AWS Grundlagen für mehrere Regionen

# AWS Präskriptive Leitlinien

# AWS Präskriptive Leitlinien: AWS Grundlagen für mehrere Regionen

# Table of Contents

# AWS Grundlagen für mehrere Regionen

John Formento, Amazon Web Services ()AWS

Dezember 2024 (Geschichte der Dokumente)

Dieser 300-stufige Leitfaden für Fortgeschrittene richtet sich an Cloud-Architekten und Führungskräfte, die Workloads auf einer multiregionalen Architektur aufbauen AWS und daran interessiert sind, die Widerstandsfähigkeit ihrer Workloads zu verbessern. In diesem Leitfaden werden Grundkenntnisse über Infrastruktur und Dienste vorausgesetzt. AWS Es beschreibt allgemeine Anwendungsfälle für mehrere Regionen, erläutert grundlegende Konzepte und Implikationen für mehrere Regionen in Bezug auf Design, Entwicklung und Bereitstellung und bietet präskriptive Anleitungen, anhand derer Sie besser bestimmen können, ob eine Architektur mit mehreren Regionen für Ihre Workloads geeignet ist.

## 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 bieten bewährte Architekturpraktiken für den Entwurf und Betrieb zuverlässiger, sicherer, effizienter, kostengünstiger und nachhaltiger Systeme. Sie können das AWS Well-Architected Tool, das kostenlos auf der Website verfügbar ist, verwenden AWS Management Console, um Ihre Workloads anhand dieser bewährten Methoden zu überprüfen, indem Sie für jede Säule eine Reihe von Fragen beantworten.

Weitere fachkundige Beratung und bewährte Methoden für Ihre Cloud-Architektur, einschließlich Bereitstellungen von Referenzarchitekturen, Diagrammen und technischen Leitfäden, finden Sie im AWS Architecture Center.

## Einführung

Jede AWS-Regionbesteht aus mehreren unabhängigen und physisch getrennten Availability Zones innerhalb eines geografischen Gebiets. Eine strikte logische Trennung zwischen den Softwarediensten in jeder Region wird beibehalten. Dieses zielgerichtete Design stellt sicher, dass ein Infrastruktur- oder Dienstausfall in einer Region nicht zu einem korrelierten Ausfall in einer anderen Region führt.

Die meisten AWS Benutzer können ihre Stabilitätsziele für einen Workload in einer einzelnen Region erreichen, indem sie mehrere Availability Zones oder Regional verwenden. AWS-Services Ein Teil der Benutzer setzt jedoch aus drei Gründen auf Architekturen mit mehreren Regionen:

- Sie haben hohe Anforderungen an Verfügbarkeit und Betriebskontinuität für ihre Workloads der höchsten Ebene und möchten eine begrenzte Wiederherstellungszeit nach Beeinträchtigungen festlegen, die sich auf die Ressourcen in einer einzelnen Region auswirken.
- Sie müssen die Anforderungen an die Datenhoheit erfüllen (z. B. die Einhaltung lokaler Gesetze und Vorschriften und die Einhaltung der Vorschriften), die voraussetzen, dass Workloads innerhalb einer bestimmten Jurisdiktion betrieben werden.
- Sie müssen die Leistung und das Kundenerlebnis für die Workloads verbessern, indem sie die Workloads an Standorten ausführen, die ihren Endbenutzern am nächsten sind.

Dieser Leitfaden konzentriert sich auf die Anforderungen an hohe Verfügbarkeit und Betriebskontinuität und hilft Ihnen, sich mit den Überlegungen zur Einführung einer multiregionalen Architektur für Workloads vertraut zu machen. Es beschreibt grundlegende Konzepte, die für den Entwurf, die Entwicklung und die Bereitstellung eines Workloads mit mehreren Regionen gelten, und bietet einen verbindlichen Rahmen, anhand dessen Sie feststellen können, ob eine Architektur mit mehreren Regionen die richtige Wahl für einen bestimmten Workload ist. Sie müssen sicherstellen, dass eine Architektur mit mehreren Regionen die richtige Wahl für Ihren Workload ist, da diese Architekturen eine Herausforderung darstellen. Wenn die Architektur mit mehreren Regionen nicht korrekt aufgebaut ist, kann es sein, dass die Gesamtverfügbarkeit des Workloads sinkt.

