

AWS Whitepaper

AWS-Fehlerisolierungsgrenzen



AWS-Fehlerisolierungsgrenzen: AWS Whitepaper

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. Alle anderen Handelsmarken, die nicht Eigentum von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise zu Amazon gehören oder nicht, mit Amazon verbunden sind oder von Amazon gesponsert werden.

Table of Contents

Zusammenfassung und Einführung	1
Überblick	1
Sind Sie Well-Architected?	1
Einführung	1
AWS-Infrastruktur	3
Availability Zones	3
Regionen	4
AWS Lokale Zonen	5
AWS Outposts	5
Points of Presence	6
Partitionen	7
Steuerebenen und Datenebenen	7
Statische Stabilität	8
Übersicht	9
AWS -Servicetypen	10
Zonengebundene Services	10
Regionale Services	13
Globale Services	14
Globale Services, die nach Partition eindeutig sind	15
Globale Services im Edge-Netzwerk	17
Globale Einzelregionsoperationen	18
Services, die globale Standardendpunkte verwenden	22
Globale Serviceübersicht	24
Schlussfolgerung	28
Anhang A — Anleitung zum partitionellen Service	29
AWSICHBIN	29
AWS Organizations	29
AWS-Kontenverwaltung	30
Route 53 Application Recovery-Controller	31
AWS-Network Manager	31
Route 53 Privates DNS	32
Anhang B — Globale Servicehinweise für Edge-Netzwerke	33
Route 53	33
Amazon CloudFront	34

Amazon Certificate Manager	34
AWSWeb Application Firewall (WAF) und WAF Classic	34
AWS Global Accelerator	35
Amazon S3 Shield	35
Anhang C — Dienste für einzelne Regionen	37
Beitragende Faktoren	38
Dokumentversionen	39
AWS-Glossar	40
Hinweise	41
.....	xlii

AWS Fault Lens

Datum der Veröffentlichung: 16. November 2022 ([Dokumentversionen](#))

Überblick

Amazon Web Services (AWS) bietet verschiedene Isolationsgrenzen wie Availability Zones (AZ), Regionen, Steuerungsebenen und Datenebenen. In diesem paper wird detailliert beschrieben, wie diese Grenzen AWS genutzt werden, um zonale, regionale und globale Dienste zu schaffen. Es enthält auch präskriptive Anleitungen zur Berücksichtigung von Abhängigkeiten von diesen verschiedenen Diensten und zur Verbesserung der Resilienz von Workloads, die Sie mithilfe dieser Dienste erstellen.

Sind Sie Well-Architected?

Das [AWS Well-Architected Framework](#) hilft Ihnen dabei, die Vor- und Nachteile der Entscheidungen zu verstehen, die Sie beim Aufbau von Systemen in der Cloud treffen. Die sechs Säulen des Frameworks ermöglichen es Ihnen, bewährte architektonische Verfahren für den Entwurf und Betrieb zuverlässiger, sicherer, effizienter, kostengünstiger und nachhaltiger Systeme zu erlernen. Mithilfe der [AWS Well-Architected Tool](#), die kostenlos in der verfügbar ist [AWS Management Console](#), können Sie Ihre Workloads anhand dieser Best Practices überprüfen, indem Sie für jede Säule eine Reihe von Fragen beantworten.

[Weitere Ratschläge von Experten und Best Practices für Ihre Cloud-Architektur — Referenzarchitekturbereitstellungen, Diagramme und Whitepapers — finden Sie im Architecture Center. AWS](#)

Einführung

AWSbetreibt eine globale Infrastruktur zur Bereitstellung von Cloud-Diensten, mit denen Kunden Workloads auf flexible, sichere, skalierbare und hochverfügbare Weise bereitstellen können. Die AWS Infrastruktur verwendet mehrere Fehlerisolutionskonstrukte, um Kunden dabei zu unterstützen, ihre Resilienzziele zu erreichen. Diese Grenzen zur Fehlerisolation ermöglichen es Kunden, ihre Workloads so zu gestalten, dass sie den vorhersehbaren Umfang der damit verbundenen Auswirkungen nutzen. Es ist auch wichtig zu verstehen, wie AWS Dienste unter Berücksichtigung

dieser Grenzen konzipiert werden, damit Sie bewusst Entscheidungen über die Abhängigkeiten treffen können, die Sie für Ihren Workload auswählen.

In diesem paper werden zunächst die AWS globale Infrastruktur und die damit verbundenen Grenzen zur Fehlerisolation sowie einige der Muster zusammengefasst, die bei der Gestaltung unserer Dienste verwendet wurden. Ausgehend von diesem grundlegenden Verständnis werden in dem paper als Nächstes die verschiedenen Leistungsumfänge skizziert: AWS zonale, regionale und globale Dienstleistungen. Außerdem werden bewährte Verfahren für den Aufbau von Architekturen vorgestellt, die diese Isolationsgrenzen und verschiedene Servicebereiche nutzen, um die Resilienz der Workloads, auf denen Sie ausgeführt werden, zu verbessern. AWS Insbesondere enthält es präskriptive Leitlinien dafür, wie Abhängigkeiten von globalen Diensten beseitigt und gleichzeitig einzelne Fehlerquellen minimiert werden können. Auf diese Weise können Sie fundierte Entscheidungen über Ihre AWS Abhängigkeiten und die Gestaltung Ihres Workloads für Hochverfügbarkeit (HA) und Disaster Recovery (DR) treffen.

AWS-Infrastruktur

Dieser Abschnitt enthält eine Zusammenfassung der AWS globalen Infrastruktur und der Grenzen der Fehlerisolierung, die sie bereitstellt. Darüber hinaus bietet dieser Abschnitt einen Überblick über das Konzept der Steuerebenen und Datenebenen, die entscheidende Unterschiede bei der AWS Gestaltung seiner Services sind. Diese Informationen bieten eine Grundlage, um zu verstehen, wie die Grenzen der Fehlerisolierung und die Steuerebene und die Datenebene eines Services auf die AWS Servicetypen angewendet werden, die wir im nächsten Abschnitt besprechen.

Themen

- [Availability Zones](#)
- [Regionen](#)
- [AWS Lokale Zonen](#)
- [AWS Outposts](#)
- [Points of Presence](#)
- [Partitionen](#)
- [Steuerebenen und Datenebenen](#)
- [Statische Stabilität](#)
- [Übersicht](#)

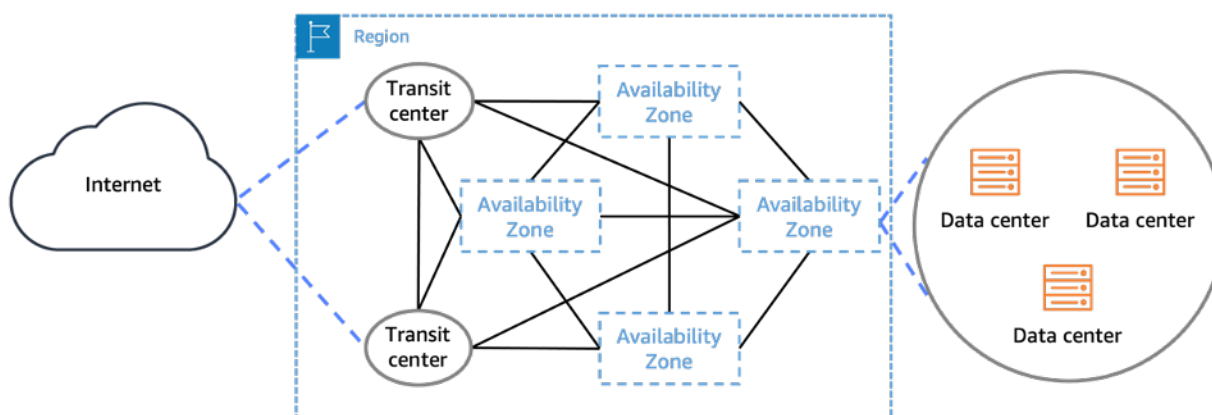
Availability Zones

AWS arbeitet über 100 Availability Zones in mehreren -Regionen weltweit (aktuelle Nummern finden Sie hier: [AWS Globale Infrastruktur](#)). Eine Availability Zone ist ein oder mehrere diskrete Rechenzentren mit unabhängiger und redundanter Strominfrastruktur, Netzwerk und Konnektivität in einem AWS-Region. Availability Zones in einer Region sind deutlich voneinander entfernt, bis zu 60 Fuß (~100 km), um korrelierte Ausfälle zu vermeiden, aber nahe genug, um die synchrone Replikation mit Latenz im einstelligen Millisekundenbereich zu verwenden. Sie sind so konzipiert, dass sie nicht gleichzeitig von einem Szenario mit gemeinsam genutztem Kabel wie Netzstrom, Wasserunterbrechung, Glasfaserisolierung, Erdbeben, Bränden, Tornados oder Überflutungen betroffen sind. Häufige Fehlerquellen, wie Generatoren und Schutz-ausrüstung, werden nicht über Availability Zones hinweg gemeinsam genutzt und sind für die Bereitstellung durch unabhängige Stromunterstations konzipiert. Wenn Updates für seine Services AWS bereitstellt, werden

Bereitstellungen in Availability Zones in derselben Region zeitlich getrennt, um korrelierte Ausfälle zu vermeiden.

Alle Availability Zones in einer Region sind mit Netzwerken mit hoher Bandbreite und niedriger Latenz über vollständig redundante, dedizierte -Metro-Glasfaser miteinander verbunden. Jede Availability Zone in einer Region stellt über zwei Transitzentren eine Verbindung zum Internet her, in denen AWS Peers mit mehreren [Tier-1-Internetanbietern](#) verbunden sind (weitere Informationen finden Sie unter [Übersicht über Amazon Web Services](#)).

Diese Funktionen bieten eine starke Isolation der Availability Zones voneinander, was wir als Availability Zone Independence (AZI) bezeichnen. Das logische Konstrukt von Availability Zones und ihre Konnektivität zum Internet ist in der folgenden Abbildung dargestellt.



Availability Zones bestehen aus einem oder mehreren physischen Rechenzentren, die redundant miteinander und mit dem Internet verbunden sind.

Regionen

Jede besteht AWS-Region aus mehreren unabhängigen und physisch separaten Availability Zones innerhalb eines geografischen Gebiets. Alle Regionen verfügen derzeit über drei oder mehr Availability Zones. Die Regionen selbst sind isoliert und unabhängig von anderen Regionen, mit einigen Ausnahmen, die später in diesem Dokument erwähnt werden ([siehe Globale Einzelregionsoperationen](#)). Diese Trennung zwischen Regionen beschränkt Servicefehler auf eine einzelne Region, wenn sie auftreten. Der normale Betrieb anderer Regionen bleibt in diesem Fall davon unberührt. Darüber hinaus sind die Ressourcen und Daten, die Sie in einer Region erstellen, in keiner anderen Region vorhanden, es sei denn, Sie verwenden explizit eine Replikations- oder Kopierfunktion, die von einem -AWSService angeboten wird, oder replizieren die Ressource selbst.



Aktuelle und geplante AWS-Regionen seit Dezember 2022

AWS Lokale Zonen

[AWS Local Zones](#) sind eine Art der Infrastrukturbereitstellung, bei der Datenverarbeitungs-, Speicher-, Datenbank- und andere [ausgewählte -AWS Services](#) in der Nähe großer Populations- und Industriezentren platziert werden. Sie können -AWS Services wie Datenverarbeitungs- und Speicherservices in der Local Zone verwenden, um Anwendungen mit niedriger Latenz am Edge auszuführen oder Hybrid-Cloud-Migrationen zu vereinfachen. Local Zones verfügen über lokales Internet, um die Latenz zu reduzieren, sind aber auch über das redundante private Netzwerk von Amazon mit hoher Bandbreite mit ihrer übergeordneten Region verbunden, sodass Anwendungen, die in AWS Local Zones ausgeführt werden, schnell, sicher und nahtlos auf das gesamte Spektrum an -Services zugreifen können.

