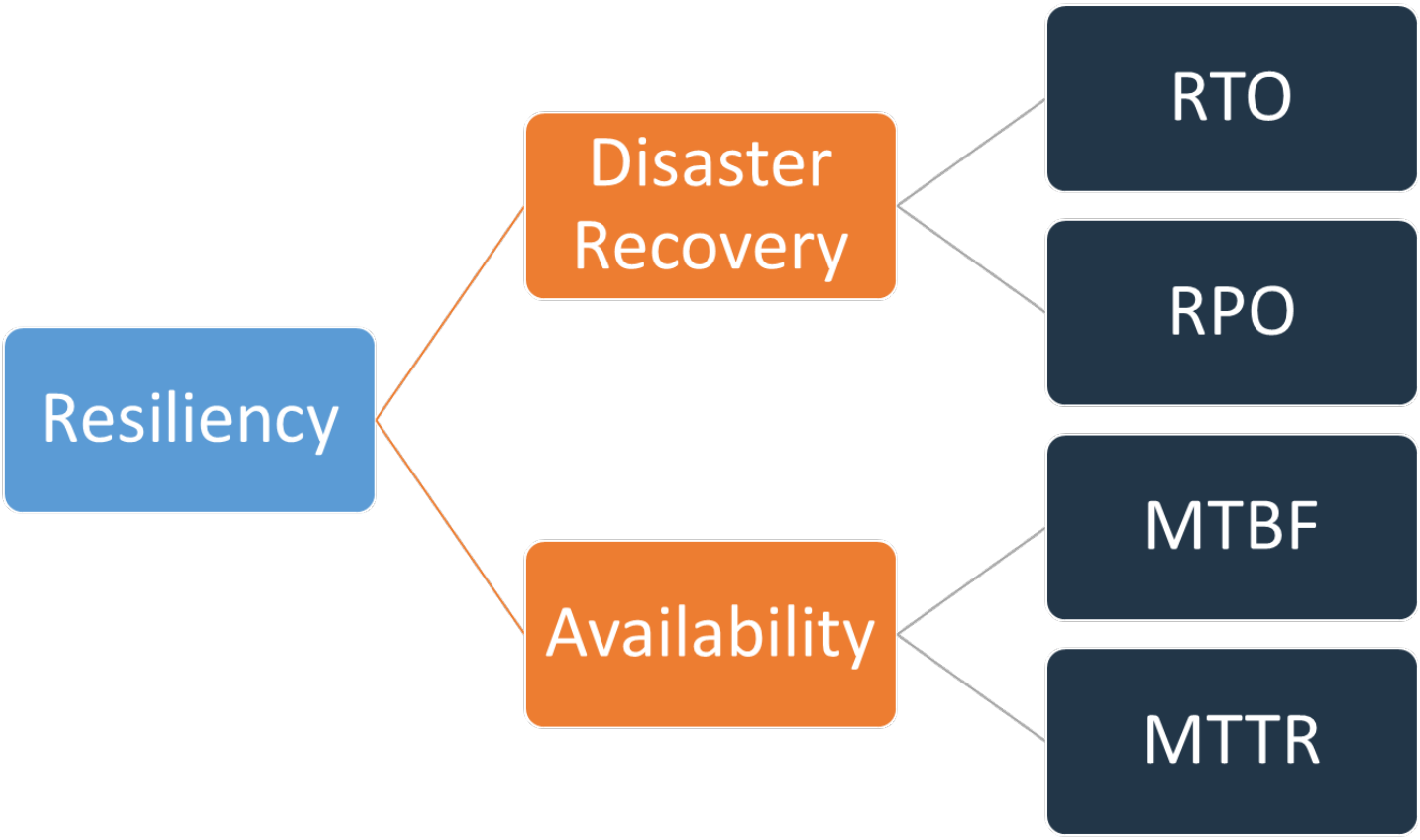

Abbildung 1 - Resilienz-Ziel

Die Verfügbarkeit wird anhand der Werte Mean Time Between Failures (MTBF) und Mean Time to Recover (MTTR) berechnet:

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

Dieser Ansatz wird oft als "nines" bezeichnet, wobei ein Verfügbarkeitsziel von 99,9 % als "three nines" bezeichnet wird.

Unter Umständen ist es für einen Workload einfacher, erfolgreiche und fehlgeschlagene Anfragen zu zählen, anstatt einen zeitbasierten Ansatz zu verwenden. In diesem Fall kann die folgende Berechnung verwendet werden:

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

Die Notfallwiederherstellung konzentriert sich auf Katastrophenereignisse, während sich die Verfügbarkeit auf häufigere Unterbrechungen kleineren Ausmaßes wie Komponentenausfälle, Netzwerkprobleme und Lastspitzen konzentriert. Das Ziel der Notfallwiederherstellung ist die Aufrechterhaltung des Geschäftsbetriebs, während es bei der Verfügbarkeit darum geht, die Zeit zu maximieren, in der ein Workload für die Ausführung der vorgesehenen Geschäftsfunktionen zur Verfügung steht. Beide sollten Teil Ihrer Resilienz-Strategie sein.

Modell der geteilten Verantwortung für Resilienz

Für die Ausfallsicherheit sind AWS und Sie, der Kunde, gemeinsam verantwortlich. Es ist wichtig, dass Sie die Funktionsweise der Notfallwiederherstellung und der Verfügbarkeit als Teil der Ausfallsicherheit im Rahmen dieses gemeinsamen Modells kennen.

AWS-Verantwortung "Ausfallsicherheit der Cloud"

AWS ist für die Ausfallsicherheit der Infrastruktur verantwortlich, auf der alle in der AWS Cloud angebotenen Services ausgeführt werden. Diese Infrastruktur umfasst die Hardware, Software, Netzwerke und Einrichtungen, die die AWS Cloud-Services ausführen. AWS unternimmt wirtschaftlich vertretbare Anstrengungen, um diese AWS Cloud-Services verfügbar zu halten und sicherzustellen, dass die Verfügbarkeit der Services die AWS-Servicelevelvereinbarungen (SLAs) erfüllt oder übertrifft.

Die globale Cloud-Infrastruktur von AWS ist so konzipiert, dass Kunden hochgradig widerstandsfähige Workload-Architekturen aufbauen können. Jede AWS-Region ist vollständig isoliert und besteht aus mehreren Availability Zones, bei denen es sich um physisch isolierte Partitionen der Infrastruktur handelt. Availability Zones isolieren Fehler, die die Ausfallsicherheit des Workloads beeinträchtigen könnten, und verhindern, dass sie sich auf andere Zonen in der Region auswirken. Gleichzeitig sind alle Zonen in einer AWS-Region mit einem Netzwerk mit hoher Bandbreite und niedriger Latenz verbunden, und zwar über vollständig redundante, dedizierte Metro-Glasfaserverbindungen, die einen hohen Durchsatz und eine niedrige Latenz zwischen den Zonen ermöglichen. Der gesamte Datenverkehr zwischen den Zonen ist verschlüsselt. Die Netzwerkleistung ist ausreichend, um eine synchrone Replikation zwischen den Zonen zu ermöglichen. Availability Zones vereinfachen den Prozess der Partitionierung von Anwendungen für hohe Verfügbarkeit.

Kundenverantwortung "Ausfallsicherheit in der Cloud"

Ihre Verantwortung wird von den AWS Cloud-Services bestimmt, die Sie auswählen. Dies bestimmt den Umfang der Konfigurationsarbeit, die Sie im Rahmen Ihrer Verantwortung für die Ausfallsicherheit durchführen müssen. Bei einem Service wie Amazon Elastic Compute Cloud (Amazon EC2) muss der Kunde zum Beispiel alle notwendigen Aufgaben zur Konfiguration und Verwaltung der Ausfallsicherheit selbst übernehmen. Kunden, die Amazon EC2-Instances bereitstellen, sind für die Bereitstellung von EC2-Instances über mehrere Standorte (wie AWS Availability Zones), die Implementierung der selbständigen Reparatur mit Services wie AWS Auto

Scaling sowie die Verwendung von bewährten Methoden für eine robuste Workload-Architektur für Anwendungen, die auf den Instances installiert sind, verantwortlich. Für verwaltete Services wie Amazon S3 und Amazon DynamoDB betreibt AWS die Infrastrukturebene, das Betriebssystem und die Plattformen und die Kunden greifen auf die Endpunkte zu, um Daten zu speichern und abzurufen. Sie sind für die Verwaltung der Ausfallsicherheit Ihrer Daten verantwortlich, einschließlich Backup, Versionierung und Replikationsstrategien.

Das Bereitstellen Ihres Workloads über mehrere Availability Zones in einer AWS-Region ist Teil einer Hochverfügbarkeitsstrategie, die darauf ausgelegt ist, Workloads zu schützen, indem Probleme auf eine Availability Zone beschränkt werden und die Redundanz der anderen Availability Zones genutzt wird, um Anfragen weiterhin zu bedienen. Eine Multi-AZ-Architektur ist auch Teil einer DR-Strategie, die darauf abzielt, Workloads besser zu isolieren und vor Problemen wie Stromausfällen, Blitzeinschlägen, Tornados, Erdbeben und mehr zu schützen. DR-Strategien können außerdem mehrere AWS-Regionen nutzen. In einer Active/Passive-Konfiguration wird zum Beispiel ein Failover für den Service des Workloads von der aktiven Region auf die DR-Region durchgeführt, wenn die aktive Region keine Anfragen mehr bedienen kann.

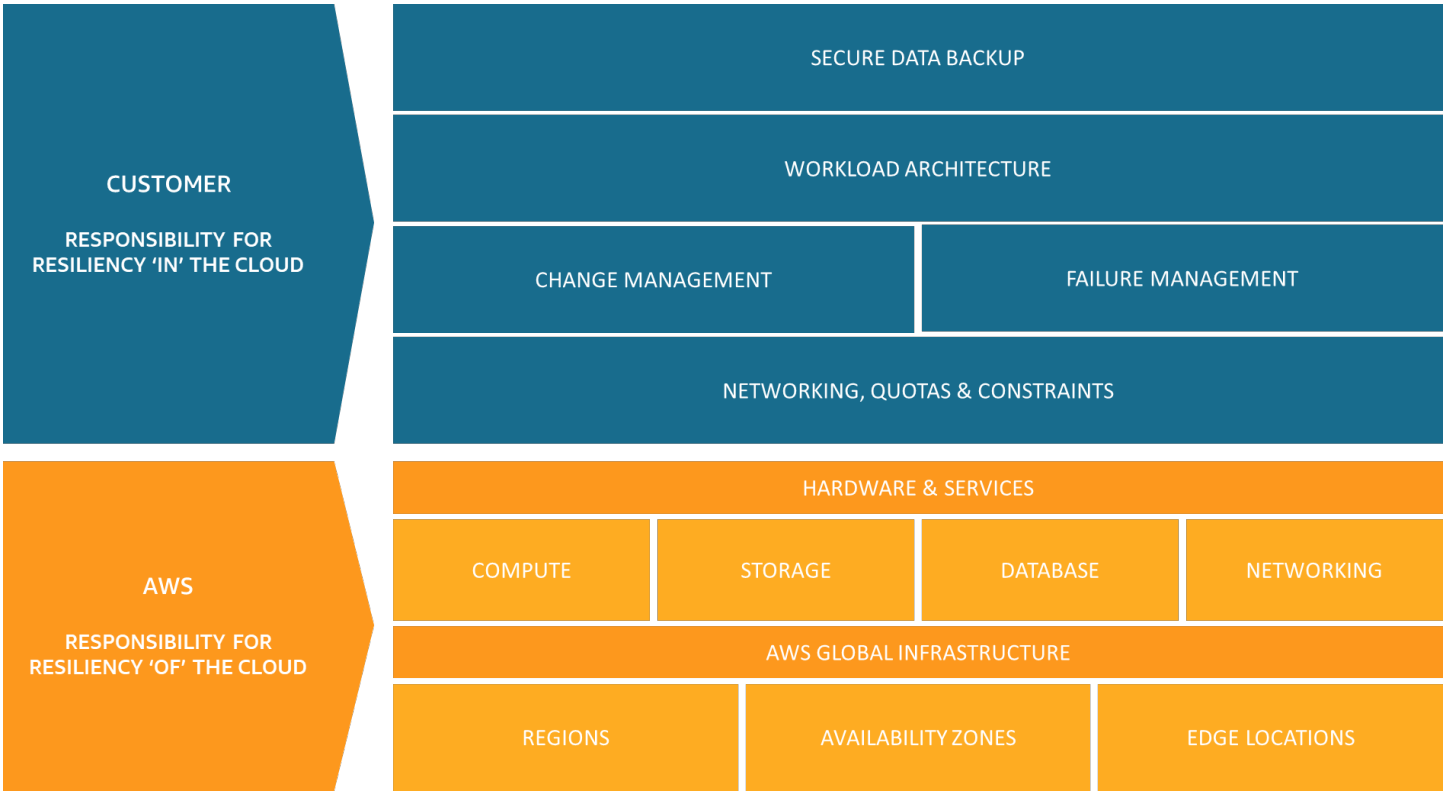

Abbildung 2 - Die Ausfallsicherheit ist eine gemeinsame Aufgabe von AWS und dem Kunden

Was ist eine Katastrophe?