# Entwicklung und Betrieb für Resilienz in einer einzigen Region

Bevor Sie sich mit Konzepten für mehrere Regionen befassen, sollten Sie zunächst sicherstellen, dass Ihr Workload in einer einzelnen Region bereits so belastbar wie möglich ist. Um dies zu erreichen, bewerten Sie Ihre Arbeitslast anhand der Säulen Zuverlässigkeit und Operational Excellence des AWS Well-Architected Framework und nehmen Sie alle erforderlichen Änderungen auf der Grundlage von Kompromissen und Risikobewertungen vor. Die folgenden Konzepte werden im AWS Well-Architected Framework behandelt:

- Workload-Segmentierung auf der Grundlage von Domänengrenzen
- Klar definierte Serviceverträge
- Verwaltung und Kopplung von Abhängigkeiten
- Umgang mit Ausfällen, Wiederholungsversuchen und Back-off-Strategien
- Idempotente Operationen und statusbehaftete und zustandslose Transaktionen
- Betriebsbereitschaft und Change-Management
- Den Zustand der Arbeitslast verstehen
- Auf Ereignisse reagieren

Um die Resilienz einzelner Regionen weiter voranzutreiben, sollten Sie die Konzepte überprüfen und anwenden, die im paper Advanced Multi-AZ Resilience Patterns: Detected and Mitigating Gray Failures erörtert werden. Dieses paper enthält bewährte Methoden für die Verwendung von Replikaten in jeder Availability Zone zur Eindämmung von Ausfällen und geht auf Multi-AZ-Konzepte ein, die im AWS Well Architected Framework eingeführt wurden. Eine Architektur mit mehreren Regionen kann zwar die Ausfallmodi, die an Availability Zones gebunden sind, minimieren, aber es gibt auch Kompromisse, die mit einem multiregionalen Ansatz einhergehen, die Sie in Betracht ziehen sollten. Aus diesem Grund empfehlen wir, mit einem Multi-AZ-Ansatz zu beginnen und dann einen bestimmten Workload anhand der Grundlagen für Architekturen mit mehreren Regionen zu bewerten, um festzustellen, ob ein Ansatz mit mehreren Regionen die Widerstandsfähigkeit des Workloads erhöhen kann.

# Multi-Region-Grundlage 1: Die Anforderungen verstehen

Wie bereits erwähnt, sind hohe Verfügbarkeit und Betriebskontinuität häufige Gründe für die Entwicklung von Architekturen mit mehreren Regionen. Verfügbarkeitsmetriken messen den Prozentsatz der Zeit, in der ein Workload über einen bestimmten Zeitraum genutzt werden kann, wohingegen Kennzahlen zur Betriebskontinuität die Wiederherstellungszeit für umfangreiche und in der Regel längere Ereignisse messen.

Die Messung der Verfügbarkeit ist ein nahezu kontinuierlicher Prozess. Spezifische Messwerte können variieren, basieren aber in der Regel auf einer Zielverfügbarkeitsmetrik, die am häufigsten als Neun bezeichnet wird (z. B. Verfügbarkeit von 99,99 Prozent). Bei Verfügbarkeitszielen gibt es keine Einheitslösung. Sie sollten Verfügbarkeitsziele auf Workload-Ebene festlegen und unkritische Komponenten von kritischen Komponenten trennen, anstatt ein einziges Ziel für alle Workloads festzulegen.