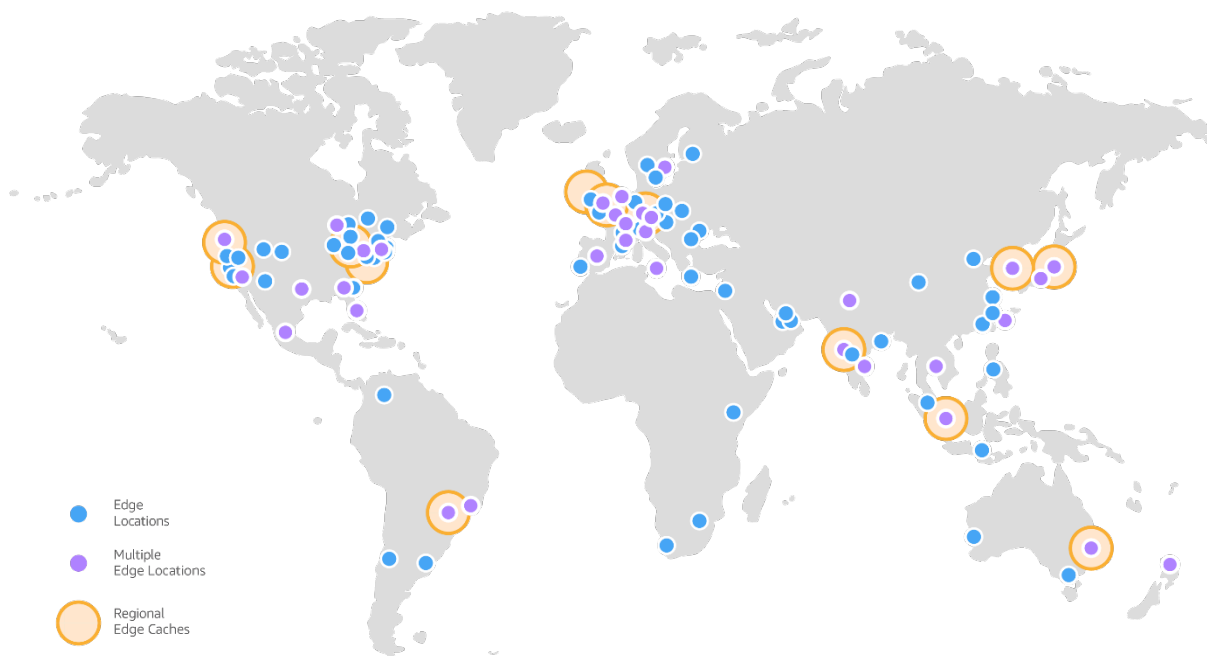
AWS Outposts

[AWS Outposts](#) ist eine Familie von vollständig verwalteten Lösungen, die AWS Infrastruktur und Services für praktisch jeden On-Premises- oder Edge-Standort für ein wirklich konsistentes Hybrid-Erlebnis bereitstellen. Outposts-Lösungen ermöglichen es Ihnen, native AWS Services On-Premises zu erweitern und auszuführen und sind in einer Vielzahl von Formfaktoren verfügbar, von 1U- und 2U-Outposts-Servern bis hin zu 42U-Outposts-Racks und mehreren Rack-Bereitstellungen.

Mit können Sie [ausgewählte -AWS Services](#) lokal ausführen und eine Verbindung zu einer Vielzahl von Services herstellen AWS Outposts, die in der übergeordneten verfügbar sind AWS-Region. AWS Outposts sind vollständig verwaltete und konfigurierbare Rechen AWS- und Speicher-Racks, die mit von entwickelter Hardware erstellt wurden, die es Kunden ermöglicht, Rechenleistung und Speicher On-Premises auszuführen und gleichzeitig nahtlos eine Verbindung mit AWS der breiten Palette von Services in der Cloud herzustellen.

Points of Presence

Zusätzlich zu den Availability Zones AWS-Regionen und betreibt AWS auch ein global verteiltes Point of Presence (PoP)-Netzwerk. Diese PoPs hosten Amazon CloudFront, ein Content Delivery Network (CDN), Amazon Route 53, einen öffentlichen Domain Name System (DNS)-Auflösungsservice, und AWS Global Accelerator (AGA), einen Edge-Netzwerkoptimierungsservice. Das globale Edge-Netzwerk besteht derzeit aus über 410 PoPs, darunter mehr als 400 Edge-Standorte, und 13 regionalen Zwischenspeichern der mittleren Ebene in über 90 Städten in 48 Ländern (der aktuelle Status finden Sie hier: [Amazon CloudFront Key Features](#)).



CloudFront Globales Edge-Netzwerk von Amazon

Jeder PoP ist von den anderen isoliert, was bedeutet, dass ein Fehler, der sich auf einen einzelnen PoP- oder metropolitanen Gebiet auswirkt, sich nicht auf den Rest des globalen Netzwerks auswirkt. Die AWS Netzwerk-Peers mit Tausenden von Tier-1/2/3-Telekommunikationsanbietern weltweit

sind gut mit allen wichtigen Zugriffsnetzwerken verbunden, um eine optimale Leistung zu erzielen, und verfügen über Hunderte von Terabit an bereitgestellter Kapazität. Edge-Standorte sind AWS-Regionen über das AWS Netzwerk-Backbone, eine vollständig redundante, mehrere 100GbE-Parallelfaser, die die Welt umkreist und mit Zehntausenden von Netzwerken verknüpft ist, um Ursprungsabrufe und dynamische Inhaltsbeschleunigung zu verbessern.

Partitionen

AWS gruppiert Regionen in [Partitionen](#). Jede -Region befindet sich genau in einer -Partition, und jede Partition hat eine oder mehrere Regionen. Partitionen haben unabhängige Instances von AWS Identity and Access Management (IAM) und bieten eine feste Grenze zwischen Regionen in verschiedenen Partitionen. AWS kommerzielle Regionen befinden sich in der -awsPartition, Regionen in China sind in der -aws-cnPartition und AWS GovCloud Regionen sind in der -aws-us-govPartition. Einige -AWSServices sind so konzipiert, dass sie regionsübergreifende Funktionen bieten, z. B. [regionsübergreifende Replikation in Amazon S3](#) oder [AWS regionsübergreifendes Transit-Gateway-Peering](#). Diese Arten von Funktionen werden nur zwischen Regionen in derselben Partition unterstützt. Sie können keine IAM-Anmeldeinformationen von einer Partition verwenden, um mit Ressourcen in einer anderen Partition zu interagieren.

Steuerebenen und Datenebenen

AWS unterteilt die meisten Services in die Konzepte der Steuerebene und der Datenebene. Diese Begriffe stammen aus der Welt des Netzwerks, insbesondere aus Routern. Die Datenebene des Routers, bei der es sich um ihre Hauptfunktionalität handelt, verschiebt Pakete basierend auf Regeln. Die Routing-Richtlinien müssen jedoch von irgendwo aus erstellt und verteilt werden, und dort kommt die Steuerebene ein.

Steuerebenen stellen die administrativen APIs bereit, die zum Erstellen, Lesen/Beschreiben, Aktualisieren, Löschen und Auflisten (CRUDL) von Ressourcen verwendet werden. Im Folgenden finden Sie beispielsweise alle Aktionen auf Steuerebene: Starten einer neuen [Amazon Elastic Compute Cloud](#) (Amazon EC2)-Instance, Erstellen eines [Amazon Simple Storage Service](#) (Amazon S3)-Buckets und Beschreiben einer [Amazon Simple Queue Service](#) (Amazon SQS)-Warteschlange. Wenn Sie eine EC2-Instance starten, muss die Steuerebene mehrere Aufgaben ausführen, z. B. das Auffinden eines physischen Hosts mit Kapazität, das Zuweisen der Netzwerkschnittstelle(n), das Vorbereiten eines [Amazon Elastic Block Store](#) (Amazon EBS)-Volumes, das Generieren von IAM-Anmeldeinformationen, das Hinzufügen der Sicherheitsgruppenregeln und vieles mehr. Steuerebenen sind in der Regel komplizierte Orchestrierungs- und Aggregationssysteme.

Die Datenebene stellt die primäre Funktion des Services bereit. Im Folgenden sind beispielsweise alle Teile der Datenebene für jeden der beteiligten Services aufgeführt: die laufende EC2-Instance selbst, das Lesen und Schreiben auf ein EBS-Volume, das Abrufen und Einfügen von Objekten in einen S3-Bucket und Route 53, das DNS-Abfragen beantwortet und Zustandsprüfungen durchführt.

Datenebenen sind absichtlich weniger kompliziert, mit weniger sich bewegenden Teilen als Steuerebenen, die normalerweise ein komplexes System von Workflows, Geschäftslogik und Datenbanken implementieren. Dadurch ist es statistisch weniger wahrscheinlich, dass Fehlerereignisse auf der Datenebene als auf der Steuerebene auftreten. Während sowohl die Daten- als auch die Steuerebene zum Gesamtbetrieb und Erfolg des Services beitragen, AWS betrachtet sie als unterschiedliche Komponenten. Diese Trennung hat sowohl Leistungs- als auch Verfügbarkeitsvorteile.

Statische Stabilität

Eines der wichtigsten Resilienzmerkmale von AWS Services ist die AWS statische Stabilität. Dieser Begriff bedeutet, dass Systeme in einem statischen Zustand betrieben werden und weiterhin normal funktionieren, ohne dass Änderungen während des Ausfalls oder der Nichtverfügbarkeit von Abhängigkeiten vorgenommen werden müssen. Eine Möglichkeit hierfür besteht darin, Zirkelbezüge in unseren Services zu verhindern, die verhindern könnten, dass einer dieser Services erfolgreich wiederhergestellt wird. Eine andere Möglichkeit, dies zu tun, besteht darin, den vorhandenen Status beizubehalten. Wir berücksichtigen die Tatsache, dass Steuerebenen statistisch wahrscheinlicher fehlschlagen als Datenebenen. Obwohl die Datenebene in der Regel von Daten abhängt, die von der Steuerebene kommen, behält die Datenebene ihren vorhandenen Zustand bei und funktioniert auch bei Beeinträchtigung der Steuerebene. Der Zugriff auf Ressourcen auf Datenebene, sobald er bereitgestellt wurde, ist nicht von der Steuerebene abhängig und ist daher nicht von einer Beeinträchtigung der Steuerebene betroffen. Mit anderen Worten, auch wenn die Fähigkeit zum Erstellen, Ändern oder Löschen von Ressourcen beeinträchtigt ist, bleiben vorhandene Ressourcen verfügbar. Dadurch sind AWS Datenebenen statisch stabil für eine Beeinträchtigung auf der Steuerebene. Sie können verschiedene Muster implementieren, um bei verschiedenen Arten von Abhängigkeitsfehlern statisch stabil zu sein.

Ein Beispiel für statische Stabilität finden Sie in Amazon EC2. Sobald eine EC2-Instance gestartet wurde, ist sie genauso verfügbar wie der physische Server in einem Rechenzentrum. Es ist nicht von APIs der Steuerebene abhängig, um weiter ausgeführt zu werden oder nach einem Neustart wieder ausgeführt zu werden. Die gleiche Eigenschaft gilt für andere AWS Ressourcen wie VPCs, Amazon-S3-Buckets und -Objekte sowie Amazon-EBS-Volumes. Amazon S3

Statische Stabilität ist ein Konzept, das bei der AWS Gestaltung seiner Services tief verwurzelt ist, aber es ist auch ein Muster, das von Kunden verwendet werden kann. Tatsächlich besteht ein Großteil der bewährten Methoden für die ausfallsichere Verwendung der verschiedenen Arten von AWS Services darin, statische Stabilität für Produktionsumgebungen zu implementieren. Die zuverlässigsten Wiederherstellungs- und Abschwächungsmechanismen sind diejenigen, die die wenigsten Änderungen erfordern, um eine Wiederherstellung zu erreichen. Anstatt sich darauf zu verlassen, dass die EC2-Steuerebene neue EC2-Instances startet, um sich von einer ausgefallenen Availability Zone zu erholen, trägt die Vorabbereitstellung dieser zusätzlichen Kapazität dazu bei, eine statische Stabilität zu erreichen. Daher trägt die Beseitigung von Abhängigkeiten von Steuerebenen (den APIs, die Änderungen an Ressourcen implementieren) in Ihrem Wiederherstellungspfad zu ausfallsichereren Workloads bei. Weitere Informationen zur statischen Stabilität, zu Steuerebenen und zu Datenebenen finden Sie im [Artikel Statische Stabilität der Amazon Builders' Library mit Availability Zones](#).

Übersicht

AWS verwendet verschiedene Fehlercontainer in unserer Infrastruktur, um eine Fehlerisolierung zu erstellen. Die Kerninfrastruktur-Fehlercontainer sind Partitionen, Regionen, Availability Zones, Steuerebenen und Datenebenen. Als Nächstes untersuchen wir verschiedene Arten von AWS Services, wie diese Fehlercontainer in ihrem Design verwendet werden und wie Sie Workloads mit ihnen so gestalten sollten, dass sie ausfallsicher sind.