Wenn Sie eine Notfallwiederherstellung planen, sollten Sie Ihren Plan auf diese drei Hauptkategorien von Katastrophen hin bewerten:

- Naturkatastrophen, wie Erdbeben oder Überschwemmungen
- Technische Ausfälle, wie z. B. Stromausfall oder Ausfall der Netzwerkkonnektivität
- Menschliche Handlungen, wie versehentliche Fehlkonfigurationen oder unbefugte/fremde Zugriffe oder Änderungen

Jede dieser potenziellen Katastrophen hat geografische Auswirkungen, die lokal, regional, landesweit, kontinental oder global sein können. Sowohl die Art der Katastrophe als auch die geografischen Auswirkungen sind wichtig, wenn Sie Ihre Strategie zur Notfallwiederherstellung bewerten. So können Sie beispielsweise ein lokales Überschwemmungsproblem, das zu einem Ausfall des Rechenzentrums führt, mit einer Multi-AZ-Strategie abmildern, da nicht mehr als eine Availability Zone betroffen wäre. Ein Angriff auf die Produktionsdaten würde jedoch erfordern, dass Sie eine Notfallwiederherstellungsstrategie nutzen, die auf Sicherungsdaten in einer anderen AWS-Region zurückgreift.

Hochverfügbarkeit ist keine Notfallwiederherstellung

Sowohl die Verfügbarkeit als auch die Notfallwiederherstellung stützen sich auf einige derselben bewährten Methoden, wie z. B. die Überwachung auf Ausfälle, das Bereitstellen an mehreren Standorten und das automatische Failover. Allerdings konzentriert sich die Verfügbarkeit auf Komponenten des Workloads, während sich die Notfallwiederherstellung auf diskrete Kopien des gesamten Workloads konzentriert. Die Notfallwiederherstellung verfolgt andere Ziele als die Verfügbarkeit, indem sie die Zeit bis zur Wiederherstellung nach größeren Ereignissen, die als Katastrophen gelten, bestimmt. Sie sollten zunächst sicherstellen, dass Ihr Workload Ihre Verfügbarkeitsziele erfüllt. Mit einer hochverfügbaren Architektur sind Sie in der Lage, die Anforderungen Ihrer Kunden zu erfüllen, wenn die Verfügbarkeit durch ein Ereignis beeinträchtigt wird. Ihre Strategie für die Notfallwiederherstellung erfordert andere Ansätze als die für die Verfügbarkeit. Sie konzentriert sich darauf, diskrete Systeme an mehreren Standorten bereitzustellen, sodass Sie bei Bedarf einen Failover für den gesamten Workload durchführen können.

Sie müssen die Verfügbarkeit Ihres Workloads bei der Notfallwiederherstellungsplanung berücksichtigen, da sie den von Ihnen gewählten Ansatz beeinflusst. Ein Workload, der auf einer einzigen Amazon EC2-Instance in einer Availability Zone läuft, ist nicht hochverfügbar. Wenn diese Availability Zone von einem lokalen Problem betroffen ist, erfordert dieses Szenario einen Failover zu einer anderen AZ, um die DR-Ziele zu erreichen. Vergleichen Sie dieses Szenario mit einem hochverfügbaren Workload, der als Multi-Site Active/Active bereitgestellt wird, wobei der Workload über mehrere aktive Regionen bereitgestellt wird und alle Regionen den Datenverkehr der Produktionsumgebung bedienen. In diesem Fall wird die DR-Strategie selbst in dem unwahrscheinlichen Fall, dass eine massive Katastrophe eine ganze Region betrifft, durch die Weiterleitung des gesamten Datenverkehrs an die verbleibenden Regionen erreicht.

Auch die Art und Weise, wie Sie mit Daten umgehen, unterscheidet sich zwischen Verfügbarkeit und Notfallwiederherstellung. Stellen Sie sich eine Speicherlösung vor, die kontinuierlich an einen anderen Standort repliziert, um eine hohe Verfügbarkeit zu erreichen (z. B. ein Multi-Site Active/Active-Workload). Wenn eine oder mehrere Dateien auf dem primären Speichergerät gelöscht oder beschädigt werden, können diese Änderungen auf das sekundäre Speichergerät repliziert werden. In diesem Szenario ist trotz hoher Verfügbarkeit die Fähigkeit zum Failover im Falle einer Datenlöschung oder -beschädigung beeinträchtigt. Stattdessen ist im Rahmen einer DR-Strategie auch ein Point-in-Time-Backup erforderlich.

Betriebskontinuitätsplan (Business Continuity Plan, BCP)

Ihr Notfallwiederherstellungsplan sollte eine Teilmenge des Betriebskontinuitätsplans (BCP) Ihres Unternehmens darstellen. Er sollte kein eigenständiges Dokument sein. Es hat keinen Sinn, aggressive Notfallwiederherstellungsziele für die Wiederherstellung eines Workloads zu verfolgen, wenn die Geschäftsziele dieses Workloads nicht erreicht werden können, weil sich die Katastrophe auf andere Elemente Ihres Unternehmens als Ihren Workload auswirkt. Ein Erdbeben könnte Sie beispielsweise daran hindern, Produkte zu transportieren, die Sie über Ihre eCommerce-Anwendung gekauft haben - selbst wenn Ihr Workload durch eine effektive Notfallwiederherstellung funktionsfähig bleibt, muss Ihr BCP die Transportanforderungen berücksichtigen. Ihre DR-Strategie sollte auf geschäftlichen Anforderungen, Prioritäten und dem Kontext basieren.

Analyse der geschäftlichen Auswirkungen und Risikobewertung

Eine Analyse der geschäftlichen Auswirkungen sollte die geschäftlichen Auswirkungen einer Unterbrechung Ihrer Workloads quantifizieren. Sie sollte ermitteln, welche Auswirkungen es auf interne und externe Kunden hat, wenn Sie Ihre Workloads nicht nutzen können, und wie sich dies auf Ihr Geschäft auswirkt. Die Analyse sollte Ihnen dabei helfen, zu bestimmen, wie schnell der Workload verfügbar gemacht werden muss und wie viel Datenverlust toleriert werden kann. Die Wahrscheinlichkeit einer Unterbrechung und die Kosten der Wiederherstellung sind Schlüsselfaktoren, die dabei helfen, den geschäftlichen Nutzen einer Notfallwiederherstellung für einen Workload zu ermitteln.

Die geschäftlichen Auswirkungen können zeitabhängig sein. Sie sollten dies bei Ihrer Notfallwiederherstellungsplanung berücksichtigen. Eine Unterbrechung Ihres Gehaltsabrechnungssystems kann beispielsweise kurz vor der Auszahlung der Löhne und Gehälter sehr große Auswirkungen auf das Unternehmen haben, während sie kurz nach der Auszahlung der Löhne und Gehälter nur noch geringe Auswirkungen haben kann.

Eine Risikobewertung der Art der Katastrophe und der geografischen Auswirkungen zusammen mit einem Überblick über die technische Implementierung Ihres Workloads bestimmt die Wahrscheinlichkeit einer Unterbrechung für jede Art von Katastrophe.

Für sehr kritische Workloads können Sie eine hohe Verfügbarkeit über mehrere Regionen hinweg mit kontinuierlichen Backups in Betracht ziehen, um die geschäftlichen Auswirkungen zu minimieren. Bei weniger kritischen Workloads kann es eine gute Strategie sein, überhaupt keine Notfallwiederherstellung einzurichten. Und für einige Katastrophenszenarien ist es sinnvoll, keine

Strategie für die Notfallwiederherstellung zu haben, da die Wahrscheinlichkeit des Auftretens gering ist. Denken Sie daran, dass Availability Zones innerhalb einer AWS-Region bereits mit einem sinnvollen Abstand zueinander und einer sorgfältigen Planung des Standorts konzipiert sind, sodass die meisten Katastrophen nur eine Zone betreffen sollten. Daher kann eine Multi-AZ-Architektur innerhalb einer AWS-Region Ihre Anforderungen an die Risikominderung bereits erfüllen.

Die Kosten der Optionen für die Notfallwiederherstellung sollten evaluiert werden, um sicherzustellen, dass die Notfallwiederherstellungsstrategie unter Berücksichtigung der geschäftlichen Auswirkungen und des Risikos das richtige Maß an geschäftlichem Nutzen bietet.

Mit all diesen Informationen können Sie die Bedrohung, das Risiko, die Auswirkungen und die Kosten der verschiedenen Katastrophenszenarien und die damit verbundenen Wiederherstellungsoptionen dokumentieren. Anhand dieser Informationen sollten Sie Ihre Wiederherstellungsziele für jeden Ihrer Workloads festlegen.

Wiederherstellungsziele (RTO und RPO)

Bei der Erstellung einer Strategie für die Notfallwiederherstellung (DR) planen Unternehmen in der Regel das Recovery Time Objective (RTO) und das Recovery Point Objective (RPO).

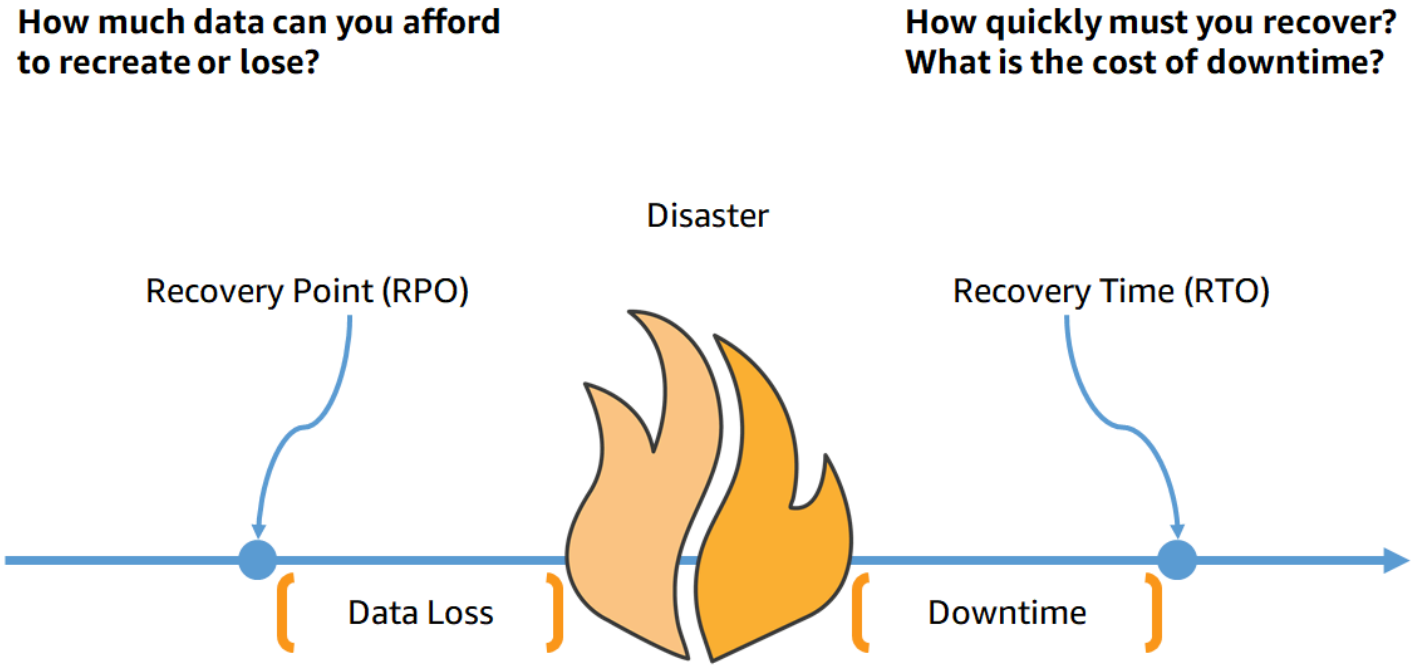

Abbildung 3 - Wiederherstellungsziele

Recovery Time Objective (RTO) ist die maximal akzeptable Verzögerung zwischen der Unterbrechung des Service und der Wiederherstellung des Service. Dieses Ziel bestimmt, welches

Zeitfenster als akzeptabel angesehen wird, wenn der Service nicht verfügbar ist, und wird von der Organisation festgelegt.