Um die Betriebskontinuität zu gewährleisten, werden in der Regel die folgenden point-in-time Messwerte verwendet:

- Recovery Time Objective (RTO) — RTO ist die maximal zulässige Verzögerung zwischen der Betriebsunterbrechung und der Wiederherstellung des Dienstes. Dieser Wert bestimmt eine akzeptable Dauer, für die der Dienst beeinträchtigt ist.
- Recovery Point Objective (RPO) — RPO ist die maximal zulässige Zeitspanne seit dem letzten Datenwiederherstellungspunkt. Dies bestimmt, was als akzeptabler Datenverlust zwischen dem letzten Wiederherstellungspunkt und einer Betriebsunterbrechung angesehen wird.

Ähnlich wie bei der Festlegung von Verfügbarkeitszielen sollten RTO und RPO auch auf Workload-Ebene definiert werden. Eine aggressivere Betriebskontinuität oder hohe Verfügbarkeit erfordern höhere Investitionen. Allerdings kann oder erfordert nicht jede Anwendung das gleiche Maß an Belastbarkeit. Als Ausgangspunkt kann es hilfreich sein, Geschäfts- und IT-Verantwortliche darauf abzustimmen, die Wichtigkeit von Anwendungen anhand ihrer geschäftlichen Auswirkungen zu bewerten und sie dann entsprechend zu klassifizieren. Die folgenden Tabellen enthalten Beispiele für Tiering.

Diese Tabelle zeigt ein Beispiel für ein Resilienz-Tiering für Service Level Agreements (). SLAs

| Stufe der Resilienz | Verfügbarkeit SLA | Akzeptable Ausfallzeiten/Jahr |
|---|---|---|
| Platin | 99,99 % | 52.60 Minuten |
| Gold | 99,90% | 8,77 Stunden |
| Silber | 99,5 % | 1,83 Tage |

Die folgende Tabelle zeigt ein Beispiel für Resilienz-Tiering für RTO und RPO.

| Stufe der Resilienz | Maximaler RTO | Maximaler RPO | Kriterien | Kosten |
|---|---|---|---|---|
| Platin | 15 Minuten | 5 Minuten | Unternehm enskritische Workloads | $$$ |
| Gold | 15 Minuten — 6 Stunden | 2 Stunden | Wichtige, aber nicht geschäfts kritische Workloads | $$ |
| Silber | 6 Stunden — ein paar Tage | 24 Stunden | Nicht kritische Workloads | $ |

Wenn Sie Workloads auf Ausfallsicherheit auslegen, sollten Sie den Zusammenhang zwischen hoher Verfügbarkeit und Betriebskontinuität berücksichtigen. Wenn ein Workload beispielsweise eine Verfügbarkeit von 99,99 Prozent erfordert, sind nicht mehr als 53 Minuten Ausfallzeit pro Jahr tolerierbar. Es kann mindestens 5 Minuten dauern, bis ein Fehler erkannt wird, und weitere 10 Minuten, bis ein Bediener eingreift, Entscheidungen über Wiederherstellungsschritte trifft und diese Schritte ausführt. Es ist nicht ungewöhnlich, dass es 30 bis 45 Minuten dauert, bis ein einzelnes Problem behoben ist. In diesem Fall ist es von Vorteil, eine Strategie für mehrere Regionen zu verfolgen, um eine isolierte Instanz bereitzustellen, die korrelierte Auswirkungen beseitigt. Auf diese Weise kann der Betrieb fortgesetzt werden, indem innerhalb einer bestimmten Zeit ein Failover vorgenommen wird, während Sie die anfängliche Beeinträchtigung selbstständig prüfen. In

diesem Fall ist es erforderlich, die angemessene begrenzte Wiederherstellungszeit festzulegen und sicherzustellen, dass eine entsprechende Anpassung erfolgt.