AWS -Servicetypen

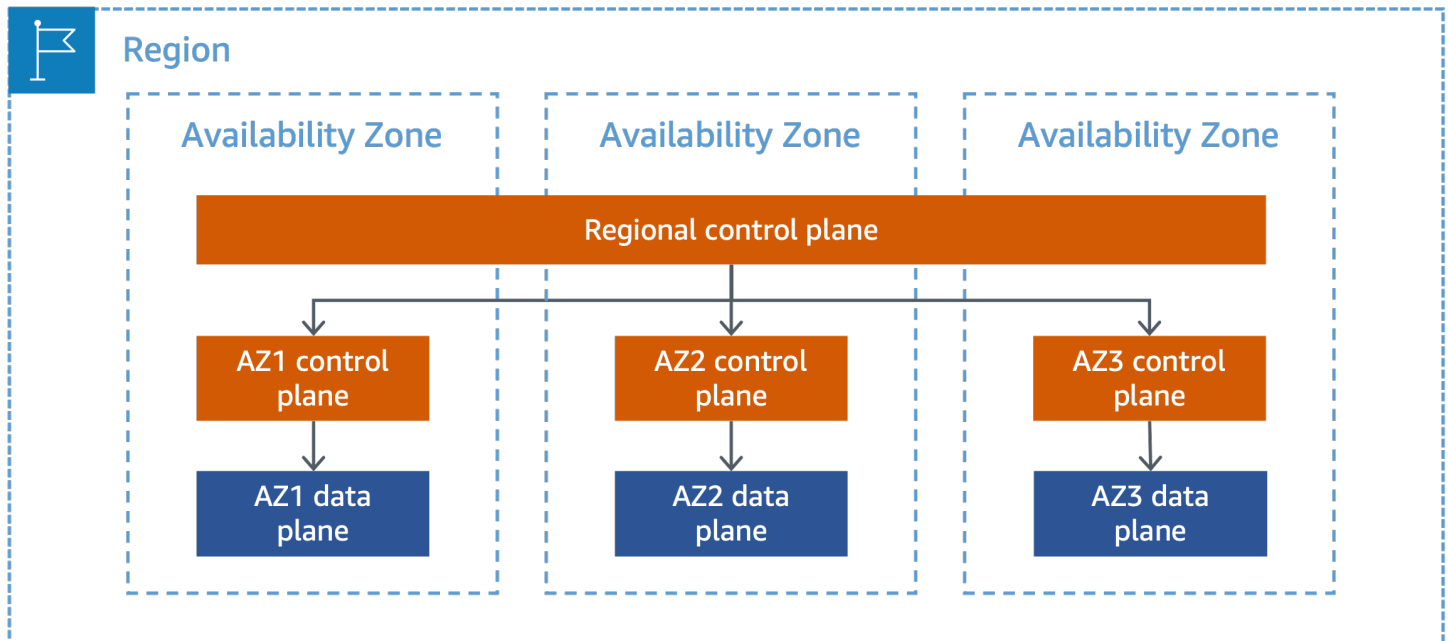
AWS betreibt drei verschiedene Kategorien von Services basierend auf ihrer Fehlerisolierungsgrenze: zonal, regional und global. In diesem Abschnitt wird ausführlicher beschrieben, wie diese verschiedenen Arten von Services konzipiert wurden, sodass Sie feststellen können, wie sich Ausfälle innerhalb eines Services eines bestimmten Servicetyps auf Ihren Workload auswirken, der auf ausgeführt wird AWS. Es bietet auch allgemeine Anleitungen dazu, wie Sie Ihre Workloads so gestalten können, dass diese Services ausfallsicher genutzt werden. Für globale -Services bietet dieses Dokument auch vorgeschriebene Anleitungen in [Anhang A — Anleitung zum partitionellen Service](#) und , [Anhang B — Globale Servicehinweise für Edge-Netzwerke](#) die Ihnen helfen können, Auswirkungen auf Ihre Workloads auf Beeinträchtigungen der Steuerebene in - AWS Services zu verhindern, und Ihnen helfen können, Abhängigkeiten von globalen Services sicher zu übernehmen und gleichzeitig die Einführung einzelner Fehlerpunkte zu minimieren.

Themen

- [Zonengebundene Services](#)
- [Regionale Services](#)
- [Globale Services](#)

Zonengebundene Services

[Availability Zone Independence](#) (A Bol) ermöglicht AWS es , zonale Services wie Amazon EC2 und Amazon EBS bereitzustellen. Ein zonaler Service bietet die Möglichkeit anzugeben, in welcher Availability Zone die Ressourcen bereitgestellt werden. Diese Services funktionieren unabhängig in jeder Availability Zone innerhalb einer Region und, was noch wichtiger ist, schlagen auch unabhängig in jeder Availability Zone fehl. Das bedeutet, dass Komponenten eines Services in einer Availability Zone keine Abhängigkeiten von Komponenten in anderen Availability Zones annehmen. Dies ist möglich, da ein zonaler Service zonale Datenebenen hat. In einigen Fällen, z. B. mit EC2, enthält der Service auch zonale Steuerebenen für zonell ausgerichtete Operationen, wie das Starten einer EC2-Instance. Für diese Services bietet AWS auch einen regionalen Endpunkt der Steuerebene, um die Interaktion mit dem Service zu vereinfachen. Die regionale Steuerebene bietet auch regionale Funktionen und dient als Aggregations- und Routing-Ebene über den zonalen Steuerebenen. Dies ist in der folgenden Abbildung dargestellt.



Ein zonaler Service mit zonenisolierten Steuerebenen und Datenebenen

Availability Zones bieten Kunden die Möglichkeit, Produktions-Workloads zu betreiben, die hochverfügbarer, fehlertoleranter und skalierbarer sind, als dies in einem einzigen Rechenzentrum möglich wäre. Wenn ein Workload mehrere Availability Zones verwendet, sind Kunden besser isoliert und vor Problemen geschützt, die sich auf die physische Infrastruktur einer einzelnen Availability Zone auswirken. Dies hilft Kunden dabei, Services zu erstellen, die über Availability Zones hinweg redundant sind, und, wenn sie korrekt konzipiert sind, betriebsbereit zu bleiben, auch wenn in einer Availability Zone Ausfälle auftreten. Kunden können die Vorteile von A nutzen, um hochverfügbare und belastbare Workloads zu erstellen. Die Implementierung von A in Ihrer Architektur hilft Ihnen, sich nach einem isolierten Ausfall der Availability Zone schnell zu erholen, da Ihre Ressourcen in einer Availability Zone die Interaktion mit Ressourcen in anderen Availability Zones minimieren oder eliminieren. Dies hilft, Abhängigkeiten zwischen Availability Zones zu entfernen, was die Evakuierung von Availability Zones vereinfacht. Weitere Informationen zur Erstellung von Availability-Zone-[Evakuierungsmechanismen finden Sie unter Erweiterte Multi-AZ-Ausfallsicherheitsmuster](#).

Darüber hinaus können Sie Availability Zones weiter nutzen, indem Sie einige der gleichen bewährten Methoden befolgen, die für ihre eigenen Services AWS verwendet, z. B. nur Änderungen gleichzeitig in einer einzelnen Availability Zone bereitstellen oder eine Availability Zone aus dem Service entfernen, wenn eine Änderung in dieser Availability Zone schlecht geht.

[Statische Stabilität](#) ist auch ein wichtiges Konzept für Architekturen mit mehreren Availability Zones. Einer der Fehlermodi, den Sie bei Architekturen mit mehreren Availability Zones einplanen sollten, ist der Verlust einer Availability Zone, was zum Verlust der Kapazität einer Availability Zone führen

kann. Wenn Sie nicht genügend Kapazität bereitgestellt haben, um den Verlust einer Availability Zone zu bewältigen, kann dies dazu führen, dass Ihre verbleibende Kapazität durch die aktuelle Last überfordert wird. Darüber hinaus müssen Sie von den Steuerebenen der Zonendienste abhängen, die Sie verwenden, um diese verlorene Kapazität zu ersetzen, was weniger zuverlässig sein kann als ein statisch stabiles Design. In diesem Fall kann Ihnen die Vorabbereitstellung von ausreichend zusätzlicher Kapazität helfen, statisch stabil gegenüber dem Verlust einer Fehlerdomäne, z. B. einer Availability Zone, zu sein, da Sie den normalen Betrieb fortsetzen können, ohne dass dynamische Änderungen erforderlich sind.

Sie können eine Auto-Scaling-Gruppe von EC2-Instances verwenden, die über mehrere Availability Zones bereitgestellt werden, um je nach den Anforderungen Ihres Workloads dynamisch zu skalieren. Auto Scaling funktioniert gut für schrittweise Nutzungsänderungen, die über Minuten bis zu Dutzenden Minuten auftreten. Das Starten neuer EC2-Instances dauert jedoch Zeit, insbesondere wenn Ihre Instances Bootstrapping erfordern (z. B. die Installation von Agenten, Anwendungsbinärdateien oder Konfigurationsdateien). Während dieser Zeit könnte Ihre verbleibende Kapazität durch die aktuelle Last überfordert sein. Darüber hinaus basiert die Bereitstellung neuer Instances durch Auto Scaling auf der EC2-Steuerebene. Dies stellt einen Kompromiss dar: Um beim Verlust einer einzelnen Availability Zone statisch stabil zu sein, müssen Sie genügend EC2-Instances in den anderen Availability Zones bereitstellen, um die Last zu bewältigen, die von der beeinträchtigten Availability Zone weg verschoben wurde, anstatt sich auf Auto Scaling zu verlassen, um neue Instances bereitzustellen. Für die Vorabbereitstellung zusätzlicher Kapazität können jedoch zusätzliche Kosten anfallen.

Nehmen wir beispielsweise an, dass Ihr Workload sechs Instances benötigt, um den Kundenverkehr über drei Availability Zones zu bedienen. Um bei einem Ausfall einer einzelnen Availability Zone statisch stabil zu sein, würden Sie drei Instances in jeder Availability Zone bereitstellen, also insgesamt neun. Wenn eine einzelne Availability-Zone-Shift von Instances ausgefallen wäre, hätten Sie immer noch sechs links und können Ihren Kundenverkehr weiterhin bedienen, ohne während des Ausfalls neue Instances bereitstellen und konfigurieren zu müssen. Das Erreichen der statischen Stabilität für Ihre EC2-Kapazität ist mit zusätzlichen Kosten verbunden, da Sie in diesem Fall 50 % zusätzliche Instances ausführen. Nicht für alle Services, für die Sie Ressourcen vorab bereitstellen können, fallen zusätzliche Kosten an, z. B. die Vorabbereitstellung eines S3-Buckets oder eines Benutzers. Sie müssen alle Nachteile der Implementierung statischer Stabilität gegen das Risiko abwägen, die gewünschte Wiederherstellungszeit für Ihren Workload zu überschreiten.

AWS Local Zones und Outposts bringen die Datenebene AWS ausgewählter Services näher an Endbenutzer. Die Steuerebenen für diese Services befinden sich in der übergeordneten Region. Ihre Local Zone- oder Outposts-Instance hat Abhängigkeiten auf Steuerebene für zonale Services

wie EC2 und EBS in der Availability Zone, in der Sie Ihr Local Zone- oder Outposts-Subnetz erstellt haben. Sie werden auch Abhängigkeiten von regionalen Steuerebenen für regionale Services wie Elastic Load Balancing (ELB), Sicherheitsgruppen und der von Amazon Elastic Kubernetes Service ([Amazon EKS](#)) verwalteten Kubernetes-Steuerebene haben (wenn Sie EKS verwenden). Weitere Informationen zu Outposts finden Sie in der [Dokumentation](#) und den häufig gestellten [Fragen zu Support und Wartung](#). Implementieren Sie statische Stabilität, wenn Sie Local Zones oder Outposts verwenden, um die Ausfallsicherheit bei Beeinträchtigungen der Steuerebene oder Unterbrechungen der Netzwerkkonnektivität zur übergeordneten Region zu verbessern.

Regionale Services

Regionale Services sind Services, die auf mehreren Availability Zones AWS basieren, sodass Kunden nicht herausfinden müssen, wie sie Zonenservices optimal nutzen können. Wir gruppieren den Service, der über mehrere Availability Zones bereitgestellt wird, logisch, um Kunden einen einzelnen regionalen Endpunkt zu präsentieren. Amazon SQS und [Amazon DynamoDB](#) sind Beispiele für regionale Services. Sie nutzen die Nähe und Redundanz von Availability Zones, um Ausfälle der Infrastruktur als Kategorie von Verfügbarkeits- und Haltbarkeitsrisiken zu minimieren. Amazon S3 verteilt beispielsweise Anfragen und Daten auf mehrere Availability Zones und ist so konzipiert, dass sie sich automatisch nach dem Ausfall einer Availability Zone wiederherstellen. Sie interagieren jedoch nur mit dem regionalen Endpunkt des Services.