In diesem Text werden im Wesentlichen vier DR-Strategien erörtert: Backup und Wiederherstellung, Pilot Light, Warm Standby und Multi-Site Active/Active (siehe Optionen für die Notfallwiederherstellung in der Cloud). Im folgenden Diagramm hat das Unternehmen sein maximal zulässiges RTO sowie die Höchstgrenze der Ausgaben für seine Service-Wiederherstellungsstrategie festgelegt. Angesichts der Ziele des Unternehmens erfüllen die DR-Strategien Pilot Light oder Warm Standby sowohl die RTO- als auch die Kostenkriterien.

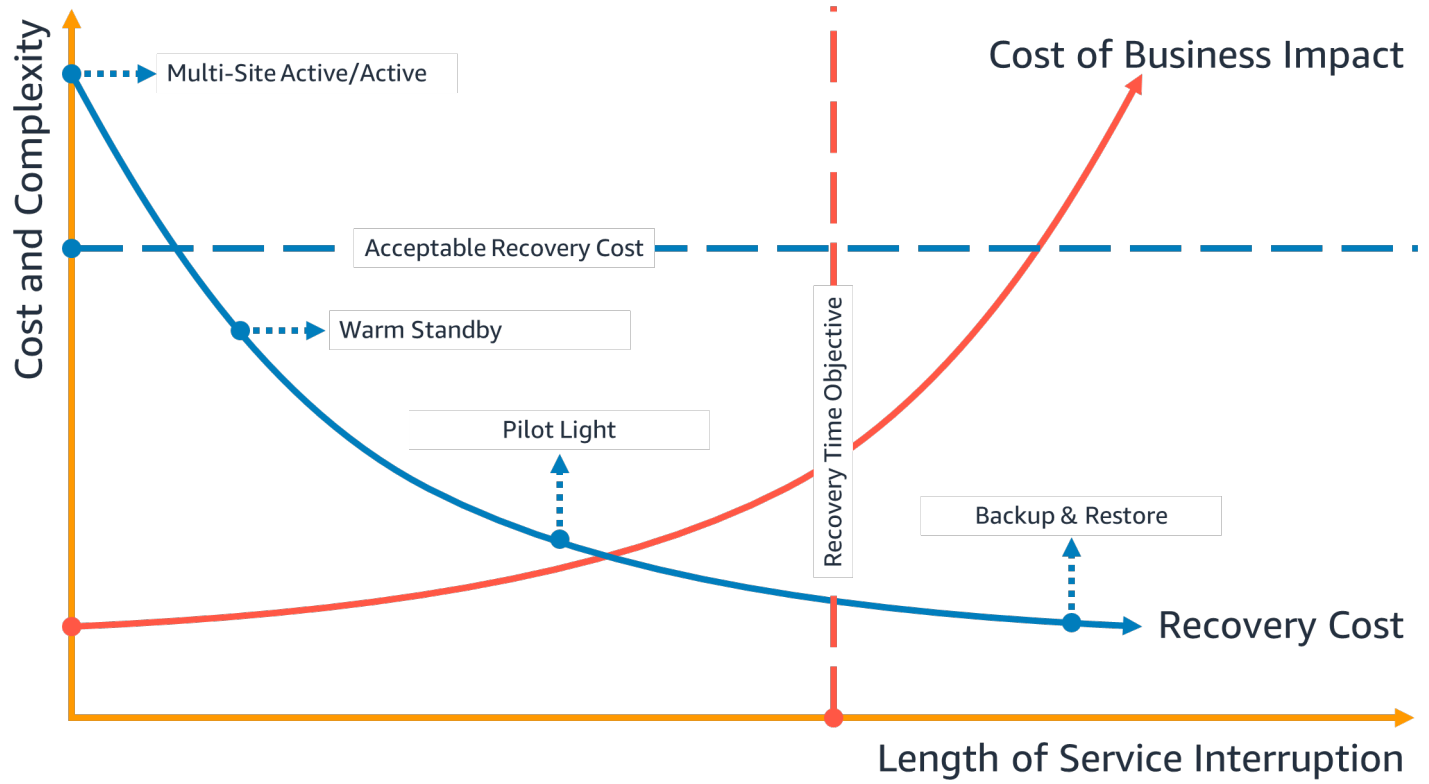

Abbildung 4 - Recovery Time Objective

Recovery Point Objective (RPO) ist die maximal akzeptable Zeitspanne seit dem letzten Datenwiederherstellungspunkt. Dieses Ziel legt fest, was als akzeptabler Datenverlust zwischen dem letzten Wiederherstellungspunkt und der Unterbrechung des Service gilt, und wird vom Unternehmen definiert.

Im folgenden Diagramm hat das Unternehmen sein maximal zulässiges RPO sowie die Höchstgrenze der Ausgaben für seine Datenwiederherstellungsstrategie festgelegt. Von den vier DR-Strategien erfüllen entweder die Pilot Light oder die Warm Standby DR-Strategie beide Kriterien für RPO und Kosten.

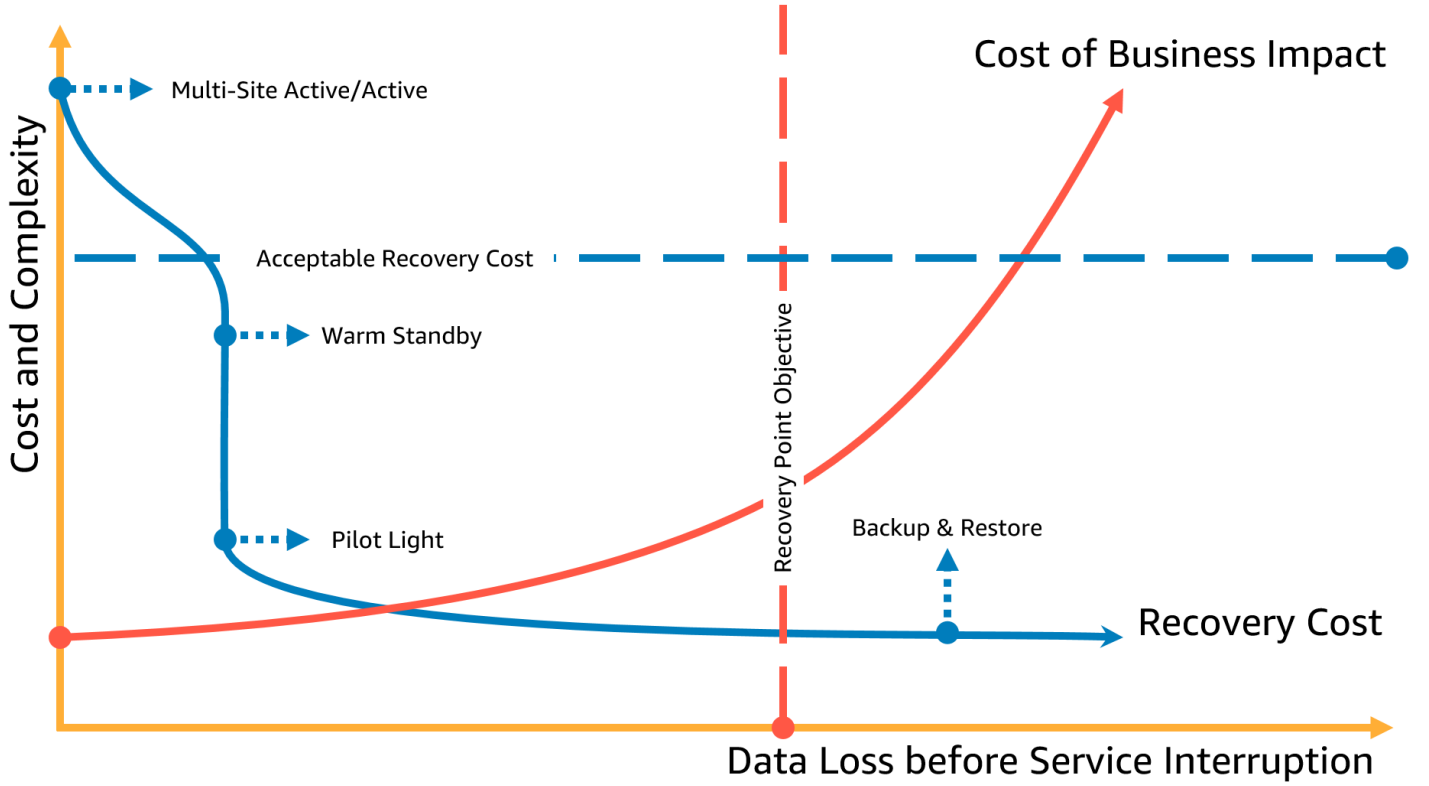

Abbildung 5 - Recovery Point Objective

> **Note**
>
> Wenn die Kosten der Wiederherstellung höher sind als die Kosten des Ausfalls oder Verlusts, sollte die Wiederherstellungsoption nicht eingeführt werden, es sei denn, es gibt einen sekundären Grund, wie z. B. gesetzliche Anforderungen.

Notfallwiederherstellung in der Cloud ist anders

Strategien zur Notfallwiederherstellung entwickeln sich mit der technischen Innovation weiter. Ein Notfallwiederherstellungsplan für lokale Umgebungen kann den physischen Transport von Bändern oder die Replikation von Daten an einen anderen Standort beinhalten. Ihr Unternehmen muss die geschäftlichen Auswirkungen, Risiken und Kosten seiner bisherigen Notfallwiederherstellungsstrategien neu bewerten, um seine Notfallwiederherstellungsziele in AWS zu erreichen. Die Notfallwiederherstellung in der AWS Cloud bietet gegenüber herkömmlichen Umgebungen die folgenden Vorteile:

- Schnelle Wiederherstellung nach einer Katastrophe bei geringerer Komplexität
- Einfache und wiederholbare Tests ermöglichen einfachere und häufigere Tests
- Geringerer Verwaltungsaufwand senkt die operative Belastung
- Automatisierungsmöglichkeiten verringern die Fehlerwahrscheinlichkeit und verkürzen die Wiederherstellungszeit

Mit AWS können Sie die fixen Kapitalkosten eines physischen Backup-Rechenzentrums gegen die variablen Betriebskosten einer Umgebung in der Cloud tauschen, was die Kosten erheblich senken kann.

Für viele Unternehmen basierte die Notfallwiederherstellung vor Ort auf dem Risiko einer Unterbrechung eines Workloads oder von Workloads in einem Rechenzentrum und der Wiederherstellung von gesicherten oder replizierten Daten in einem sekundären Rechenzentrum. Wenn Unternehmen Workloads auf AWS bereitstellen, können sie einen sorgfältig konzipierten Workload implementieren und sich auf das Design der globalen AWS Cloud-Infrastruktur verlassen, um die Auswirkungen solcher Unterbrechungen zu mildern. Im Whitepaper AWS Well-Architected Framework - Reliability Pillar finden Sie weitere Informationen zu bewährten Architekturverfahren für die Entwicklung und den Betrieb zuverlässiger, sicherer, effizienter und kostengünstiger Workloads in der Cloud.

Wenn sich Ihre Workloads auf AWS befinden, müssen Sie sich keine Gedanken über die Konnektivität des Rechenzentrums (mit Ausnahme Ihrer Zugriffsmöglichkeiten), die Stromversorgung, die Klimatisierung, die Brandbekämpfung und die Hardware machen. All dies wird für Sie verwaltet. Sie haben Zugriff auf mehrere voneinander isolierte Availability Zones (die jeweils aus einem oder mehreren diskreten Rechenzentren bestehen).