Ein regionsübergreifender Ansatz eignet sich möglicherweise für geschäftskritische Workloads mit extremen Verfügbarkeitsanforderungen (z. B. 99,99 Prozent oder mehr Verfügbarkeit) oder strengen Anforderungen an die Betriebskontinuität, die nur durch einen Failover in eine andere Region erfüllt werden können. Diese Anforderungen gelten jedoch in der Regel nur für einen kleinen Teil des Workload-Portfolios eines Unternehmens, für das eine begrenzte Wiederherstellungszeit gilt, die in Minuten oder Stunden gemessen wird. Sofern für eine Anwendung keine Wiederherstellungszeit von Minuten oder einigen Stunden erforderlich ist, ist es möglicherweise besser, abzuwarten, bis eine regionale Unterbrechung der Anwendung in der betroffenen Region behoben ist. Dieser Ansatz ist in der Regel auf Workloads niedrigerer Stufen ausgerichtet.

Vor der Implementierung einer Architektur mit mehreren Regionen sollten sich Entscheidungsträger und technische Teams im Hinblick auf die Auswirkungen auf die Kosten, einschließlich betrieblicher und infrastruktureller Kostentreiber, einig sein. Eine typische Architektur mit mehreren Regionen kann doppelt so hohe Kosten verursachen wie ein Ansatz mit nur einer Region. Zwar gibt es mehrere regionsübergreifende Muster für die Geschäftskontinuität, wie z. B. den Betrieb mit Hot-Standby -, Warm-Standby - oder Kontrolllampe, aber das Muster mit dem geringsten Risiko, die Wiederherstellungsziele zu erreichen, beinhaltet den Betrieb im Hot-Standby-Modus, wodurch sich die Kosten für Ihre Arbeitslast verdoppeln.

## Wichtige Leitlinien

- Ziele für Verfügbarkeit und Kontinuität des Betriebs wie RTO und RPO sollten pro Workload festgelegt und mit den Geschäfts- und IT-Stakeholdern abgestimmt werden.
- Die meisten Ziele in Bezug auf Verfügbarkeit und Kontinuität des Betriebs können innerhalb einer einzigen Region erreicht werden. Für Ziele, die innerhalb einer einzelnen Region nicht erreicht werden können, sollten Sie mehrere Regionen in Betracht ziehen, wobei ein klares Bild von Kompromissen zwischen Kosten, Komplexität und Nutzen gezogen werden muss.

# Multi-Region-Grundlagen 2: Die Daten verstehen

Die Verwaltung von Daten ist ein nicht triviales Problem, wenn Sie Architekturen mit mehreren Regionen einsetzen. Die geografische Entfernung zwischen Regionen führt zu einer unvermeidlichen Latenz, die sich in der Zeit äußert, die für die Replikation von Daten zwischen Regionen benötigt wird. Kompromisse zwischen Verfügbarkeit, Datenkonsistenz und der Einführung einer höheren Latenz bei einem Workload, der eine Architektur mit mehreren Regionen verwendet, werden notwendig sein. Unabhängig davon, ob Sie asynchrone oder synchrone Replikation verwenden, müssen Sie Ihre Anwendung an die Verhaltensänderungen anpassen, die die Replikationstechnologie mit sich bringt. Probleme im Zusammenhang mit Datenkonsistenz und Latenz machen es sehr schwierig, eine bestehende Anwendung, die für eine einzelne Region konzipiert wurde, in mehrere Regionen umzuwandeln. Es ist wichtig, die Anforderungen an die Datenkonsistenz und die Datenzugriffsmuster für bestimmte Workloads zu verstehen, um die Kompromisse abzuwägen.

## 2.a: Die Anforderungen an die Datenkonsistenz verstehen

Das CAP-Theorem bietet eine Referenz für Überlegungen zu den Kompromissen zwischen Datenkonsistenz, Verfügbarkeit und Netzwerkpartitionen. Für einen Workload können nur zwei dieser Anforderungen gleichzeitig erfüllt werden. Per Definition umfasst eine Architektur mit mehreren Regionen Netzwerkpartitionen zwischen Regionen, sodass Sie sich zwischen Verfügbarkeit und Konsistenz entscheiden müssen.