AWS ist der Meinung, dass die meisten Kunden ihre Resilienzziele in einer einzelnen Region erreichen können, indem sie regionale -Services oder Multi-AZ-Architekturen verwenden, die auf zonalen Services basieren. Einige Workloads erfordern jedoch möglicherweise zusätzliche Redundanz, und Sie können die Isolation von verwenden, AWS-Regionen um Multi-Regions-Architekturen für HA- oder Geschäftskontinuitätszwecke zu erstellen. Die physische und logische Trennung zwischen AWS-Regionen vermeidet korrelierte Ausfälle zwischen ihnen. Mit anderen Worten, ähnlich wie Sie ein EC2-Kunde sind und von der Isolation von Availability Zones profitieren könnten, indem Sie sie über mehrere Regionen hinweg bereitstellen, können Sie denselben Vorteil für regionale Services erzielen, indem Sie die Bereitstellung über mehrere Regionen hinweg durchführen. Dazu müssen Sie eine multiregionale Architektur für Ihre Anwendung implementieren, die Ihnen helfen kann, gegenüber der Beeinträchtigung eines regionalen Services ausfallsicher zu sein.

Es kann jedoch schwierig sein, die Vorteile einer Multi-Region-Architektur zu erzielen. Es erfordert sorgfältige Arbeit, um die regionale Isolation zu nutzen und gleichzeitig nichts auf Anwendungsebene rückgängig zu machen. Wenn Sie beispielsweise ein Failover für eine Anwendung zwischen

Regionen durchführen, müssen Sie die strikte Trennung zwischen Ihren Anwendungs-Stacks in jeder Region aufrechterhalten, alle Anwendungsabhängigkeiten kennen und alle Teile der Anwendung zusammen ausfallen. Um dies mit einer komplexen, Microservices-basierten Architektur zu erreichen, die viele Abhängigkeiten zwischen Anwendungen aufweist, müssen viele Engineering- und Geschäftsteams geplant und koordiniert werden. Wenn Sie einzelnen Workloads erlauben, ihre eigenen Failover-Entscheidungen zu treffen, ist die Koordination weniger komplex, führt jedoch ein modales Verhalten ein, da die Latenz zwischen Regionen im Vergleich zu innerhalb einer einzelnen Region erheblich anders ist.

AWS bietet derzeit kein synchrones regionsübergreifendes Replikationsfeature. Wenn Sie einen asynchron replizierten Datenspeicher (bereitgestellt von AWS) über -Regionen hinweg verwenden, besteht die Möglichkeit eines Datenverlusts oder einer Inkonsistenz, wenn Sie ein Failover Ihrer Anwendung zwischen Regionen durchführen. Um mögliche Inkonsistenzen zu minimieren, benötigen Sie einen zuverlässigen Datenabgleichsprozess, dem Sie vertrauen und den Sie möglicherweise für mehrere Datenspeicher in Ihrem gesamten Workload-Portfolio ausführen müssen, oder Sie müssen bereit sein, Datenverlust zu akzeptieren. Schließlich müssen Sie das Failover üben, um zu wissen, dass es funktioniert, wenn Sie es benötigen. Das regelmäßige Rotieren Ihrer Anwendung zwischen Regionen, um ein Failover zu üben, ist eine erhebliche Zeit- und Ressourceninvestitionen. Wenn Sie sich dafür entscheiden, einen synchron replizierten Datenspeicher über Regionen hinweg zu verwenden, um Ihre Anwendungen zu unterstützen, die von mehr als einer Region gleichzeitig ausgeführt werden, unterscheiden sich die Leistungsmerkmale und Latenz einer solchen Datenbank, die sich über 100 oder 1000 Kilometer erstreckt, stark von einer Datenbank, die in einer einzigen Region ausgeführt wird. Dazu müssen Sie Ihren Anwendungs-Stack von Grund auf so planen, dass er dieses Verhalten berücksichtigt. Außerdem wird die Verfügbarkeit beider Regionen zu einer festen Abhängigkeit, was zu einer verringerten Ausfallsicherheit Ihrer Workload führen kann.

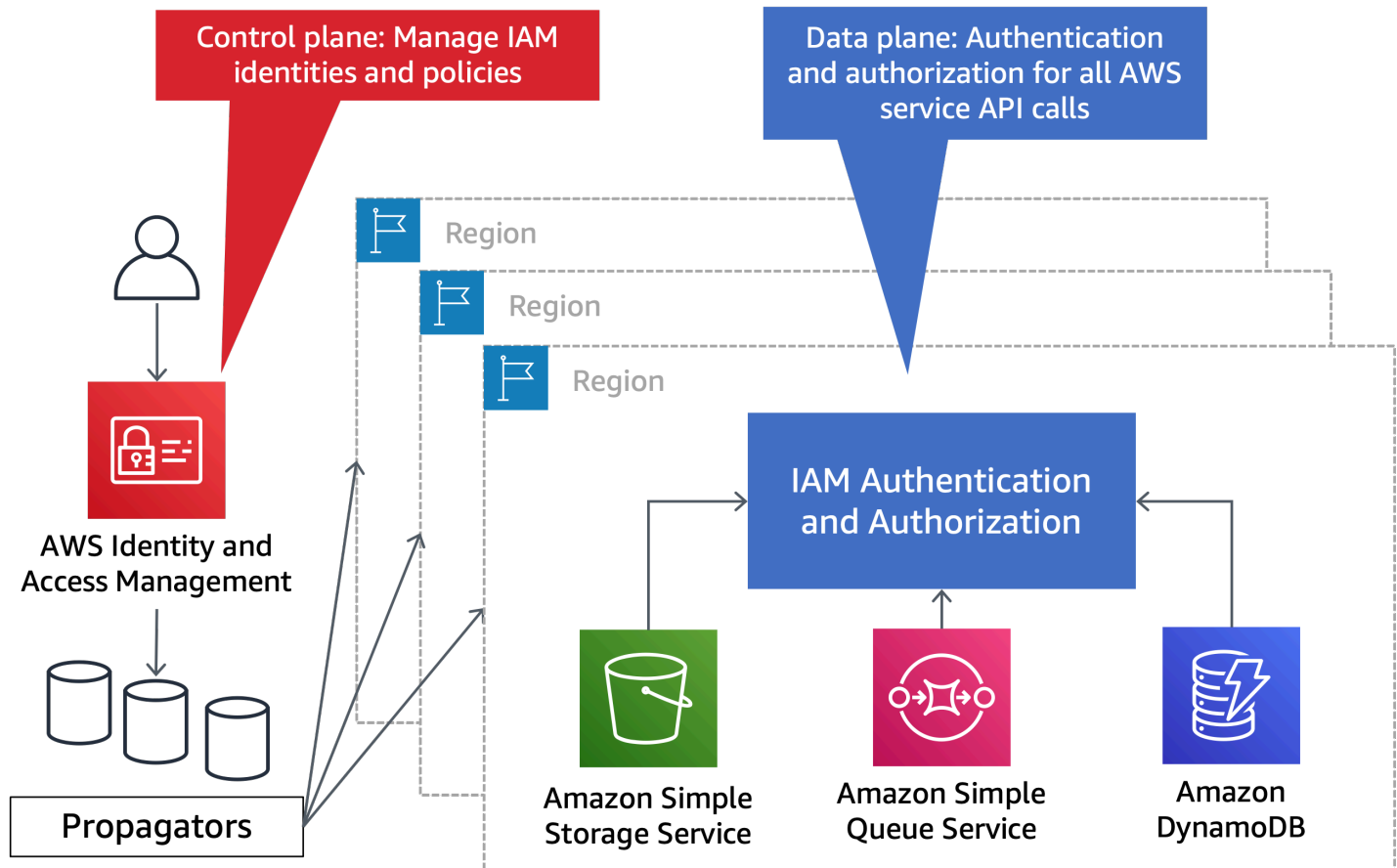
Globale Services

Zusätzlich zu regionalen und zonalen AWS Services gibt es eine kleine Gruppe von AWS Services, deren Steuerebenen und Datenebenen nicht in jeder Region unabhängig voneinander existieren. Da ihre Ressourcen nicht regionsspezifisch sind, werden sie allgemein als globale bezeichnet. Globale AWS Services folgen weiterhin dem herkömmlichen AWS Designmuster, bei dem die Steuerebene und die Datenebene getrennt werden, um eine statische Stabilität zu erreichen. Der wesentliche Unterschied für die meisten globalen Services besteht darin, dass ihre Steuerebene in einer einzigen AWS-Region gehostet wird, während ihre Datenebene global verteilt ist. Es gibt drei verschiedene Arten von globalen Services und eine Reihe von Services, die auf der Grundlage Ihrer ausgewählten Konfiguration global zu sein scheinen.

In den folgenden Abschnitten werden die einzelnen Arten von globalen Services sowie deren Trennung von Steuerebenen und Datenebenen beschrieben. Sie können diese Informationen verwenden, um zu steuern, wie Sie zuverlässige Mechanismen für Hochverfügbarkeit (HA) und Notfallwiederherstellung (DR) erstellen, ohne von einer globalen Service-Steuerebene abhängen zu müssen. Dieser Ansatz hilft dabei, einzelne Fehlerpunkte in Ihrer Architektur zu entfernen und potenzielle regionsübergreifende Auswirkungen zu vermeiden, auch wenn Sie in einer Region arbeiten, die sich von der unterscheidet, in der die globale Service-Steuerebene gehostet wird. Es hilft Ihnen auch, Failover-Mechanismen sicher zu implementieren, die nicht auf globalen Service-Kontrollebenen basieren.

Globale Services, die nach Partition eindeutig sind

In jeder Partition gibt es einige globale AWS Services (in diesem Dokument als Partitionsservices bezeichnet). Partitionale Services stellen ihre Steuerebene in einem einzigen bereit AWS-Region. Einige partitionale Services, wie AWS Network Manager, sind nur auf der Steuerebene verfügbar und orchestrieren die Datenebene anderer Services. Andere partitionale Services wie IAM haben ihre eigene Datenebene, die isoliert und über alle AWS-Regionen in der Partition verteilt ist. Fehler in einem Partitionsservice wirken sich nicht auf andere Partitionen aus. In der `-awsPartition` befindet sich die Steuerebene des IAM-Service in der `-us-east-1Region` mit isolierten Datenebenen in jeder Region der Partition. Partitionale Services haben auch unabhängige Steuerebenen und Datenebenen in den `aws-cn` Partitionen `aws-us-gov` und `aws-iso`. Die Trennung von Steuerebene und Datenebene für IAM ist im folgenden Diagramm dargestellt.



IAM verfügt über eine einzelne Steuerebene und eine regionale Datenebene

Im Folgenden sind Partitionsdienste und deren Speicherort auf der Steuerebene in der aws Partition aufgeführt:

- AWS IAM (us-east-1)
- AWS Organizations (us-east-1)
- AWS Kontoverwaltung (us-east-1)
- Route 53 Application Recovery Controller (ARC) (us-west-2) – Dieser Service ist nur in der aws Partition vorhanden
- AWS Network Manager (us-west-2)
- Route 53 Private DNS (us-east-1)

Wenn eine dieser Service-Steuerebenen ein verfügbarkeitsbeeinträchtigungseignis aufweist, können Sie möglicherweise die von diesen Services bereitgestellten CRUDL-Operationen nicht verwenden. Wenn Ihre Wiederherstellungsstrategie von diesen Vorgängen abhängig ist, verringert

eine Auswirkung auf die Verfügbarkeit der Steuerebene oder der Region, die die Steuerebene hostet, Ihre Wahrscheinlichkeit einer erfolgreichen Wiederherstellung. [Anhang A — Anleitung zum partitionellen Service](#) bietet Strategien zum Entfernen von Abhängigkeiten von globalen Service-Steuerebenen während der Wiederherstellung.

Empfehlung

Verlassen Sie sich nicht auf die Steuerebenen von Partitionsdiensten in Ihrem Wiederherstellungspfad. Verlassen Sie sich stattdessen auf die Operationen dieser Services auf Datenebene. [Anhang A — Anleitung zum partitionellen Service](#) Weitere Informationen dazu, wie Sie partitionale Services entwerfen sollten, finden Sie unter .