Einzelne AWS-Region

Bei einem Katastrophenereignis, das auf der Unterbrechung oder dem Verlust eines physischen Rechenzentrums beruht, trägt die Implementierung eines hochverfügbaren Workloads in mehreren Availability Zones innerhalb einer einzigen AWS Region dazu bei, natürliche und technische Katastrophen abzufedern und das Risiko menschlicher Bedrohungen wie Fehler oder unbefugte Aktivitäten, die zu Datenverlusten führen können, zu verringern. Jede AWS Region besteht aus mehreren Availability Zones, die jeweils von den Fehlern in den anderen Zonen isoliert sind. Jede Availability Zone wiederum besteht aus mehreren physischen Rechenzentren. Um die Auswirkungen von Problemen besser zu isolieren und eine hohe Verfügbarkeit zu erreichen, können Sie Workloads über mehrere Zonen in derselben Region verteilen. Availability Zones sind auf physische Redundanz ausgelegt und bieten Ausfallsicherheit, sodass selbst bei Stromausfällen, Internetausfällen, Überschwemmungen und anderen Naturkatastrophen eine unterbrechungsfreie Leistung gewährleistet ist. Unter Globale AWS Cloud-Infrastruktur erfahren Sie, wie AWS dies erreicht.

Indem Sie mehrere Availability Zones in einer einzigen AWS-Region bereitstellen, ist Ihr Workload besser gegen den Ausfall eines einzelnen Rechenzentrums (oder sogar mehrerer) Rechenzentren geschützt. Für zusätzliche Sicherheit bei der Bereitstellung in einer einzigen Region können Sie Daten und Konfiguration (einschließlich der Infrastrukturdefinition) in einer anderen Region sichern. Mit dieser Strategie reduziert sich der Umfang Ihres Notfallwiederherstellungsplans auf die Sicherung und Wiederherstellung von Daten. Die Nutzung der Ausfallsicherheit mehrerer Regionen durch eine Sicherung in einer anderen AWS-Region ist im Vergleich zu den anderen im folgenden Abschnitt beschriebenen Optionen für mehrere Regionen einfach und kostengünstig. Wenn Sie beispielsweise auf Amazon Simple Storage Service (Amazon S3) sichern, können Sie Ihre Daten sofort abrufen. Wenn Ihre DR-Strategie für Teile Ihrer Daten jedoch geringere Anforderungen an die Abrufzeiten vorsieht (Minuten oder Stunden), können Sie mit Amazon S3 Glacier oder Amazon S3 Glacier Deep Archive die Kosten für Ihre Backup- und Wiederherstellungsstrategie erheblich senken.

Für einige Workloads gelten möglicherweise gesetzliche Anforderungen an die Datenaufbewahrung. Wenn dies auf Ihre Workloads an einem Standort zutrifft, der derzeit nur eine AWS Region hat, können Sie zusätzlich zu der oben beschriebenen Entwicklung von Multi-AZ-Workloads für Hochverfügbarkeit auch die AZs innerhalb dieser Region als diskrete Standorte verwenden, was hilfreich sein kann, um die Anforderungen an die Datenspeicherung für Ihre Workloads innerhalb dieser Region zu erfüllen. Die in den folgenden Abschnitten beschriebenen DR-Strategien verwenden mehrere AWS-Regionen, können aber auch mit Availability Zones anstelle von Regionen implementiert werden.

Mehrere AWS-Regionen

Für einen Katastrophenfall, bei dem das Risiko besteht, dass mehrere Rechenzentren in beträchtlicher Entfernung voneinander ausfallen, sollten Sie Notfallwiederherstellungsoptionen in Betracht ziehen, um sich gegen natürliche und technische Katastrophen abzusichern, die eine ganze Region innerhalb von AWS betreffen. Alle in den folgenden Abschnitten beschriebenen Optionen können als Multi-Region-Architekturen zum Schutz vor solchen Katastrophen implementiert werden.

Optionen zur Notfallwiederherstellung in der Cloud

Die Strategien zur Notfallwiederherstellung, die Ihnen innerhalb von AWS zur Verfügung stehen, lassen sich grob in vier Ansätze einteilen, die von der kostengünstigen und wenig komplexen Erstellung von Backups bis hin zu komplexeren Strategien mit mehreren aktiven Regionen reichen. Es ist von entscheidender Bedeutung, dass Sie Ihre Strategie zur Notfallwiederherstellung regelmäßig testen, sodass Sie sie im Bedarfsfall sicher auslösen können.

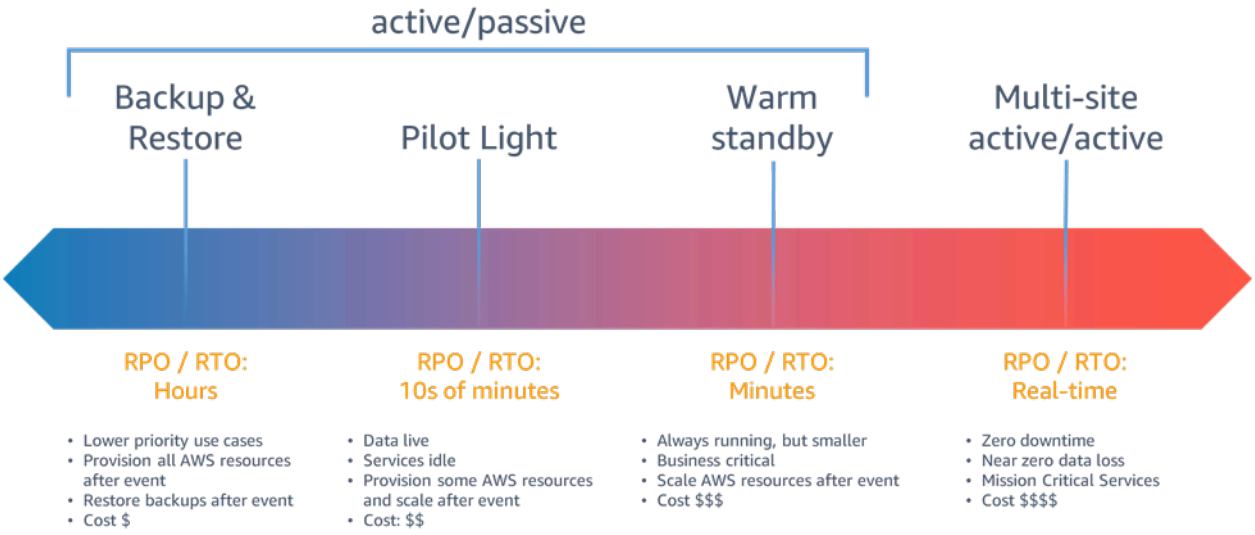

Abbildung 6 - Strategien zur Notfallwiederherstellung

Bei einem Katastrophenereignis, das durch die Unterbrechung oder den Verlust eines physischen Rechenzentrums für einen gut strukturierten, hochverfügbaren Workload verursacht wird, benötigen Sie möglicherweise nur einen Backup- und Wiederherstellungsansatz für die Notfallwiederherstellung. Wenn Ihre Definition einer Katastrophe über die Unterbrechung oder den Verlust eines physischen Rechenzentrums hinausgeht und sich auf eine Region erstreckt oder wenn Sie gesetzlichen Anforderungen unterliegen, die dies erfordern, dann sollten Sie Pilot Light, Warm Standby oder Multi-Site Active/Active in Betracht ziehen.

Backup und Wiederherstellung

Backup und Wiederherstellung ist ein geeigneter Ansatz, um Datenverlusten oder -beschädigungen vorzubeugen. Dieser Ansatz kann auch verwendet werden, um eine regionale Katastrophe abzumildern, indem Daten in andere AWS-Regionen repliziert werden, oder um eine fehlende Redundanz für Workloads abzumildern, die in einer einzigen Availability Zone bereitgestellt werden. Zusätzlich zu den Daten müssen Sie auch die Infrastruktur, die Konfiguration und den

Anwendungscode in der Wiederherstellungsregion neu bereitstellen. Damit die Infrastruktur schnell und fehlerfrei neu bereitgestellt werden kann, sollten Sie die Bereitstellung als Infrastruktur as Code (IaC) mit Services wie AWS CloudFormation oder AWS Cloud Development Kit (AWS CDK) durchführen. Ohne IaC kann die Wiederherstellung von Workloads in der Wiederherstellungsregion komplex sein, was zu längeren Wiederherstellungszeiten führt und möglicherweise Ihr RTO überschreitet. Sichern Sie neben den Benutzerdaten auch den Code und die Konfiguration, einschließlich Amazon Machine Images (AMIs), die Sie zur Erstellung von Amazon EC2-Instances verwenden. Sie können AWS CodePipeline verwenden, um die Neuverteilung von Anwendungscode und Konfiguration zu automatisieren.

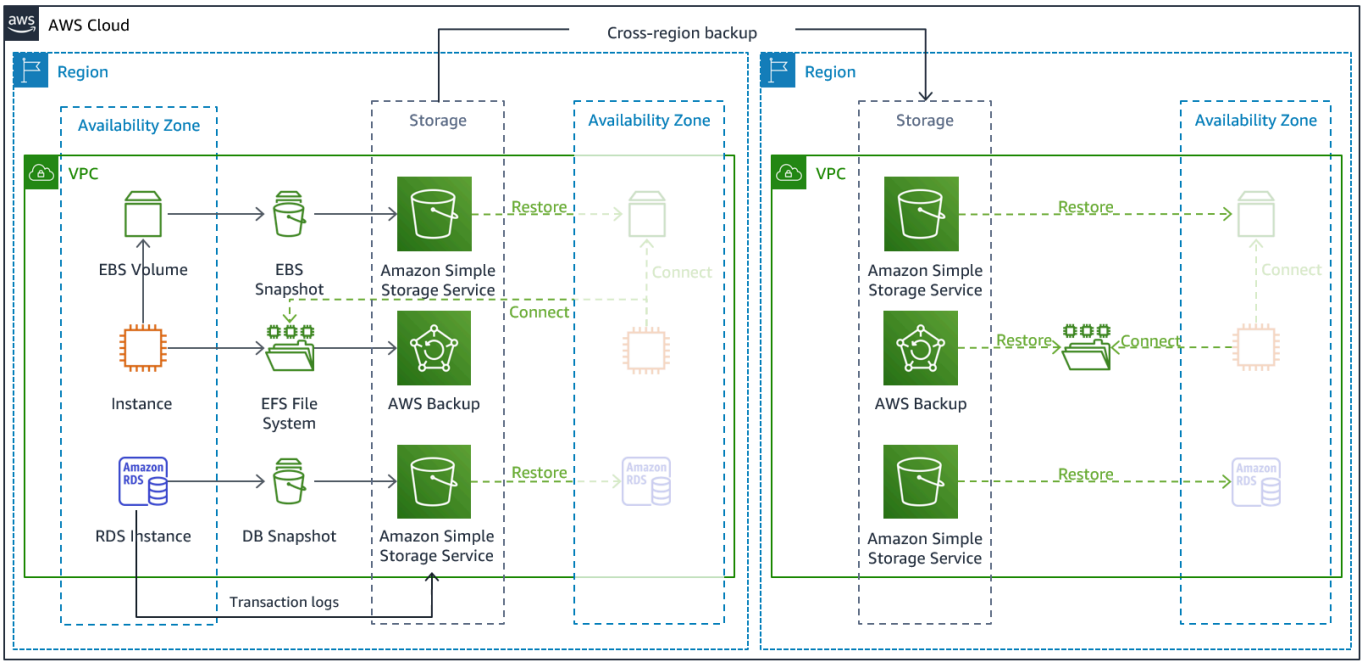

Abbildung 7 - Sicherungs- und Wiederherstellungsarchitektur

AWS-Services

Ihre Workload-Daten erfordern eine Backup-Strategie, die in regelmäßigen Abständen oder kontinuierlich durchgeführt wird. Wie oft Sie Ihr Backup ausführen, bestimmt den erreichbaren Wiederherstellungspunkt (der sich an Ihrem RPO orientieren sollte). Das Backup sollte außerdem die Möglichkeit bieten, es zu dem Zeitpunkt wiederherzustellen, zu dem es erstellt wurde. Backups mit Point-in-Time-Recovery sind über die folgenden Services und Ressourcen verfügbar:

- Amazon Elastic Block Store (Amazon EBS)-Snapshot
- Amazon DynamoDB-Backup

- Amazon RDS-Snapshot
- Amazon Aurora DB-Snapshot
- Amazon EFS-Backup (bei Verwendung von AWS Backup)
- Amazon Redshift-Snapshot
- Amazon Neptune-Snapshot

Für Amazon Simple Storage Service (Amazon S3) können Sie Amazon S3 Cross-Region-Replication (CRR) verwenden, um Objekte asynchron und kontinuierlich in einen S3-Bucket in der DR-Region zu kopieren und gleichzeitig eine Versionierung für die gespeicherten Objekte bereitzustellen, sodass Sie den Wiederherstellungspunkt wählen können. Die kontinuierliche Replikation von Daten hat den Vorteil, dass sie die kürzeste Zeit (nahezu null) für die Sicherung Ihrer Daten benötigt, schützt aber möglicherweise nicht so gut vor Katastrophenereignissen wie Datenbeschädigung oder Angriffen (z. B. unbefugtes Löschen von Daten) wie die zeitpunktbezogene Sicherungen. Die kontinuierliche Replikation wird im Abschnitt AWS-Services für Pilot Light behandelt.