Wenn Sie sich für die regionsübergreifende Verfügbarkeit von Daten entscheiden, treten bei transaktionalen Schreibvorgängen keine nennenswerten Latenzen auf, da die Abhängigkeit von der asynchronen Replikation festgeschriebener Daten zwischen Regionen zu einer verringerten Konsistenz zwischen den Regionen führt, bis die Replikation abgeschlossen ist. Bei der asynchronen Replikation besteht bei einem Ausfall in der primären Region eine hohe Wahrscheinlichkeit, dass Schreibvorgänge noch auf die Replikation aus der primären Region warten. Dies führt zu einem Szenario, in dem die neuesten Daten erst verfügbar sind, wenn die Replikation wieder aufgenommen wird, und ein Abgleichprozess erforderlich ist, um laufende Transaktionen zu verarbeiten, die nicht aus der Region repliziert wurden, in der die Unterbrechung aufgetreten ist. Für dieses Szenario müssen Sie Ihre Geschäftslogik verstehen und einen speziellen Prozess erstellen, um die Transaktion erneut abzuspielen oder Datenspeicher zwischen Regionen zu vergleichen.

Für Workloads, bei denen asynchrone Replikation bevorzugt wird, können Sie Services wie Amazon Aurora und Amazon DynamoDB für die asynchrone regionsübergreifende Replikation verwenden. Sowohl die globalen Amazon Aurora Aurora-Datenbanken als auch die globalen

Amazon DynamoDB-Tabellen verfügen über CloudWatch Standardkennzahlen von Amazon, um die Überwachung von Replikationsverzögerungen zu unterstützen. Eine globale Aurora-Datenbank besteht aus einer primären Region, in die Ihre Daten geschrieben werden, und bis zu fünf schreibgeschützten sekundären Regionen. Globale DynamoDB-Tabellen bestehen aus multiaktiven Replikattabellen in einer beliebigen Anzahl von Regionen, in die Ihre Daten geschrieben und aus denen Ihre Daten gelesen werden.

Bei einer Strategie mit mehreren Regionen ist es von Vorteil, den Workload so zu gestalten, dass er die Vorteile ereignisgesteuerter Architekturen nutzt, da der Workload die asynchrone Replikation von Daten umfassen kann und der Status durch die Wiedergabe von Ereignissen wiederhergestellt werden kann. Da Streaming- und Messaging-Dienste Nachrichtennutzdaten in einer einzigen Region zwischenspeichern, muss ein regionaler Failover- oder Failback-Prozess einen Mechanismus zur Umleitung von Client-Eingabedatenströmen beinhalten. Der Prozess muss auch Payloads abgleichen, die während der Übertragung übertragen wurden oder nicht zugestellt wurden, die in der Region gespeichert sind, in der die Störung aufgetreten ist.

Wenn Sie sich für die GAP-Konsistenzanforderung entscheiden und eine synchron replizierte Datenbank in allen Regionen verwenden, um Ihre Anwendungen zu unterstützen, die gleichzeitig in mehreren Regionen ausgeführt werden, vermeiden Sie das Risiko eines Datenverlusts und sorgen dafür, dass die Daten zwischen den Regionen synchron bleiben. Dies führt jedoch zu höheren Latenzeigenschaften, da Schreibvorgänge auf mehr als eine Region übertragen werden müssen und die Regionen Hunderte oder Tausende von Meilen voneinander entfernt sein können. Sie müssen dieses Latenzmerkmal in Ihrem Anwendungsdesign berücksichtigen. Darüber hinaus kann die synchrone Replikation das Risiko korrelierter Fehler mit sich bringen, da Schreibvorgänge in mehr als einer Region gespeichert werden müssen, um erfolgreich zu sein. Wenn es innerhalb einer Region zu einer Beeinträchtigung kommt, müssen Sie ein Quorum bilden, damit Schreibvorgänge erfolgreich sind. Dies beinhaltet in der Regel die Einrichtung Ihrer Datenbank in drei Regionen und die Einrichtung eines Quorums für zwei von drei Regionen. Technologien wie Paxos können dabei helfen, Daten synchron zu replizieren und zu übertragen, erfordern jedoch erhebliche Investitionen von Entwicklern.