Globale Services im Edge-Netzwerk

Die nächsten globalen AWS Services verfügen über eine Steuerebene in der -awsPartition und hosten ihre Datenebenen in der globalen [Points of Presence](#) (PoP)-Infrastruktur (und möglicherweise AWS-Regionen auch). Auf die in gehosteten Datenebenen PoPs kann von Ressourcen in jeder Partition sowie vom Internet aus zugegriffen werden. Beispielsweise betreibt Route 53 seine Steuerebene in der us-east-1 Region, aber seine Datenebene ist über Hunderte von PoPs global sowie über jedes verteilt AWS-Region (um öffentliches und privates DNS von Route 53 innerhalb der Region zu unterstützen). Route 53-Zustandsprüfungen sind ebenfalls Teil der Datenebene und werden von acht AWS-Regionen in der aws Partition durchgeführt. Clients können DNS mithilfe von öffentlich gehosteten Route-53-Zonen von überall im Internet auflösen, einschließlich anderer Partitionen wie GovCloudsowie aus einer AWS Virtual Private Cloud (VPC). Im Folgenden sind globale Edge-Netzwerksservices und ihr Standort auf der Steuerebene in der aws Partition aufgeführt:

- Route 53 Öffentliches DNS (us-east-1)
- Amazon CloudFront (us-east-1)
- AWS WAF Classic für CloudFront (us-east-1)
- AWS WAF für CloudFront (us-east-1)
- Amazon Certificate Manager (ACM) für CloudFront (us-east-1)
- AWS Global Accelerator (AGA) (us-west-2)
- AWS Shield Advanced (us-east-1)

Wenn Sie AGA-Zustandsprüfungen für EC2-Instances oder Elastic IP-Adressen verwenden, verwenden diese Route 53-Zustandsprüfungen. Das Erstellen oder Aktualisieren von AGA-Zustandsprüfungen würde von der Route-53-Steuerebene in `abhängenus-east-1`. Bei der Ausführung der AGA-Zustandsprüfungen wird die Datenebene der Route 53-Zustandsprüfung verwendet.

Während eines Ausfalls, der sich auf die Region auswirkt, in der die Steuerebenen für diese Services gehostet werden, oder eines Ausfalls, der sich auf die Steuerebene selbst auswirkt, können Sie möglicherweise die von diesen Services bereitgestellten CRUDL-Operationen nicht verwenden. Wenn Sie in Ihrer Wiederherstellungsstrategie Abhängigkeiten von diesen Vorgängen übernommen haben, ist diese Strategie möglicherweise weniger wahrscheinlich erfolgreich als wenn Sie sich nur auf die Datenebene dieser Services verlassen.

Empfehlung

Verlassen Sie sich in Ihrem Wiederherstellungspfad nicht auf die Steuerebene von Edge-Netzwerkdiensten. Verlassen Sie sich stattdessen auf die Operationen dieser Services auf Datenebene. [Anhang B — Globale Servicehinweise für Edge-Netzwerke](#) Weitere Informationen zum Entwerfen globaler Services im Edge-Netzwerk finden Sie unter .

Globale Einzelregionsoperationen

Die letzte Kategorie besteht aus bestimmten Operationen auf Steuerebene innerhalb eines Services, die einen globalen Wirkungsbereich haben, nicht aus ganzen Services wie die vorherigen Kategorien. Während Sie mit zonalen und regionalen Services in der von Ihnen angegebenen Region interagieren, haben bestimmte Operationen eine zugrunde liegende Abhängigkeit von einer einzelnen Region, die sich von der unterscheidet, in der sich die Ressource befindet. Diese unterscheiden sich von Services, die nur in einer einzigen Region bereitgestellt werden. [Anhang C — Dienste für einzelne Regionen](#) Eine Liste dieser Services finden Sie unter .

Während eines Fehlers, der sich auf die zugrunde liegende globale Abhängigkeit auswirkt, können Sie die CRUDL-Aktionen der abhängigen Operationen möglicherweise nicht verwenden. Wenn Sie in Ihrer Wiederherstellungsstrategie Abhängigkeiten von diesen Vorgängen übernommen haben, ist diese Strategie möglicherweise weniger wahrscheinlich erfolgreich als wenn Sie sich nur auf die Datenebene dieser Services verlassen. Sie sollten für Ihre Wiederherstellungsstrategie Abhängigkeiten von diesen Operationen vermeiden.

Im Folgenden finden Sie eine Liste von Services, von denen andere Services möglicherweise Abhängigkeiten annehmen, die einen globalen Umfang haben:


- Route 53

Mehrere - AWS Services erstellen Ressourcen, die einen ressourcenspezifischen DNS-Name(n) bereitstellen. Wenn Sie beispielsweise einen Elastic Load Balancer (ELB) bereitstellen, erstellt der Service öffentliche DNS-Datensätze und Zustandsprüfungen in Route 53 für den ELB. Dies hängt von der Route-53-Steuerebene in `abus-east-1`. Andere Services, die Sie verwenden, müssen möglicherweise auch einen ELB bereitstellen, öffentliche Route-53-DNS-Datensätze erstellen oder Route-53-Zustandsprüfungen als Teil ihrer Workflows auf Steuerebene erstellen. Beispielsweise führt die Bereitstellung einer REST-API-Ressource von Amazon API Gateway, einer Amazon Relational Database Service (Amazon RDS)-Datenbank oder einer Amazon-OpenSearch Service-Domain zum Erstellen von DNS-Datensätzen in Route 53. Im Folgenden finden Sie eine Liste von Services, deren Steuerebene von der Route-53-Steuerebene in `abus-east-1` um DNS-Datensätze, gehostete Zonen und/oder Route-53-Zustandsprüfungen zu erstellen, zu aktualisieren oder zu löschen. Diese Liste ist nicht vollständig. Sie soll einige der am häufigsten verwendeten Services hervorheben, deren Aktionen auf Steuerebene zum Erstellen, Aktualisieren oder Löschen von Ressourcen von der Route-53-Steuerebene abhängen:

- Amazon API Gateway REST- und HTTP-APIs
- Amazon-RDS-Instances
- Amazon-Aurora-Datenbanken
- Amazon ELB Load Balancer
- AWS PrivateLink VPC-Endpunkte
- AWS Lambda URLs
- Amazon ElastiCache
- Amazon OpenSearch Service
- Amazon CloudFront
- Amazon MemoryDB für Redis
- Amazon Neptune
- Amazon DynamoDB Accelerator (DAX)
- AGA
- Amazon Elastic Container Service (Amazon ECS) mit DNS-basierter Service Discovery (die die AWS Cloud Map API zur Verwaltung von Route-53-DNS verwendet)

- Amazon-EKS-Kubernetes-Steuerebene

Es ist wichtig zu beachten, dass der VPC-DNS-Service für [EC2-Instance-Hostnamen](#) unabhängig in jedem vorhanden ist AWS-Region und nicht von der Route-53-Steuerebene abhängt. Datensätze, die für EC2-Instances im VPC-DNS-Service AWS erstellt, wie `ip-10-0-10.ec2.internal`, `ip-10-0-1-5.compute.us-west-2.compute.internal`, `i-0123456789abcdef.ec2.internal`, und `i-0123456789abcdef.us-west-2.compute.internal`, verlassen sich nicht auf die Route-53-Steuerebene in `us-east-1`.

 Empfehlung

Verlassen Sie sich nicht auf das Erstellen, Aktualisieren oder Löschen von Ressourcen, die das Erstellen, Aktualisieren oder Löschen von Route-53-Ressourcendatensätzen, gehosteten Zonen oder Zustandsprüfungen in Ihrem Wiederherstellungspfad erfordern. Stellen Sie diese Ressourcen wie ELBs vorab bereit, um eine Abhängigkeit von der Route-53-Steuerebene in Ihrem Wiederherstellungspfad zu verhindern.

- Amazon S3

Die folgenden Operationen der Amazon S3-Steuerebene haben eine zugrunde liegende Abhängigkeit von `us-east-1` in der `-awsPartition`. Ein Fehler, der sich auf Amazon S3 oder andere Services in auswirkt, `us-east-1` kann dazu führen, dass diese Aktionen auf Steuerebene in anderen Regionen beeinträchtigt werden:

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



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

Die Steuerebene für Amazon S3 Multi-Region Access Points (MRAP) wird [nur in und Anforderungen us-west-2](#) zum Erstellen, Aktualisieren oder Löschen von MRAPs gehostet, die diese Region direkt anvisieren. Die Steuerebene für MRAP hat auch zugrunde liegende Abhängigkeiten von AGA in us-west-2, Route 53 in und ACM in jeder Region us-east-1, aus der das MRAP für die Bereitstellung von Inhalten konfiguriert ist. Sie sollten nicht von der Verfügbarkeit der MRAP-Stuerebene in Ihrem Wiederherstellungspfad oder in den Datenebenen Ihrer eigenen Systeme abhängen. Dies unterscheidet sich von [MRAP-Failover-Kontrollen](#), die verwendet werden, um den aktiven oder passiven Routing-Status für jeden Ihrer Buckets im MRAP anzugeben. Diese APIs werden in [fünf AWS-Regionen](#) gehostet und können verwendet werden, um den Datenverkehr mithilfe der Datenebene des Services effektiv zu verschieben.


Darüber hinaus sind Amazon S3-Bucket-Namen global eindeutig und alle Aufrufe der DeleteBucket APIs CreateBucket und hängen von us-east-1 in der -awsPartition ab, um die Eindeutigkeit des Namens zu gewährleisten, obwohl der API-Aufruf an die spezifische Region gerichtet ist, in der Sie den Bucket erstellen möchten. <https://docs.aws.amazon.com/AmazonS3/latest/userguide/UsingBucket.html> Wenn Sie kritische Workflows zur Bucket-Erstellung haben, sollten Sie nicht von der Verfügbarkeit einer bestimmten Schreibweise eines Bucket-Namens abhängig sein, insbesondere von denen, die einem erkennbaren Muster folgen.

Empfehlung

Verlassen Sie sich nicht auf das Löschen oder Erstellen neuer S3-Buckets oder das Aktualisieren von S3-Bucket-Konfigurationen als Teil Ihres Wiederherstellungspfads. Stellen Sie alle erforderlichen S3-Buckets mit den erforderlichen Konfigurationen bereit, damit Sie keine Änderungen vornehmen müssen, um nach einem Fehler eine Wiederherstellung durchzuführen. Dieser Ansatz gilt auch für MRAPs.

- CloudFront

Amazon API Gateway bietet [Edge-optimierte API-Endpunkte](#). Das Erstellen dieser Endpunkte hängt von der CloudFront Steuerebene in `abus-east-1`, um die Verteilung vor dem Gateway-Endpunkt zu erstellen.

 **Empfehlung**

Verlassen Sie sich nicht darauf, neue Edge-optimierte API Gateway-Endpunkte als Teil Ihres Wiederherstellungspfads zu erstellen. Stellen Sie alle erforderlichen API Gateway-Endpunkte vorab bereit.


Alle in diesem Abschnitt beschriebenen Abhängigkeiten sind Aktionen auf Steuerebene, nicht Aktionen auf Datenebene. Wenn Ihre Workloads so konfiguriert sind, dass sie statisch stabil sind, sollten sich diese Abhängigkeiten nicht auf Ihren Wiederherstellungspfad auswirken. Beachten Sie dabei, dass die statische Stabilität zusätzliche Arbeit oder Services für die Implementierung erfordert.

Services, die globale Standardendpunkte verwenden

In einigen Fällen stellen AWS Services einen globalen Standardendpunkt bereit, z. B. AWS Security Token Service ([AWS STS](#)). Andere -Services verwenden diesen globalen Standardendpunkt in ihrer Standardkonfiguration. Das bedeutet, dass ein regionaler Service, den Sie verwenden, eine globale Abhängigkeit von einem einzelnen haben könnte AWS-Region. In den folgenden Details wird erläutert, wie Sie unbeabsichtigte Abhängigkeiten von globalen Standardendpunkten entfernen, die Ihnen helfen, den Service regional zu verwenden.

AWS STS: STS ist ein Webservice, mit dem Sie temporäre Anmeldeinformationen mit eingeschränkten Berechtigungen für IAM-Benutzer oder für Benutzer anfordern können, die Sie authentifizieren (Verbundbenutzer). Die STS-Nutzung aus dem AWS Software Development Kit (SDK) und der Befehlszeilenschnittstelle (CLI) ist standardmäßig `us-east-1`. Der STS-Service bietet auch regionale Endpunkte. Diese Endpunkte sind in Regionen, die ebenfalls standardmäßig aktiviert sind, standardmäßig aktiviert. Sie können diese jederzeit nutzen, indem Sie Ihr SDK oder Ihre CLI gemäß den folgenden Anweisungen konfigurieren: [AWS Regionalisierte STS-Endpunkte](#). Die Verwendung von SigV4A [erfordert auch temporäre Anmeldeinformationen, die von einem](#)


[regionalen STS-Endpunkt angefordert](#) werden. Sie können den globalen STS-Endpunkt für diesen Vorgang nicht verwenden.