AWS Backup bietet einen zentralen Ort zur Konfiguration, Planung und Überwachung der AWS-Backup-Funktionen für die folgenden Services und Ressourcen:

- Amazon Elastic Block Store (Amazon EBS)-Volumes
- Amazon EC2-Instances
- Amazon Relational Database Service (Amazon RDS)-Datenbanken (einschließlich Amazon Aurora-Datenbanken)
- Amazon DynamoDB-Tabellen
- Amazon Elastic File System (Amazon EFS)-Dateisysteme
- AWS Storage Gateway-Volumes
- Amazon FSx for Windows File Server und Amazon FSx for Lustre

AWS Backup unterstützt das Kopieren von Backups über Regionen hinweg, z. B. in eine Notfallwiederherstellungsregion.

Als zusätzliche Notfallwiederherstellungsstrategie für Ihre Amazon S3-Daten aktivieren Sie die S3-Objektversionierung. Die Objektversionierung schützt Ihre Daten in S3 vor den Folgen von Lösch- oder Änderungsaktionen, indem sie die ursprüngliche Version vor der Aktion beibehält. Die Objektversionierung kann eine nützliche Abhilfe für Katastrophen durch menschliches Versagen sein. Wenn Sie die S3-Replikation verwenden, um Daten in Ihrer DR-Region zu sichern, dann fügt Amazon

S3 standardmäßig eine Löschmarkierung nur im Quell-Bucket hinzu, wenn ein Objekt im Quell-Bucket gelöscht wird. Dieser Ansatz schützt die Daten in der DR-Region vor böswilligen Löschungen in der Quellregion.

Zusätzlich zu den Daten müssen Sie auch die Konfiguration und die Infrastruktur sichern, die erforderlich sind, um Ihren Workload neu zu verteilen und Ihr Recovery Time Objective (RTO) einzuhalten. AWS CloudFormation bietet IaC-Lösungen (Infrastructure as Code, mit denen Sie alle AWS-Ressourcen in Ihrem Workload definieren können, sodass Sie diesen zuverlässig für mehrere AWS-Konten und AWS-Regionen bereitstellen und neu bereitstellen können). Sie können die von Ihrem Workload verwendeten Amazon EC2-Instances als Amazon Machine Images (AMIs) sichern. Das AMI wird aus Snapshots des Root-Volumes Ihrer Instance und aller anderen EBS-Volumes erstellt, die mit Ihrer Instance verknüpft sind. Sie können dieses AMI verwenden, um eine wiederhergestellte Version der EC2-Instance zu starten. Ein AMI kann innerhalb einer Region oder regionsübergreifend kopiert werden. Oder Sie können AWS Backup verwenden, um Backups kontoübergreifend und in andere AWS-Regionen zu kopieren. Die kontoübergreifende Backup-Funktion schützt Sie vor Katastrophenereignissen wie internen Bedrohungen oder einer Kontokompromittierung. AWS Backup bietet außerdem zusätzliche Funktionen für EC2-Backups. Zusätzlich zu den einzelnen EBS-Volumes der Instances speichert und verfolgt AWS Backup auch die folgenden Metadaten: Instance-Typ, konfigurierte Virtual Private Cloud (VPC), Sicherheitsgruppe, IAM-Rolle, Überwachungskonfiguration und Tags. Diese zusätzlichen Metadaten werden jedoch nur bei der Wiederherstellung des EC2-Backups in derselben AWS Region verwendet.

Alle Daten, die in der Notfallwiederherstellungsregion als Backups gespeichert sind, müssen zum Zeitpunkt des Failovers wiederhergestellt werden. AWS Backup bietet eine Wiederherstellungsfunktion, ermöglicht aber derzeit keine geplante oder automatische Wiederherstellung. Sie können eine automatische Wiederherstellung in der DR-Region implementieren, indem Sie das AWS SDK verwenden, um APIs für AWS Backup aufzurufen. Sie können dies als regelmäßig wiederkehrenden Auftrag einrichten oder die Wiederherstellung auslösen, sobald ein Backup abgeschlossen ist. Die folgende Abbildung zeigt ein Beispiel für die automatische Wiederherstellung mit Amazon Simple Notification Service (Amazon SNS) und AWS Lambda.

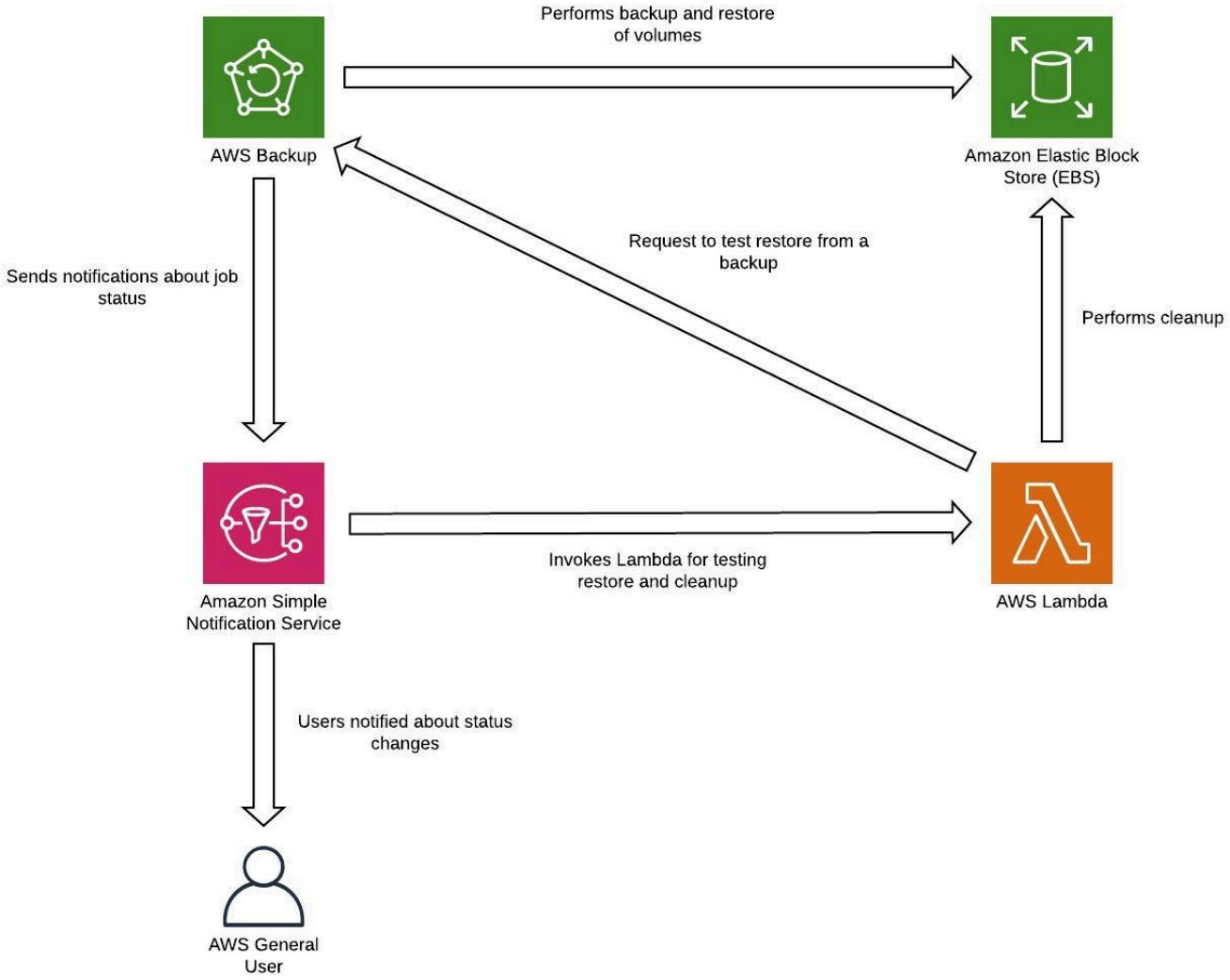

Abbildung 8 - Wiederherstellen und Testen von Backups

> **Note**
>
> Ihre Backup-Strategie muss das Testen Ihrer Backups einbeziehen. Weitere Informationen finden Sie im Abschnitt Testen der Notfallwiederherstellung. Lesen Sie außerdem den Artikel AWS Well-Architected Lab: Testen von Backups und Wiederherstellung von Daten. Dort finden Sie eine praktische Demonstration der Implementierung.

Pilot Light

Mit dem Pilot Light-Ansatz replizieren Sie Ihre Daten von einer Region in eine andere und stellen eine Kopie Ihrer zentralen Workload-Infrastruktur bereit. Ressourcen, die zur Unterstützung der Datenreplikation und des Backups erforderlich sind, wie Datenbanken und Objektspeicher, sind immer betriebsbereit. Andere Elemente, wie z. B. Anwendungsserver, werden mit Anwendungscode und Konfigurationen geladen, sind aber offline und werden nur während der Tests oder beim Aufrufen des Failovers für die Notfallwiederherstellung verwendet. Im Gegensatz zum Backup- und Wiederherstellungsansatz ist Ihre Kerninfrastruktur immer verfügbar und Sie haben jederzeit die Möglichkeit, durch die Aktivierung und Skalierung Ihrer Anwendungsserver schnell eine vollwertige Produktionsumgebung bereitzustellen.

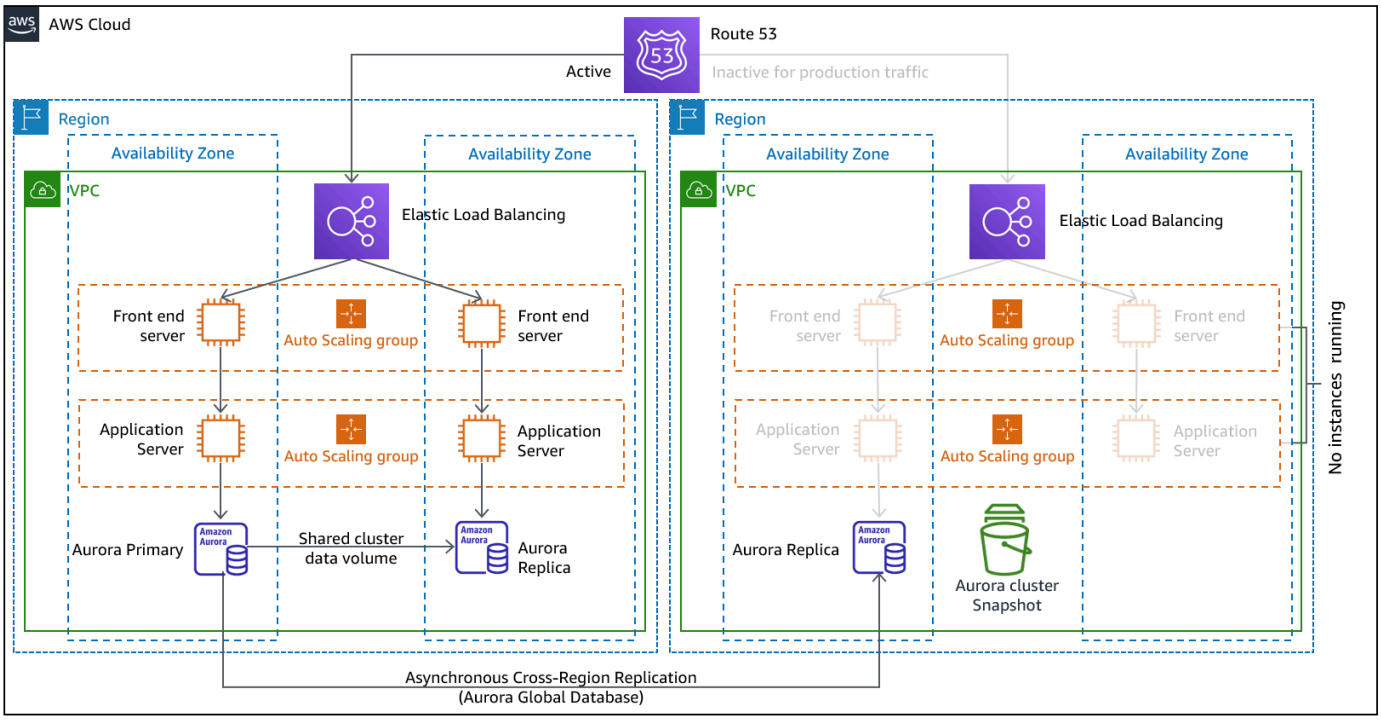

Abbildung 9 - Pilot Light-Architektur

Ein Pilot Light-Ansatz minimiert die laufenden Kosten für die Notfallwiederherstellung, indem er die aktiven Ressourcen minimiert und die Wiederherstellung zum Zeitpunkt einer Katastrophe vereinfacht, da die Kernanforderungen an die Infrastruktur bereits vorhanden sind. Diese Wiederherstellungsoption erfordert, dass Sie Ihren Bereitstellungsansatz ändern. Sie müssen Änderungen an der Kerninfrastruktur in jeder Region vornehmen und gleichzeitig Änderungen am Workload (Konfiguration, Code) in jeder Region bereitstellen. Dieser Schritt lässt sich vereinfachen, indem Sie Ihre Bereitstellungen automatisieren und Infrastruktur as Code (IaC) verwenden, um die

Infrastruktur über mehrere Konten und Regionen hinweg bereitzustellen (vollständige Bereitstellung der Infrastruktur in der primären Region und verkleinerte/deaktivierte Bereitstellung der Infrastruktur in den DR-Regionen). Es wird empfohlen, für jede Region ein anderes Konto zu verwenden, um ein Höchstmaß an Ressourcen- und Sicherheitsisolierung zu gewährleisten (falls kompromittierte Anmeldeinformationen Teil Ihrer Notfallwiederherstellungspläne sind).

Bei diesem Ansatz müssen Sie sich auch gegen eine Datenpanne wappnen. Die kontinuierliche Datenreplikation schützt Sie zwar vor einigen Arten von Katastrophen, aber möglicherweise nicht vor einer Datenbeschädigung oder -zerstörung, es sei denn, Ihre Strategie umfasst auch die Versionierung der gespeicherten Daten oder Optionen für eine Point-in-Time-Wiederherstellung. Sie können die replizierten Daten in der Katastrophenregion sichern, um Point-in-Time-Backups in derselben Region zu erstellen.

AWS-Services

Zusätzlich zu den AWS-Services für Point-in-Time-Backups, die im Abschnitt Backup und Wiederherstellung beschrieben sind, sollten Sie auch die folgenden Services für Ihre Pilot Light-Strategie in Betracht ziehen.

Für Pilot Light ist die kontinuierliche Datenreplikation auf Live-Datenbanken und Datenspeicher in der DR-Region der beste Ansatz für ein niedriges RPO (wenn sie zusätzlich zu den zuvor besprochenen Point-in-Time-Backups verwendet wird). AWS bietet eine kontinuierliche, regionsübergreifende, asynchrone Datenreplikation für Daten über die folgenden Services und Ressourcen:

- Amazon Simple Storage Service (Amazon S3)-Replikation
- Amazon RDS-Read-Replicas
- Globale Amazon Aurora-Datenbanken
- Globale Amazon DynamoDB-Tabellen

Mit der kontinuierlichen Replikation sind die Versionen Ihrer Daten fast sofort in Ihrer DR-Region verfügbar. Die tatsächlichen Replikationszeiten können mithilfe von Service-Funktionen wie S3 Replication Time Control (S3 RTC) für S3-Objekte und Verwaltungsfunktionen für globale Amazon Aurora-Datenbanken überwacht werden.

Bei einem Failover, bei dem Sie Ihren Lese-/Schreib-Workload von der Notfallwiederherstellungsregion aus ausführen, müssen Sie eine RDS-Lesereplik zur primären Instance heraufstufen. Bei DB-Instances, die nicht auf Aurora basieren, dauert dieser Prozess

einige Minuten und umfasst einen Neustart. Für die regionsübergreifende Replikation (Cross-Region Replication, CRR) und den Failover mit RDS bietet die Verwendung der globalen Datenbank von Amazon Aurora mehrere Vorteile. Die globale Datenbank verwendet eine dedizierte Infrastruktur, die Ihre Datenbanken vollständig für Ihre Anwendung zur Verfügung stellt und die Replikation in die sekundäre Region mit einer typischen Latenzzeit von weniger als einer Sekunde ermöglicht (innerhalb einer AWS-Region sogar weniger als 100 Millisekunden). Mit der globalen Amazon Aurora-Datenbank können Sie bei einer Leistungsverschlechterung oder einem Ausfall Ihrer primären Region eine der sekundären Regionen hochstufen, um die Lese-/Schreibaufgaben in weniger als 1 Minute zu übernehmen, selbst im Falle eines kompletten regionalen Ausfalls. Die Hochstufung kann automatisch erfolgen und es ist kein Neustart erforderlich.

In Ihrer DR-Region muss eine verkleinerte Version Ihrer Kern-Workload-Infrastruktur mit weniger oder kleineren Ressourcen bereitgestellt werden. Mit AWS CloudFormation können Sie Ihre Infrastruktur definieren und sie konsistent über AWS-Konten und AWS-Regionen hinweg bereitstellen. AWS CloudFormation verwendet vordefinierte Pseudoparameter, um das AWS-Konto und die AWS-Region zu identifizieren, in der der Service bereitgestellt wird. Daher können Sie eine Bedingungslogik in Ihre CloudFormation-Vorlagen implementieren, um nur die verkleinerte Version Ihrer Infrastruktur in der DR-Region bereitzustellen. Für die Bereitstellung von EC2-Instances liefert ein Amazon Machine Image (AMI) Informationen zur Hardwarekonfiguration und installierten Software. Sie können eine Image Builder-Pipeline implementieren, die die von Ihnen benötigten AMIs erstellt und diese sowohl in Ihre primäre als auch in Ihre Backup-Region kopiert. So stellen Sie sicher, dass diese Golden AMIs alles enthalten, was Sie brauchen, um Ihren Workload in einer neuen Region bereitzustellen oder zu skalieren, falls es zu einem Notfall kommt. Amazon EC2-Instances werden in einer verkleinerten Konfiguration bereitgestellt (weniger Instances als in Ihrer primären Region). Sie können hibernate verwenden, um EC2-Instance in einen angehaltenen Zustand zu versetzen, in dem Sie keine EC2-Kosten verursachen. Sie zahlen dann nur für den verwendeten Speicherplatz. Um EC2-Instance zu starten, können Sie Skripte mit der AWS Command Line Interface (CLI) oder dem AWS SDK erstellen. Um die Infrastruktur zur Unterstützung des Datenverkehrs in der Produktion zu skalieren, lesen Sie AWS Auto Scaling im Abschnitt Warm Standby.

Bei einer Active/Standby-Konfiguration wie Pilot Light geht der gesamte Datenverkehr zunächst an die primäre Region und wechselt zur Notfallwiederherstellungsregion, wenn die primäre Region nicht mehr verfügbar ist. Es gibt zwei Optionen für die Verwaltung des Datenverkehrs mit AWS-Services. Die erste Option ist die Verwendung von Amazon Route 53. Mit Amazon Route 53 können Sie mehrere IP-Endpunkte in einer oder mehreren AWS-Regionen mit einer Route 53-Domäne verknüpfen. Dann können Sie den Datenverkehr an den entsprechenden Endpunkt unter diesem Domänennamen weiterleiten. Amazon Route 53-Status-Checks überwachen diese Endpunkte.

Mithilfe dieser Status-Checks können Sie DNS-Failovers konfigurieren, um sicherzustellen, dass der Datenverkehr an funktionsfähige Endpunkte geleitet wird.

Die zweite Möglichkeit ist die Verwendung von AWS Global Accelerator. Mit AnyCast IP können Sie mehrere Endpunkte in einer oder mehreren AWS-Regionen mit derselben statischen IP-Adresse bzw. denselben statischen IP-Adressen verknüpfen. AWS Global Accelerator leitet den Datenverkehr dann an den entsprechenden Endpunkt weiter, der mit dieser Adresse verknüpft ist. Global Accelerator-Status-Checks überwachen die Endpunkte. Mithilfe dieser Status-Checks prüft AWS Global Accelerator automatisch den Zustand Ihrer Anwendungen und leitet den Datenverkehr der Benutzer nur an einen funktionsfähigen Anwendungsendpunkt weiter. Global Accelerator bietet geringere Latenzzeiten zum Anwendungsendpunkt, da es das umfangreiche AWS-Edge-Netzwerk nutzt, um den Datenverkehr so schnell wie möglich auf das AWS-Netzwerk-Backbone zu routen. Global Accelerator vermeidet außerdem Caching-Probleme, die bei DNS-Systemen (wie Route 53) auftreten können.

CloudEndure Disaster Recovery

CloudEndure Disaster Recovery ist über AWS Marketplace verfügbar und repliziert kontinuierlich servergehostete Anwendungen und servergehostete Datenbanken aus jeder beliebigen Quelle in AWS, indem der zugrunde liegende Server auf Blockebene repliziert wird. Mit CloudEndure Disaster Recovery können Sie die AWS Cloud als Notfallwiederherstellungsregion für einen lokalen Workload und seine Umgebung nutzen. Der Service kann auch für die Notfallwiederherstellung von in AWS gehosteten Workloads verwendet werden, wenn diese nur aus Anwendungen und Datenbanken bestehen, die in EC2 gehostet werden (d. h. nicht RDS). CloudEndure Disaster Recovery verwendet die Pilot Light-Strategie, bei der eine Kopie der Daten und der deaktivierten Ressourcen in einer Amazon Virtual Private Cloud (Amazon VPC) aufbewahrt wird, die als Staging-Area dient. Wenn ein Failover-Ereignis ausgelöst wird, werden die bereitgestellten Ressourcen verwendet, um automatisch eine Bereitstellung mit voller Kapazität in der Amazon VPC zu erstellen, die als Wiederherstellungsort dient.

Abbildung 10 - CloudEndure Disaster Recovery-Architektur

Warm Standby

Beim Warm Standby-Ansatz wird sichergestellt, dass eine verkleinerte, aber voll funktionsfähige Kopie Ihrer Produktionsumgebung in einer anderen Region vorhanden ist. Dieser Ansatz erweitert das Pilot Light-Konzept und verkürzt die Zeit bis zur Wiederherstellung, da Ihr Workload in einer

anderen Region ständig aktiv ist. Mit diesem Ansatz können Sie auch leichter Tests durchführen oder kontinuierliche Tests implementieren, um das Vertrauen in Ihre Fähigkeit zur Wiederherstellung nach einer Katastrophe zu erhöhen.