 Empfehlung

Aktualisieren Sie Ihre SDK- und CLI-Konfiguration, um die regionalen STS-Endpunkte zu verwenden.

Security Assertion Markup Language (SAML)-Anmeldung: SAML-Services sind in allen vorhanden AWS-Regionen. Um diesen Service zu verwenden, wählen Sie den entsprechenden regionalen SAML-Endpunkt aus, z. B. <https://us-west-2.signin.aws.amazon.com/saml>. Sie müssen Aktualisierungen an Konfigurationen in Ihren Vertrauensrichtlinien und Identitätsanbietern (IdP) vornehmen, um die regionalen Endpunkte verwenden zu können. Spezifische Details finden Sie in der [AWS SAML-Dokumentation](#).

Wenn Sie einen IdP verwenden, der auch auf gehostet wird AWS, besteht die Gefahr, dass er auch während eines AWS Ausfallereignisses betroffen ist. Dies kann dazu führen, dass Sie Ihre IdP-Konfiguration nicht aktualisieren können oder dass Sie möglicherweise keinen vollständigen Verbund herstellen können. Sie sollten „Break-Glass“-Benutzer vorab bereitstellen, falls Ihr IdP beeinträchtigt oder nicht verfügbar ist. [Anhang A — Anleitung zum partitionellen Service](#) Einzelheiten zum Erstellen von Break-Glass-Benutzern auf statisch stabile Weise finden Sie unter .

 Empfehlung

Aktualisieren Sie Ihre Vertrauensrichtlinien für IAM-Rollen, um SAML-Anmeldungen aus mehreren Regionen zu akzeptieren. Aktualisieren Sie während eines Fehlers Ihre IdP-Konfiguration, um einen anderen regionalen SAML-Endpunkt zu verwenden, wenn Ihr bevorzugter Endpunkt beeinträchtigt ist. Erstellen Sie einen/eine Break-Glass-Benutzer(e), falls Ihr IdP beeinträchtigt ist oder nicht verfügbar ist.

AWS IAM Identity Center: Identity Center ist ein cloudbasierter Service, der die zentrale Verwaltung des Single-Sign-On-Zugriffs auf die AWS-Konten und Cloud-Anwendungen eines Kunden vereinfacht. Identity Center muss in einer einzigen Region Ihrer Wahl bereitgestellt werden. Das Standardverhalten für den Service besteht jedoch darin, den globalen SAML-Endpunkt (<https://signin.aws.amazon.com/saml>) zu verwenden, der in gehostet wird us-east-1. Wenn Sie Identity Center in einem anderen bereitgestellt haben AWS-Region, sollten Sie die [Relaystate](#)-URL jedes

Berechtigungssatzes aktualisieren, um denselben regionalen Konsolenendpunkt wie Ihre Identity-Center-Bereitstellung zu erreichen. Wenn Sie beispielsweise Identity Center in bereitgestellt habenus-west-2, sollten Sie den Relaystatus Ihrer Berechtigungssätze aktualisieren, um <https://us-west-2.console.aws.amazon.com> zu verwenden. Dadurch werden alle Abhängigkeiten von us-east-1 aus Ihrer Identity-Center-Bereitstellung entfernt.

Da IAM Identity Center nur in einer einzigen Region bereitgestellt werden kann, sollten Sie außerdem „Break-Glass“-Benutzer bereitstellen, falls Ihre Bereitstellung beeinträchtigt ist. Einzelheiten [Anhang A — Anleitung zum partitionellen Service](#) zum Erstellen von Break-Glass-Benutzern auf statisch stabile Weise finden Sie unter .

Empfehlung

Legen Sie die Relaystate-URL Ihrer Berechtigungssätze in IAM Identity Center so fest, dass sie mit der Region übereinstimmt, in der Sie den Service bereitgestellt haben. Erstellen Sie einen/eine Break-Glass-Benutzer(e), falls Ihre IAM-Identity-Center-Bereitstellung nicht verfügbar ist.

Amazon S3 Storage Lens: Storage Lens bietet ein Standard-Dashboard namens default-account-dashboard. Die Dashboard-Konfiguration und die zugehörigen Metriken werden in gespeichertus-east-1. Sie können zusätzliche Dashboards in anderen Regionen erstellen, indem Sie die [Heimatregion](#) für die Dashboard-Konfiguration und Metrikdaten angeben.

Empfehlung

Wenn Sie während eines Fehlers, der sich auf den Service in auswirkt, Daten aus dem Standard-Dashboard von S3 Storage Lens benötigenus-east-1, erstellen Sie ein zusätzliches Dashboard in einer alternativen Heimatregion. Sie können auch alle anderen benutzerdefinierten Dashboards duplizieren, die Sie in zusätzlichen Regionen erstellt haben.

Globale Serviceübersicht

Die Datenebenen für globale Services wenden ähnliche Isolations- und Isolationsprinzipien an wie regionale AWS Services. Ein Fehler, der sich auf die Datenebene von IAM in einer Region auswirkt, hat keinen Einfluss auf den Betrieb der IAM-Datenebene in einem anderen AWS-Region. Ebenso wirkt sich ein Fehler, der sich auf die Datenebene von Route 53 in einem PoP

auswirkt, nicht auf den Betrieb der Route-53-Datenebene im Rest von aus PoPs. Daher müssen wir berücksichtigen, dass es sich um Ereignisse zur Serviceverfügbarkeit handelt, die sich auf die Region auswirken, in der die Steuerebene ausgeführt wird, oder auf die Steuerebene selbst. Da für jeden globalen Service nur eine einzige Steuerebene vorhanden ist, kann ein Fehler, der sich auf diese Steuerebene auswirkt, regionsübergreifende Auswirkungen auf CRUDL-Operationen haben (d. h. die Konfigurationsoperationen, die normalerweise verwendet werden, um einen Service im Gegensatz zur direkten Verwendung des Services einzurichten oder zu konfigurieren).

Die effektivste Möglichkeit, Workloads für die ausfallsichere Nutzung globaler Services zu entwerfen, ist die Verwendung statischer Stabilität. Entwerfen Sie Ihren Workload während eines Ausfallszenarios, um keine Änderungen an einer Steuerebene vornehmen zu müssen, um die Auswirkungen oder das Failover an einem anderen Ort zu minimieren. Unter [Anhang A — Anleitung zum partitionellen Service](#) und [Anhang B — Globale Servicehinweise für Edge-Netzwerke](#) finden Sie eine ausführliche Anleitung zur Verwendung dieser Arten von globalen Services, um Abhängigkeiten auf Steuerebene zu entfernen und einzelne Fehlerpunkte zu eliminieren. Wenn Sie die Daten aus einer Operation auf Steuerebene für die Wiederherstellung benötigen, speichern Sie diese Daten in einem Datenspeicher zwischen, auf den über die Datenebene zugegriffen werden kann, z. B. einen [AWS Systems Manager](#) Parameter Store (SSM Parameter Store)-Parameter, eine DynamoDB-Tabelle oder einen S3-Bucket. Aus Redundanzgründen können Sie diese Daten auch in einer zusätzlichen Region speichern. Wenn Sie beispielsweise die [bewährten Methoden](#) für Route 53 Application Recovery Controller (ARC) befolgen, sollten Sie Ihre fünf regionalen Cluster-Endpunkte fest codieren oder mit einem Lesezeichen versehen. Während eines Ausfallereignisses können Sie möglicherweise nicht auf einige API-Operationen zugreifen, einschließlich Route-53-ARC-API-Operationen, die nicht auf dem extrem zuverlässigen Datenebenen-Cluster gehostet werden. Sie können die Endpunkte für Ihre Route-53-ARC-Cluster mithilfe der `DescribeCluster`-API-Operation auflisten.

Im Folgenden finden Sie eine Zusammenfassung einiger der häufigsten Fehlkonfigurationen oder Anti-Muster, die zu Abhängigkeiten von den Steuerebenen globaler Services führen:

- Änderungen an Route-53-Datensätzen vornehmen, z. B. den Wert eines Datensatzes aktualisieren oder die Gewichtungen eines gewichteten Datensatzes ändern, um ein Failover durchzuführen.
- Erstellen oder Aktualisieren von IAM-Ressourcen, einschließlich IAM-Rollen und -Richtlinien, während eines Failovers. Dies ist in der Regel nicht beabsichtigt, kann aber auf einen nicht getesteten Failover-Plan zurückzuführen sein.
- Verlassen Sie sich darauf, dass IAM Identity Center für Operatoren den Zugriff auf Produktionsumgebungen während eines Ausfallereignisses erhält.

- Verlassen Sie sich auf die Standardkonfiguration von IAM Identity Center, um die Konsole in zu verwendenus-east-1, wenn Sie Identity Center in einer anderen Region bereitgestellt haben.
- Nehmen Sie Änderungen an den AGA-Datenverkehrs-Wärlgewichtungen vor, um ein regionales Failover manuell durchzuführen.
- Aktualisieren der Ursprungskonfiguration einer CloudFront Verteilung, um von einem beeinträchtigten Ursprung wegzufallen.
- Bereitstellung von Notfallwiederherstellungs-Ressourcen (DR), wie ELBs und RDS-Instances während eines Ausfallereignisses, die von der Erstellung von DNS-Datensätzen in Route 53 abhängen.

Im Folgenden finden Sie eine Zusammenfassung der Empfehlungen in diesem Abschnitt für die ausfallsichere Verwendung globaler -Services, die dazu beitragen würden, die vorherigen häufigen Anti-Muster zu verhindern.

Zusammenfassung der Empfehlungen

Verlassen Sie sich nicht auf die Steuerebenen von Partitionsdiensten in Ihrem Wiederherstellungspfad. Verlassen Sie sich stattdessen auf die Operationen dieser Services auf Datenebene. [Anhang A — Anleitung zum partitionellen Service](#) Weitere Informationen dazu, wie Sie partitionale Services entwerfen sollten, finden Sie unter .

Verlassen Sie sich in Ihrem Wiederherstellungspfad nicht auf die Steuerebene von Edge-Netzwerkdiensten. Verlassen Sie sich stattdessen auf die Operationen dieser Services auf Datenebene. [Anhang B — Globale Servicehinweise für Edge-Netzwerke](#) Weitere Informationen zum Entwerfen globaler Services im Edge-Netzwerk finden Sie unter .

Verlassen Sie sich nicht auf das Erstellen, Aktualisieren oder Löschen von Ressourcen, die das Erstellen, Aktualisieren oder Löschen von Route-53-Ressourcendatensätzen, gehosteten Zonen oder Zustandsprüfungen in Ihrem Wiederherstellungspfad erfordern. Stellen Sie diese Ressourcen wie ELBs vorab bereit, um eine Abhängigkeit von der Route-53-Steuerebene in Ihrem Wiederherstellungspfad zu verhindern.

Verlassen Sie sich nicht auf das Löschen oder Erstellen neuer S3-Buckets oder das Aktualisieren von S3-Bucket-Konfigurationen als Teil Ihres Wiederherstellungspfads. Stellen Sie alle erforderlichen S3-Buckets mit den erforderlichen Konfigurationen bereit, damit Sie keine Änderungen vornehmen müssen, um nach einem Fehler eine Wiederherstellung durchzuführen. Dieser Ansatz gilt auch für MRAPs.

Verlassen Sie sich nicht darauf, neue Edge-optimierte API Gateway-Endpunkte als Teil Ihres Wiederherstellungspfads zu erstellen. Stellen Sie vorab alle erforderlichen API Gateway-Endpunkte bereit.

Aktualisieren Sie Ihre SDK- und CLI-Konfiguration, um die regionalen STS-Endpunkte zu verwenden.

Aktualisieren Sie Ihre Vertrauensrichtlinien für IAM-Rollen, um SAML-Anmeldungen aus mehreren Regionen zu akzeptieren. Aktualisieren Sie während eines Fehlers Ihre IdP-Konfiguration, um einen anderen regionalen SAML-Endpunkt zu verwenden, wenn Ihr bevorzugter Endpunkt beeinträchtigt ist. Erstellen Sie Break-Glass-Benutzer, falls Ihr IdP beeinträchtigt oder nicht verfügbar ist.