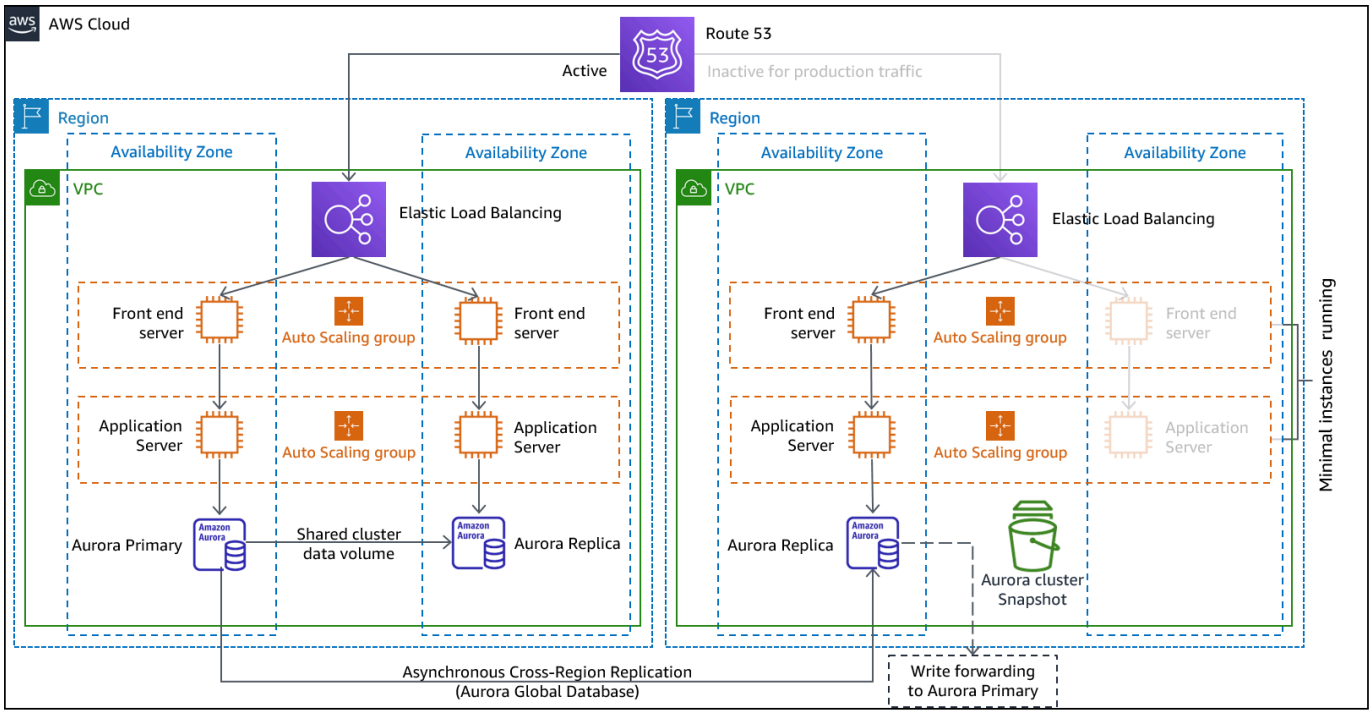

Abbildung 11 - Warm Standby-Architektur

Hinweis: Der Unterschied zwischen Pilot Light und Warm Standby kann mitunter nicht eindeutig sein. Beide beinhalten eine Umgebung in Ihrer DR-Region mit Kopien der Assets Ihrer primären Region. Der Unterschied besteht darin, dass Pilot Light keine Anfragen bearbeiten kann, ohne dass zuvor zusätzliche Maßnahmen ergriffen werden, während Warm Standby den Datenverkehr (mit reduzierter Kapazität) sofort bearbeiten kann. Der Pilot Light-Ansatz erfordert, dass Sie Server "einschalten" und möglicherweise zusätzliche (nicht zum Kerngeschäft gehörende) Infrastruktur bereitstellen und hochskalieren, während Sie bei Warm Standby nur hochskalieren müssen (alles ist bereits bereitgestellt und läuft). Verwenden Sie Ihre RTO- und RPO-Anforderungen, um zwischen diesen Ansätzen zu wählen.

AWS-Services

Alle AWS-Services, die unter Backup und Wiederherstellung und Pilot Light aufgeführt sind, werden auch im Warm Standby-Konzept für die Datensicherung, die Datenreplikation, die Weiterleitung des Datenverkehrs im aktiven Standby-Modus und die Bereitstellung der Infrastruktur einschließlich EC2-Instances verwendet.

AWS Auto Scaling wird zur Skalierung von Ressourcen wie Amazon EC2-Instance, Amazon ECS-Aufgaben, Amazon DynamoDB-Durchsatz und Amazon Aurora-Replikate innerhalb einer AWS Region verwendet. Amazon EC2 Auto Scaling skaliert die Bereitstellung von EC2-Instances über Availability Zones innerhalb einer AWS Region und sorgt für Ausfallsicherheit innerhalb dieser Region. Nutzen Sie Auto Scaling, um Ihre DR-Region im Rahmen einer Pilot Light- oder Warm Standby-Strategie auf volle Produktionsleistung zu skalieren. Erhöhen Sie zum Beispiel für EC2 die Einstellung Gewünschte Kapazität in der Auto Scaling-Gruppe. Sie können diese Einstellung manuell über die AWS Management Console, automatisch über das AWS SDK oder durch erneute Bereitstellung Ihrer AWS CloudFormation-Vorlage mit dem neuen Wert für die gewünschte Kapazität anpassen. Sie können AWS CloudFormation-Parameter verwenden, um die Neubereitstellung der CloudFormation-Vorlage zu erleichtern. Stellen Sie sicher, dass die Service-Kontingente in Ihrer DR-Region hoch genug sind, sodass die Hochskalierung auf die Produktionsleistung nicht behindert wird.

Multi-Site Active/Active

Sie können Ihren Workload im Rahmen einer Multi-Site Active/Active- oder Hot Standby Active/Passive-Strategie gleichzeitig in mehreren Regionen ausführen. Multi-Site Active/Active bedient den Datenverkehr aus allen Regionen, für die es bereitgestellt wird, während Hot Standby den Datenverkehr nur aus einer einzigen Region bedient und die andere(n) Region(en) nur für die Notfallwiederherstellung verwendet wird/werden. Bei einem Multi-Site Active/Active-Ansatz können Benutzer auf Ihren Workload in jeder der Regionen zugreifen, in denen er bereitgestellt ist. Dieser Ansatz ist der komplexeste und kostspieligste Ansatz für die Notfallwiederherstellung, kann aber bei den meisten Katastrophen mit der richtigen Technologieauswahl und -implementierung die Wiederherstellungszeit auf nahezu null reduzieren (bei einer Datenbeschädigung müssen Sie sich jedoch möglicherweise auf Backups verlassen, was in der Regel zu einem Wiederherstellungspunkt ungleich null führt). Hot Standby verwendet eine Active/Passive-Konfiguration, bei der die Benutzer nur zu einer einzigen Region geleitet werden und die DR-Regionen keinen Datenverkehr aufnehmen. Die meisten Kunden sind der Ansicht, dass beim Aufbau einer vollständigen Umgebung in der zweiten Region eine Active/Active-Konfiguration sinnvoll ist. Wenn Sie den Datenverkehr nicht über beide Regionen abwickeln möchten, bietet Warm Standby einen wirtschaftlicheren und operativ weniger komplexen Ansatz.

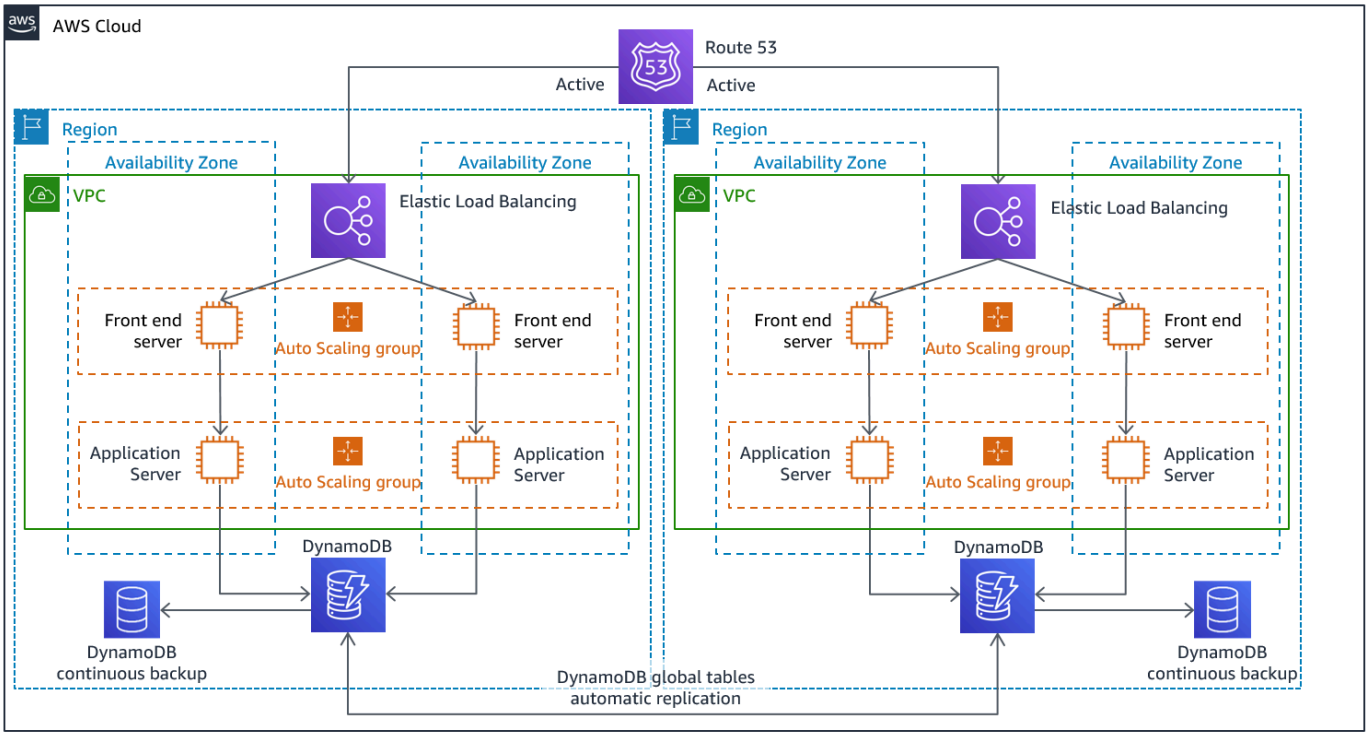

Abbildung 12 - Multi-Site Active/Active-Architektur (für Hot-Standby müssen Sie einen Active-Pfad in Inactive ändern)

Bei Multi-Site Active/Active gibt es in diesem Szenario kein Failover, da der Workload in mehr als einer Region ausgeführt wird. Notfallwiederherstellungstests würden sich in diesem Fall darauf konzentrieren, wie der Workload auf den Ausfall einer Region reagiert: Wird der Datenverkehr von der ausgefallenen Region weggeleitet? Kann/können die andere(n) Region(en) den gesamten Datenverkehr bewältigen? Auch Tests für eine Datenkatastrophe sind erforderlich. Backup und Wiederherstellung sind weiterhin erforderlich und sollten regelmäßig getestet werden. Es sollte auch beachtet werden, dass die Wiederherstellungszeiten für eine Datenkatastrophe mit Datenbeschädigung, -löschung oder -verschleierung immer größer als null sein werden und der Wiederherstellungspunkt immer an einem Punkt vor der Entdeckung der Katastrophe liegen wird. Wenn die zusätzliche Komplexität und die Kosten eines Multi-Site Active/Active- (oder Hot-Standby)-Ansatzes erforderlich sind, um die Wiederherstellungszeiten nahe null zu halten, dann sollten zusätzliche Anstrengungen unternommen werden, um die Sicherheit aufrechtzuerhalten und menschliches Versagen zu verhindern, um menschliche Katastrophen abzufedern.

AWS-Services

Alle AWS-Services, die unter Backup und Wiederherstellung, Pilot Light und Warm Standby behandelt werden, werden hier auch für die Point-in-Time-Datensicherung, die Datenreplikation, die Weiterleitung des Active/Active-Datenverkehrs und die Bereitstellung und Skalierung der Infrastruktur einschließlich EC2-Instances verwendet.

Für die bereits besprochenen Active/Passive-Szenarien (Pilot Light und Warm Standby) können sowohl Amazon Route 53 als auch AWS Global Accelerator für die Weiterleitung des Datenverkehrs zur aktiven Regionen verwendet werden. Für die Active/Active-Strategie ermöglichen beide Services auch die Definition von Richtlinien, die festlegen, welche Benutzer zu welchem aktiven regionalen Endpunkt gehen. Mit AWS Global Accelerator legen Sie eine Datenverkehrsverteilung fest, um den Prozentsatz des Datenverkehrs zu steuern, der an jeden Anwendungsendpunkt geleitet wird. Amazon Route 53 unterstützt diesen prozentualen Ansatz und darüber hinaus mehrere andere verfügbare Richtlinien, einschließlich Richtlinien für die geografische Verteilung und die Latenz. Global Accelerator nutzt automatisch das umfangreiche Netzwerk von AWS-Edge-Servern, um den Datenverkehr so schnell wie möglich an das AWS-Netzwerk-Backbone weiterzuleiten, was zu geringeren Latenzen bei Anfragen führt.

Die Datenreplikation mit dieser Strategie ermöglicht ein RPO nahe null. AWS-Services wie die globale Amazon Aurora-Datenbank nutzen eine dedizierte Infrastruktur, die Ihre Datenbanken vollständig für Ihre Anwendung verfügbar macht, und können mit einer typischen Latenz von unter einer Sekunde in eine sekundäre Region replizieren. Bei Active/Passive-Strategien erfolgen Schreibvorgänge nur in der primären Region. Der Unterschied zu Active/Active besteht darin, wie die Schreibvorgänge in jeder aktiven Region gehandhabt werden. Üblicherweise werden Lesevorgänge von der Region ausgeführt, die dem Benutzer am nächsten ist, was als Read Local bezeichnet wird. Bei Schreibvorgängen haben Sie mehrere Möglichkeiten:

- Eine Write Global Strategie routet alle Schreibvorgänge an eine einzige Region. Sollte diese Region ausfallen, wird eine andere Region zur Annahme von Schreibvorgängen befördert. Die globale Aurora-Datenbank eignet sich gut für Write Global, da sie die Synchronisierung mit Lese-Replikaten über die Regionen hinweg unterstützt und Sie eine der sekundären Regionen in weniger als 1 Minute zur Übernahme von Lese-/Schreibaufgaben hochstufen können.
- Eine Write Local-Strategie routet Schreibvorgänge an die nächstgelegene Region (genau wie Lesevorgänge). Globale Amazon DynamoDB-Tabellen ermöglichen eine solche Strategie und Lese- und Schreibzugriffe aus jeder Region, in der Ihre globale Tabelle bereitgestellt wird. Globale Amazon DynamoDB-Tabellen verwenden einen Last Writer Wins-Abgleich bei gleichzeitigen Aktualisierungen.

- Eine Write Partitioned-Strategie weist Schreibvorgänge einer bestimmten Region auf der Grundlage eines Partitionsschlüssels (wie der Benutzer-ID) zu, um Schreibkonflikte zu vermeiden. In diesem Fall kann die bidirektional konfigurierte Amazon S3-Replikation verwendet werden. Sie unterstützt derzeit die Replikation zwischen zwei Regionen. Stellen Sie bei der Implementierung dieses Ansatzes sicher, dass Sie replica modification sync für beide Buckets (A und B) aktivieren, um Änderungen an Replikat-Metadaten wie Objekt-Zugriffskontrolllisten (ACLs), Objekt-Tags oder Objektsperren für die replizierten Objekten zu replizieren. Sie können außerdem konfigurieren, ob Löschmarkierungen zwischen Buckets in Ihren aktiven Regionen repliziert werden sollen. Neben der Replikation muss Ihre Strategie auch Point-in-Time-Backups zum Schutz vor Datenbeschädigung oder -zerstörung vorsehen.

AWS CloudFormation ist ein leistungsstarkes Tool, um eine einheitlich bereitgestellte Infrastruktur zwischen AWS-Konten in mehreren AWS-Regionen durchzusetzen. AWS CloudFormation StackSets erweitert diese Funktionalität, indem es Ihnen ermöglicht, CloudFormation-Stacks über mehrere Konten und Regionen hinweg mit einem einzigen Vorgang zu erstellen, zu aktualisieren oder zu löschen. AWS CloudFormation nutzt YAML oder JSON, um Infrastruktur as Code zu definieren. Sie können mit AWS Cloud Development Kit (AWS CDK) jedoch Ihnen bekannte Programmiersprachen nutzen, um die Infrastructure as Code-Lösungen zu definieren. Ihr Code wird in CloudFormation umgewandelt und dann zum Bereitstellen von Ressourcen in AWS verwendet.

Erkennung

Es ist wichtig, so früh wie möglich zu wissen, dass Ihre Workloads nicht die geschäftlichen Ergebnisse liefern, die sie liefern sollten. Auf diese Weise können Sie schnell einen Katastrophenfall auslösen und für eine Wiederherstellung nach einem Vorfall sorgen. Bei aggressiven Wiederherstellungszielen ist diese Reaktionszeit in Verbindung mit angemessenen Informationen entscheidend für das Erreichen der Wiederherstellungsziele. Wenn Ihr Recovery Point Objective eine Stunde beträgt, müssen Sie den Vorfall erkennen, die zuständigen Mitarbeiter benachrichtigen, Ihre Eskalationsprozesse einleiten, Informationen (falls vorhanden) über die voraussichtliche Zeit bis zur Wiederherstellung auswerten (ohne den Notfallwiederherstellungsplan auszuführen), einen Notfall auslösen und innerhalb einer Stunde wiederherstellen.

> **Note**
>
> Wenn die Beteiligten beschließen, die DR nicht auszulösen, obwohl das RTO gefährdet wäre, sollten Sie die DR-Pläne und -Ziele neu bewerten. Die Entscheidung, die DR-Pläne nicht auszulösen, kann darauf zurückzuführen sein, dass die Pläne unzureichend sind oder dass es an Vertrauen in die Ausführung mangelt.

Es ist von entscheidender Bedeutung, dass Sie die Erkennung, Benachrichtigung, Eskalation, Ermittlung und Auslösung von Vorfällen in Ihre Planung und Ziele einbeziehen, um realistische, erreichbare Ziele zu schaffen, die einen geschäftlichen Nutzen bieten.

AWS veröffentlicht die aktuellen Informationen zur Verfügbarkeit von Services im Dashboard zum Service-Status. Sie können sich jederzeit über den aktuellen Status informieren oder einen RSS-Feed abonnieren, um über Unterbrechungen bei den einzelnen Services informiert zu werden. Wenn Sie ein Echtzeit-Problem mit einem unserer Services haben, das nicht im Dashboard zum Service-Status angezeigt wird, können Sie eine Supportanfrage stellen.

Das AWS Health Dashboard bietet Informationen zu AWS Health-Ereignissen, die Ihr Konto betreffen können. Diese Informationen werden auf zweierlei Weise bereitgestellt: in einem Dashboard, in dem aktuelle und anstehende Ereignisse sortiert nach Kategorie angezeigt werden, und in einem vollständigen Protokoll, in dem alle Ereignisse der letzten 90 Tage angezeigt werden.

Für die strengsten RTO-Anforderungen können Sie automatische Failovers auf der Basis von Status-Checks implementieren. Entwerfen Sie Status-Checks, die repräsentativ für die Benutzererfahrung

sind und auf KPIs (Key Performance Indicator) basieren. Tiefgreifende Status-Checks überprüfen wichtige Funktionen Ihres Workloads und gehen über oberflächliche Heartbeat-Prüfungen hinaus. Verwenden Sie tiefgreifende Status-Checks, die auf mehreren Signalen basieren. Seien Sie bei diesem Ansatz vorsichtig, sodass Sie keinen Fehlalarm auslösen. Ein unnötiger Failover kann an sich schon ein Verfügbarkeitsrisiko darstellen.

Testen der Notfallwiederherstellung

Testen Sie die Implementierung der Notfallwiederherstellung, um die Implementierung zu validieren. Testen Sie regelmäßig den Failover zur DR-Region Ihres Workloads, um sicherzustellen, dass RTO und RPO eingehalten werden.

Ein Modell zur Vermeidung ist die Entwicklung von Wiederherstellungspfaden, die selten ausgeführt werden. So könnten Sie beispielsweise einen zweiten Datenspeicher unterhalten, der nur für Leseabfragen verwendet wird. Wenn Sie Daten in einen Datenspeicher schreiben und der primäre Datenspeicher einen Fehler ausgibt, können Sie einen Failover auf den zweiten Datenspeicher durchführen. Wenn Sie diesen Failover nicht regelmäßig testen, werden Sie möglicherweise feststellen, dass Ihre Annahmen zu den Möglichkeiten des sekundären Datenspeichers unzutreffend sind. Die Kapazität der sekundären Region, die beim letzten Test vielleicht noch ausreichend war, kann die Last in diesem Szenario möglicherweise nicht mehr bewältigen, oder die Service-Kontingente in der sekundären Region reichen nicht aus.

Unsere Erfahrungen haben gezeigt, dass bei einer Wiederherstellung nach einem Fehler nur der Pfad funktioniert, den Sie regelmäßig testen. Aus diesem Grund ist es am besten, eine kleine Anzahl von Wiederherstellungspfaden zu nutzen.

Sie können Wiederherstellungsmuster erstellen und diese regelmäßig testen. Wenn Sie einen komplexen oder kritischen Wiederherstellungspfad nutzen, müssen Sie diesen Ausfall dennoch regelmäßig in der Produktion testen, um zu überprüfen, ob der Wiederherstellungspfad funktioniert.

Verwalten Sie die Konfigurationsabweichung in der DR-Region. Stellen Sie sicher, dass Ihre Infrastruktur, Daten und Konfiguration in der DR-Region den Anforderungen entsprechen. Prüfen Sie zum Beispiel, ob AMIs und Service-Kontingente auf dem neuesten Stand sind.

Sie können AWS Config verwenden, um Ihre AWS-Ressourcenkonfigurationen kontinuierlich zu überwachen und aufzuzeichnen. AWS Config kann eine Abweichung erkennen und AWS Systems Manager Automation zur Behebung von Abweichungen und für Alarme nutzen. AWS CloudFormation kann zusätzlich Abweichungen in bereitgestellten Stacks erkennen.

Fazit

Die Kunden sind für die Verfügbarkeit ihrer Anwendungen in der Cloud verantwortlich. Sie müssen definieren, was eine Katastrophe ist, und einen Notfallwiederherstellungsplan aufstellen, der diese Definition und die möglichen Auswirkungen auf die Geschäftsergebnisse widerspiegelt. Erstellen Sie ein Recovery Time Objective (RTO) und Recovery Point Objective (RPO) auf Basis von Auswirkungsanalysen und Risikobewertungen und wählen Sie dann die geeignete Architektur, um sich gegen Katastrophen abzusichern. Stellen Sie sicher, dass die Erkennung von Katastrophen möglich ist und rechtzeitig erfolgt. Dies ist von entscheidender Bedeutung, um festzustellen, wann die Ziele gefährdet sind. Stellen Sie sicher, dass Sie einen Plan haben und diesen durch Tests validieren. Bei nicht validierten Notfallwiederherstellungsplänen besteht die Gefahr, dass sie nicht umgesetzt werden, weil das Vertrauen fehlt oder die Ziele für die Notfallwiederherstellung nicht erreicht werden.

Mitwirkende

An diesem Dokument haben folgende Personen mitgewirkt:

- Alex Livingstone, Practice Lead Cloud Operations, AWS Enterprise Support
- Seth Eliot, Principal Reliability Solutions Architect, Amazon Web Services

Weitere Informationen

Weitere Informationen finden Sie unter:

- Reliability Pillar, AWS Well-Architected Framework
- Checkliste für Notfallwiederherstellungspläne
- Implementierung von Status-Checks
- AWS-Lösungen implementieren: Multi-Regionale-Anwendungsarchitektur
- AWS re:Invent 2018: Architekturmuster für Active-Active-Anwendungen in mehreren Regionen (ARC209-R2)

Dokumentverlauf

Änderung	Beschreibung	Datum
Erste Veröffentlichung	Erstveröffentlichung	12. Februar 2021

Abonnieren Sie den RSS-Feed, um über Aktualisierungen des Whitepapers benachrichtigt zu werden.

Hinweise

Kunden sind eigenverantwortlich für die unabhängige Bewertung der Informationen in diesem Dokument zuständig. Dieses Dokument: (a) dient rein zu Informationszwecken, (b) spiegelt die aktuellen Produktangebote und Verfahren von AWS wider, die sich ohne vorherige Mitteilung ändern können, und (c) impliziert keinerlei Verpflichtungen oder Zusicherungen seitens AWS und dessen Tochtergesellschaften, Lieferanten oder Lizenzgebern. AWS-Produkte oder -Services werden im vorliegenden Zustand und ohne ausdrückliche oder stillschweigende Gewährleistungen, Zusicherungen oder Bedingungen bereitgestellt. Die Verantwortung und Haftung von AWS gegenüber seinen Kunden wird durch AWS-Vereinbarungen geregelt. Dieses Dokument ist weder ganz noch teilweise Teil der Vereinbarungen zwischen AWS und seinen Kunden und ändert diese Vereinbarungen auch nicht.