Legen Sie die Relaystate-URL Ihrer Berechtigungssätze in IAM Identity Center so fest, dass sie mit der Region übereinstimmt, in der Sie den Service bereitgestellt haben. Erstellen Sie einen/eine Break-Glass-Benutzer(e), falls Ihre Identity-Center-Bereitstellung nicht verfügbar ist/sind.

Wenn Sie während eines Fehlers, der sich auf den Service in auswirkt, Daten aus dem Standard-Dashboard von S3 Storage Lens benötigen `us-east-1`, erstellen Sie ein zusätzliches Dashboard in einer alternativen Heimatregion. Sie können auch alle anderen benutzerdefinierten Dashboards duplizieren, die Sie in zusätzlichen Regionen erstellt haben.

Schlussfolgerung

AWS bietet mehrere verschiedene Konstrukte für Fehlerisolationsgrenzen. Sie sollten darüber nachdenken, wie Sie zonale, regionale und globale Dienste einrichten und welche potenziellen Auswirkungen dies auf Ihre Arbeitslast und die Fähigkeit Ihres Workloads hat, sich bei Beeinträchtigungen der Steuerungsebene zu erholen. Statische Stabilität ist eine der wichtigsten Möglichkeiten, Abhängigkeiten auf der Steuerungsebene zu vermeiden und zuverlässige und belastbare HA- und DR-Mechanismen zu schaffen, wenn Sie AWS Dienste verwenden.

Anhang A — Anleitung zum partitionellen Service

Für partitionelle Dienste sollten Sie statische Stabilität implementieren, um die Resilienz Ihrer Arbeitslast auch bei einer Beeinträchtigung der AWS Dienststeuerungsebene aufrechtzuerhalten. Im Folgenden finden Sie eine Anleitung zur Berücksichtigung von Abhängigkeiten von partitionellen Diensten sowie dazu, was bei einer Beeinträchtigung der Steuerungsebene funktioniert und was nicht.

AWS Identity and Access Management (IAM)

Die AWS Identity and Access Management (IAM-) Steuerungsebene besteht aus allen öffentlichen IAM-APIs (einschließlich Access Advisor, aber nicht Access Analyzer oder IAM Roles Anywhere). Dazu gehören Aktionen wie `CreateRoleAttachRolePolicy`, `ChangePassword`, `UpdateSAMLProvider`, und `UpdateLoginProfile`. Die IAM-Datenebene ermöglicht die Authentifizierung und Autorisierung für jeweils IAM-Prinzipale. AWS-Region Während einer Beeinträchtigung der Steuerungsebene funktionieren CRUDL-Operationen für IAM möglicherweise nicht, aber die Authentifizierung und Autorisierung für bestehende Principals funktionieren weiterhin. STS ist ein reiner Dienst auf Datenebene, der von IAM getrennt ist und nicht von der IAM-Steuerungsebene abhängig ist.

Dies bedeutet, dass Sie sich bei der Planung von Abhängigkeiten von IAM in Ihrem Wiederherstellungspfad nicht auf die IAM-Steuerungsebene verlassen sollten. Ein statisch stabiles Design für einen „Bruchglas“-Administratorbenutzer würde beispielsweise darin bestehen, einen Benutzer mit den entsprechenden Berechtigungen zu erstellen, das Passwort festzulegen und den Zugriffsschlüssel und den geheimen Zugriffsschlüssel bereitzustellen und diese Anmeldeinformationen dann in einem physischen oder virtuellen Tresor zu sperren. Rufen Sie bei Bedarf in einem Notfall die Benutzeranmeldeinformationen aus dem Tresor ab und verwenden Sie sie nach Bedarf. Ein non-statically-stable Design wäre, den Benutzer während eines Fehlers bereitzustellen oder dass der Benutzer eine Vorbereitstellung vornimmt, die Administratorrichtlinie jedoch nur angehängt wird, wenn dies erforderlich ist. Diese Ansätze würden von der IAM-Steuerungsebene abhängen.

AWS Organizations

Die AWS Organizations Steuerungsebene besteht aus allen APIs öffentlicher Organizations wie `AcceptHandshake`, `AttachPolicyCreateAccount`, `CreatePolicy`, und `ListAccounts`. Es gibt keine Datenebene für AWS Organizations. Es orchestriert die Datenebene für andere Dienste

wie IAM. Während einer Beeinträchtigung der Kontrollebene funktionieren CRUDL-Operationen für Organizations möglicherweise nicht, aber die Richtlinien, wie Service Control Policies (SCP) und Tag Policies, funktionieren weiterhin und werden im Rahmen des IAM-Autorisierungsprozesses bewertet. Delegierte Administratorfunktionen und Funktionen für mehrere Konten in anderen AWS Diensten, die von Organizations unterstützt werden, werden ebenfalls weiterhin funktionieren.

Dies bedeutet, dass Sie sich bei der Planung von Abhängigkeiten auf AWS Organizations Ihrem Wiederherstellungspfad nicht auf die Steuerungsebene der Organizations verlassen sollten. Implementieren Sie stattdessen statische Stabilität in Ihrem Wiederherstellungsplan. Ein non-statically-stable Ansatz könnte beispielsweise darin bestehen, SCPs zu aktualisieren, um die Beschränkungen für die `aws:RequestedRegion` Bedingung „Zulässige AWS-Regionen Via the Condition“ aufzuheben, oder um Administratorberechtigungen für bestimmte IAM-Rollen zu aktivieren. Dies hängt davon ab, dass die Kontrollebene der Organizations diese Aktualisierungen vornimmt. Ein besserer Ansatz wäre die Verwendung von [Sitzungs-Tags](#), um die Verwendung von Administratorberechtigungen zu gewähren. Ihr Identity Provider (IdP) kann Sitzungs-Tags enthalten, die anhand der `aws:PrincipalTag` Bedingung ausgewertet werden können. Auf diese Weise können Sie Berechtigungen für bestimmte Prinzipale dynamisch konfigurieren und gleichzeitig Ihre SCPs dabei unterstützen, statisch zu bleiben. Dadurch werden Abhängigkeiten von Steuerungsebenen entfernt und nur Aktionen auf der Datenebene verwendet.

AWS-Kontenverwaltung

Die AWS Account-Management-Kontrollebene wird in us-east-1 gehostet und besteht aus allen [öffentlichen APIs](#) für die Verwaltung einer AWS-Konto, wie `GetContactInformation` z. B. und `PutContactInformation`. Dazu gehört auch das Erstellen oder Schließen eines neuen AWS-Konto über die Verwaltungskonsole. Die APIs für `CloseAccount`, `CreateAccount`, `CreateGovCloudAccount`, und `DescribeAccount` sind Teil der AWS Organizations Steuerungsebene, die auch in us-east-1 gehostet wird. Darüber hinaus AWS Organizations hängt die [Erstellung eines GovCloud Kontos außerhalb von der AWS-Konto Management-Kontrollebene in us-east-1](#) ab. Außerdem [müssen GovCloud Konten 1:1 mit einem AWS-Konto in der aws Partition verknüpft sein](#). Zum Erstellen von -Konten in der `aws-cn` Partition ist us-east-1 nicht abhängig. Die Datenebene für AWS-Konten sind die Konten selbst. Während einer Beeinträchtigung der Steuerungsebene funktionieren Vorgänge vom Typ CRUDL (wie das Erstellen eines neuen Kontos oder das Abrufen und Aktualisieren von Kontaktinformationen) für AWS-Konten möglicherweise nicht. Verweise auf das Konto in den IAM-Richtlinien funktionieren weiterhin.

Das bedeutet, dass Sie sich bei der Planung von Abhängigkeiten von der AWS Kontoverwaltung bei der Wiederherstellung nicht auf die Account-Management-Steuerungsebene verlassen sollten.

Die Account-Management-Steuerungsebene bietet zwar keine direkten Funktionen, die Sie normalerweise in einer Wiederherstellungssituation verwenden würden, es kann jedoch vorkommen, dass Sie dies tun würden. Ein statisch stabiles Design würde beispielsweise darin bestehen, alles, was AWS-Konten Sie für ein Failover benötigen, vorab bereitzustellen. Ein non-statically-stable Design wäre, AWS-Konten im Falle eines Fehlers neu zu erstellen, um Ihre DR-Ressourcen zu hosten.

Route 53 Application Recovery-Controller

Die Steuerungsebene für Route 53 ARC besteht aus den APIs für die Wiederherstellungssteuerung und die Wiederherstellungsbereitschaft, wie sie unter [Amazon Route 53 Application Recovery Controller-Endpoints und Kontingente](#) angegeben sind. Sie verwalten Bereitschaftsprüfungen, Routing-Kontrollen und Clusteroperationen mithilfe der Steuerungsebene. Die Datenebene von ARC ist Ihr Wiederherstellungscluster, der die Routing-Kontrollwerte verwaltet, die bei Route 53-Gesundheitschecks abgefragt werden, und der auch die Sicherheitsregeln implementiert. Auf die [Datenebenenfunktionalität](#) von Route 53 ARC wird über Ihre Recovery-Cluster-APIs wie `https://aaaaaaa.route53-recovery-cluster.eu-west-1.amazonaws.com` zugegriffen.

Das bedeutet, dass Sie sich bei Ihrem Wiederherstellungspfad nicht auf die Route 53 ARC-Steuerebene verlassen sollten. Es gibt zwei [bewährte Verfahren](#), die bei der Umsetzung dieser Leitlinien helfen:

- Markieren Sie zunächst die fünf regionalen Cluster-Endpunkte mit einem Lesezeichen oder schreiben Sie sie fest. Dadurch entfällt die Notwendigkeit, während eines Failover-Szenarios die Operation auf der DescribeCluster Steuerungsebene zu verwenden, um die Endpunktwerte zu ermitteln.
- Verwenden Sie zweitens die Route 53 ARC-Cluster-APIs, indem Sie die CLI oder das SDK verwenden, um Aktualisierungen der Routing-Kontrollen durchzuführen, und nicht die AWS Management Console. Dadurch wird die Managementkonsole als Abhängigkeit von Ihrem Failoverplan entfernt und sichergestellt, dass sie nur von Aktionen auf der Datenebene abhängt.

AWS-Network Manager

Der AWS Network Manager-Dienst ist in erster Linie ein System, das nur für die Steuerungsebene bestimmt ist und in us-west-2 gehostet wird. Damit können Sie die Konfiguration Ihres AWS Cloud WAN-Kernnetzwerk (WAN-Kernnetzwerk) und Ihr AWS -Transit-Gateway-Netzwerk über AWS-Konten -Regionen und lokale Standorte verwalten. Es aggregiert auch Ihre Cloud-WAN-Metriken

in us-west-2, auf die auch über die CloudWatch Datenebene zugegriffen werden kann. Wenn Network Manager beeinträchtigt ist, wird die Datenebene der Dienste, die er orchestriert, nicht beeinträchtigt. Die CloudWatch Metriken für -Cloud-WAN sind auch in us-west-2 verfügbar. Wenn Sie historische Metrikdaten wie ein- und ausgehende Byte pro Region benötigen, um zu verstehen, wie viel Traffic während eines Fehlers, der sich auf US-West-2 auswirkt, oder für andere betriebliche Zwecke in andere Regionen verlagert werden könnte, können Sie diese Metriken als CSV-Daten direkt von der CloudWatch Konsole exportieren oder diese Methode verwenden: [Veröffentlichen Sie CloudWatch Amazon-Metriken in einer CSV-Datei](#). Die Daten befinden sich im AWS/Network Manager Namespace und Sie können dies nach einem von Ihnen ausgewählten Zeitplan ausführen und in S3 oder in einem anderen von Ihnen ausgewählten Datenspeicher speichern. Um einen statisch stabilen Wiederherstellungsplan zu implementieren, sollten Sie den AWS Network Manager nicht verwenden, um Aktualisierungen an Ihrem Netzwerk vorzunehmen, und verlassen Sie sich nicht auf Daten aus den Vorgängen auf der Steuerungsebene als Failover-Eingabe.

Route 53 Privates DNS

Private gehostete Route 53-Zonen werden in jeder Partition unterstützt. Die Überlegungen für privat gehostete Zonen und öffentlich gehostete Zonen in Route 53 sind jedoch dieselben. Weitere Informationen finden Sie unter Amazon Route 53 in [Anhang B — Globale Servicerichtlinien für Edge-Netzwerke](#).

Anhang B — Globale Servicehinweise für Edge-Netzwerke

Für globale Dienste in Edge-Netzwerken sollten Sie statische Stabilität implementieren, um die Resilienz Ihres Workloads während einer Beeinträchtigung der AWS Dienststeuerungsebene aufrechtzuerhalten.

Route 53

Die Route 53-Steuerungsebene besteht aus allen öffentlichen Route 53-APIs, die Funktionen für gehostete Zonen, Datensätze, Integritätsprüfungen, DNS-Abfrageprotokolle, wiederverwendbare Delegierungssätze, Verkehrsrichtlinien und Kostenzuweisungskennzeichnungen abdecken. Es wird in den us-east-1 gehostet. Die Datenebene ist der maßgebliche DNS-Dienst, der über 200 AWS-Region Points-of-Points-of-Points-of-east-Standorte und Daten der Zustandsprüfung beantwortet. Darüber hinaus verfügt Route 53 über eine Datenebene für Gesundheitschecks, bei der es sich ebenfalls um einen global verteilten Dienst handelt, der auf mehrere Dienste verteilt ist. AWS-Regionen Diese Datenebene führt Zustandsprüfungen durch, aggregiert und an die Datenebenen des öffentlichen und privaten DNS der Route 53 und liefert. Während einer Beeinträchtigung der Steuerungsebene funktionieren Crudl-Operationen für Route 53 möglicherweise nicht, aber DNS-Auflösungs- und Integritätsprüfungen sowie Routing-Aktualisierungen, die sich aus Änderungen bei den Integritätsprüfungen ergeben, funktionieren weiterhin.

Dies bedeutet, dass Sie sich bei der Planung von Abhängigkeiten von Route 53 in Ihrem Wiederherstellungspfad nicht auf die Route 53-Steuerebene verlassen sollten. Ein statisch stabiles Design wäre beispielsweise, den Status von Integritätsprüfungen zu verwenden, um Failover zwischen Regionen durchzuführen oder eine Availability Zone zu evakuieren. Sie können die [Routingkontrollen des Route 53 Application Recovery Controller \(ARC\)](#) verwenden, um den Status von Integritätsprüfungen und die Antworten auf DNS-Abfragen manuell zu ändern. Es gibt ähnliche Muster wie das, was ARC bietet, die Sie auf der Grundlage Ihrer Anforderungen implementieren können. Einige dieser Muster werden unter [Creating Disaster Recovery Mechanisms using Route 53](#) und im [Abschnitt Advanced Multi-AZ Resilience Patterns Health Check Circuit Breaker beschrieben](#). Wenn Sie sich für einen DR-Plan mit mehreren Regionen entschieden haben, stellen Sie Ressourcen, für die DNS-Einträge erstellt werden müssen, wie ELBs und RDS-Instances, vorab bereit. Ein non-statically-stable Design wäre, den Wert eines Route 53-Ressourcendatensatzes über die ChangeResourceRecordSets API zu aktualisieren, die Gewichtung eines gewichteten Datensatzes zu ändern oder neue Datensätze zu erstellen, um ein Failover durchzuführen. Diese Ansätze hängen von der Route 53-Steuerebene ab.

Amazon CloudFront

Die CloudFront Amazon-Kontrollebene besteht aus allen öffentlichen CloudFront APIs für die Verwaltung von Distributionen und wird in us-east-1 gehostet. Die Datenebene ist die Verteilung selbst, die PoPs vom In-the-Edge-Netzwerk aus bedient wird. Es übernimmt die Bearbeitung, das Routing und das Caching Ihrer ursprünglichen Inhalte. Während einer Beeinträchtigung der Steuerungsebene funktionieren Crudl-Operationen für CloudFront (einschließlich Invalidierungsanfragen) möglicherweise nicht, aber Ihre Inhalte werden weiterhin zwischengespeichert und bereitgestellt, und die [Origin-Failover](#) funktionieren weiterhin.

Das bedeutet, dass Sie sich bei der Planung von Abhängigkeiten auf CloudFront Ihrem Wiederherstellungspfad nicht auf die CloudFront Steuerungsebene verlassen sollten. Ein statisch stabiles Design wäre beispielsweise die Verwendung automatisierter Origin-Failover, um die Auswirkungen einer Beeinträchtigung eines Ihrer Origins zu mildern. Sie können sich auch dafür entscheiden, Origin Load Balancing oder Failover mit Lambda @Edge zu erstellen. Weitere Informationen zu diesem [Muster finden Sie unter Drei erweiterte Entwurfsmuster für hochverfügbare Anwendungen mit Amazon CloudFront und Verwenden von Amazon CloudFront und Amazon S3 zum Erstellen von aktiv-aktiven Geo-Proximity-Anwendungen für mehrere Regionen](#). Ein non-statically-stable Design wäre, die Konfiguration Ihrer Distribution als Reaktion auf einen Origin-Fehler manuell zu aktualisieren. Dieser Ansatz würde von der CloudFront Steuerungsebene abhängen.

Amazon Certificate Manager

Wenn Sie benutzerdefinierte Zertifikate für Ihre CloudFront Distribution verwenden, sind Sie auch von ACM abhängig. Die Verwendung CloudFront benutzerdefinierter Zertifikate in der Region us-east-1 Während einer Beeinträchtigung der Kontrollebene funktionieren Ihre vorhandenen Zertifikate, die in Ihrer Distribution konfiguriert wurden, ebenso wie automatische Zertifikatserneuerungen. Verlassen Sie sich nicht darauf, die Konfiguration der Distribution zu ändern oder neue Zertifikate als Teil Ihres Wiederherstellungspfads zu erstellen.

AWSWeb Application Firewall (WAF) und WAF Classic

Wenn Sie es AWS WAF mit Ihrer CloudFront Distribution verwenden, sind Sie von der WAF-Kontrollebene abhängig, die ebenfalls in der Region us-east-1 gehostet wird. Bei einer Beeinträchtigung der Steuerungsebene funktionieren die konfigurierten Web Access Control Lists (ACLs) und die zugehörigen Regeln weiterhin. Verlassen Sie sich nicht darauf, Ihre WAF-Web-ACLs als Teil Ihres Wiederherstellungspfads zu aktualisieren.

AWS Global Accelerator

Die AGA-Steuerungsebene besteht aus allen öffentlichen AGA-APIs und wird in us-west-2 gehostet. Die Datenebene ist das Netzwerk-Routing der Anycast-IP-Adressen, die von AGA an Ihre registrierten Endpunkte bereitgestellt werden. AGA verwendet auch Route 53-Gesundheitschecks, um den Zustand Ihrer AGA-Endpunkte zu ermitteln, die Teil der Route 53-Datenebene sind. Während einer Beeinträchtigung der Steuerungsebene funktionieren Operationen vom Typ CRUDL für AGA möglicherweise nicht. Das Routing zu Ihren vorhandenen Endpunkten sowie die bestehenden Integritätsprüfungen, Wähltasten und Konfigurationen zur Gewichtung von Endpunkten, die verwendet werden, um den Datenverkehr an andere Endpunkte und Endpunktgruppen weiterzuleiten oder zu verlagern, werden weiterhin funktionieren.

Dies bedeutet, dass Sie sich bei der Planung von Abhängigkeiten von AGA bei Ihrem Wiederherstellungspfad nicht auf die AGA-Steuerungsebene verlassen sollten. Ein statisch stabiles Design würde beispielsweise darin bestehen, den Status der konfigurierten Integritätsprüfungen zu verwenden, um an fehlerhaften Endpunkten ein Failaway durchzuführen. Beispiele für diese [Konfiguration finden Sie unter Bereitstellen von Anwendungen für mehrere Regionen in der AWS Verwendung von AWS Global Accelerator](#). Ein non-statically-stable Plan wäre, die Prozentsätze für die AGA-Traffic Wählvorgänge zu ändern, Endpunktgruppen zu bearbeiten oder einen Endpunkt aus einer Endpunktgruppe zu entfernen, wenn eine Beeinträchtigung auftritt. Diese Ansätze würden von der AGA-Steuerungsebene abhängen.

Amazon S3 Shield

Die Amazon Shield Advanced-Steuerungsebene besteht aus allen öffentlichen Shield Advanced-APIs und wird in us-east-1 gehostet. Dazu gehören Funktionen wie `CreateProtection`, `CreateProtectionGroup`, `AssociateHealthCheck`, `DescribeDRTAccess`, und `ListProtections`. Die Datenebene ist der von Shield Advanced bereitgestellte DDoS-Schutz sowie die Erstellung von Shield Advanced-Metriken. Shield Advanced verwendet auch Route 53-Gesundheitschecks (die Teil der Route 53-Datenebene sind), sofern Sie sie konfiguriert haben. Während einer Beeinträchtigung der Steuerungsebene funktionieren Crudl-Operationen für Shield Advanced möglicherweise nicht, aber der für Ihre Ressourcen konfigurierte DDoS-Schutz sowie die Reaktionen auf Änderungen bei den Zustandsprüfungen funktionieren weiterhin.

Das bedeutet, dass Sie sich bei Ihrem Wiederherstellungsprozess nicht auf die Shield Advanced-Steuerungsebene verlassen sollten. Die Shield Advanced-Steuerungsebene bietet zwar keine direkten Funktionen, die Sie normalerweise in einer Wiederherstellungssituation verwenden

würden, es kann jedoch vorkommen, dass Sie dies tun würden. Ein statisch stabiles Design würde beispielsweise darin bestehen, dass Ihre DR-Ressourcen bereits so konfiguriert sind, dass sie Teil einer Schutzgruppe sind, und dass ihnen Gesundheitschecks zugeordnet sind, anstatt diesen Schutz nach dem Auftreten des Fehlers zu konfigurieren. Dadurch wird verhindert, dass Sie bei der Wiederherstellung auf die Shield Advanced-Steuerebene angewiesen sind.

Anhang C — Dienste für einzelne Regionen

Im Folgenden finden Sie eine Liste von Diensten oder spezifischen Funktionen in diesem Dienst (die in Klammern hinter dem Dienstenamen aufgeführt sind), die nur in einer einzigen Region verfügbar sind. Dieselben Richtlinien für die Implementierung statischer Stabilität, die für andere globale Dienste gelten, gelten für diese Dienste, wenn Sie Abhängigkeiten von ihren Steuerungsebenen und Datenebenen planen müssen.

- [Alexa for Business](#)
- [AWS Marketplace](#) (AWS Marketplace Katalog-API, AWS Marketplace Commerce Analytics, AWS Marketplace Entitlement Service)
- [Billing and Cost Management](#) (AWS Cost Explorer, AWS Kosten- und Nutzungsberichte, AWS Budgets, Savings Plans)
- [AWS BugBust](#)
- [Amazon Mechanical Turk](#)
- [Amazon Chime](#)
- [Amazon Chime SDK](#) (PSTN-Audio, Messaging, Identität)
- [AWSChatbot](#)
- [AWS DeepRacer](#)
- [AWSDevice Farm](#)
- [Amazon GameSparks](#)
- [Amazon Honeycode](#)

Beitragende Faktoren

Zu den Mitwirkenden an diesem Dokument gehören:

- Michael Haken, Principal Solutions Architect, Amazon Web Services

Dokumentversionen

Um Benachrichtigungen über Aktualisierungen dieser Veröffentlichung zu erhalten, können Sie den RSS-Feed abonnieren.

Änderung	Beschreibung	Datum
Geringfügige Überarbeitung	Aktualisierter Leitfaden, angepasst an die bewährten IAM-Methoden. Weitere Informationen finden Sie unter Bewährte IAM-Methoden .	9. Februar 2023
Erstveröffentlichung	Whitepaper veröffentlicht.	16. November 2022

AWS-Glossar

Die neueste AWS-Terminologie finden Sie im [AWS-Glossar](#) in der AWS-Glossar-Referenz.

Hinweise

Die Kunden sind dafür verantwortlich, die Informationen in diesem Dokument selbst unabhängig zu bewerten. Dieses Dokument: (a) dient nur zu Informationszwecken, (b) stellt aktuelle AWS Produktangebote und Praktiken dar, die ohne vorherige Ankündigung geändert werden können, und (c) enthält keine Verpflichtungen oder Zusicherungen von AWS und seinen verbundenen Unternehmen, Lieferanten oder Lizenzgebern. AWS-Produkte oder Dienstleistungen werden „wie sie sind“ ohne ausdrückliche oder stillschweigende Garantien, Zusicherungen oder Bedingungen jeglicher Art bereitgestellt. Die Verantwortlichkeiten und Verbindlichkeiten AWS gegenüber seinen Kunden werden durch AWS Vereinbarungen geregelt, und dieses Dokument ist nicht Teil einer Vereinbarung zwischen seinen Kunden AWS und seinen Kunden und ändert diese auch nicht.

© 2022 Amazon Web Services, Inc. oder verbundene Unternehmen. Alle Rechte vorbehalten.

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.