Wenn Schreibvorgänge eine synchrone Replikation über mehrere Regionen hinweg beinhalten, um hohe Konsistenzanforderungen zu erfüllen, erhöht sich die Schreiblatenz um eine Größenordnung. Eine höhere Schreiblatenz ist normalerweise nicht etwas, das Sie in einer Anwendung nachrüsten können, ohne wesentliche Änderungen, wie z. B. die Timeout- und Wiederholungsstrategie für Ihre Anwendung, zu ändern. Im Idealfall muss dies bei der ersten Entwicklung der Anwendung berücksichtigt werden. Für Workloads mit mehreren Regionen, bei denen die synchrone Replikation Priorität hat, können AWS Partner Lösungen helfen.

# 2.b: Datenzugriffsmuster verstehen

Arbeitslast-Datenzugriffsmuster sind entweder leseintensiv oder schreibintensiv. Wenn Sie dieses Merkmal für einen bestimmten Workload verstehen, können Sie eine geeignete Architektur für mehrere Regionen auswählen.

Für leseintensive Workloads wie statische Inhalte, die vollständig schreibgeschützt sind, können Sie eine aktiv-aktive Multiregions-Architektur erreichen, die im Vergleich zu einem schreibintensiven Workload eine geringere technische Komplexität aufweist. Die Bereitstellung statischer Inhalte am Netzwerkrand mithilfe eines Content Delivery Network (CDN) gewährleistet die Verfügbarkeit, indem Inhalte zwischengespeichert werden, die dem Endbenutzer am nächsten sind. Die Verwendung von Funktionen wie Origin Failover innerhalb von Amazon CloudFront kann dazu beitragen, dies zu erreichen. Eine weitere Option ist die Bereitstellung von statusfreiem Computing in mehreren Regionen und die Verwendung von DNS, um Benutzer zur nächstgelegenen Region weiterzuleiten, um die Inhalte zu lesen. Sie können Amazon Route 53 mit einer Geolocation-Routing-Richtlinie verwenden, um dies zu erreichen.

Für leseintensive Workloads, die einen höheren Prozentsatz an Lesetraffic als an Schreibverkehr haben, können Sie eine lokale Lesestrategie und eine globale Schreibstrategie verwenden. Das bedeutet, dass alle Schreibanforderungen an eine Datenbank in einer bestimmten Region gesendet werden, die Daten asynchron in alle anderen Regionen repliziert werden und Lesevorgänge in jeder Region ausgeführt werden können. Dieser Ansatz erfordert einen Workload, der letztendlich für Konsistenz sorgt, da lokale Lesevorgänge aufgrund der erhöhten Latenz bei der regionsübergreifenden Replikation von Schreibvorgängen veraltet sein können.

Die globalen Aurora-Datenbanken können dabei helfen, Lesereplikate in einer Standby-Region bereitzustellen, die ausschließlich den gesamten Lesetraffic lokal verarbeiten kann, und einen einzigen primären Datenspeicher in einer bestimmten Region für den Schreibverkehr bereitzustellen. Daten werden asynchron von der Primärdatenbank in Standby-Datenbanken (Read Replicas) repliziert, und die Standby-Datenbanken können zur Primärdatenbank heraufgestuft werden, wenn Sie einen Failover von Vorgängen auf die Standby-Region durchführen müssen. Sie können bei diesem Ansatz auch DynamoDB verwenden. DynamoDB-Tabellen können regionsübergreifende Replikattabellen bereitstellen, die jeweils skaliert werden können, um ein beliebiges Volumen an lokalem Lese- oder Schreibverkehr zu unterstützen. Wenn eine Anwendung Daten in eine Replikattabelle in einer Region schreibt, verteilt DynamoDB den Schreibvorgang automatisch auf die anderen Replikattabellen in den übrigen -Regionen. Mit dieser Konfiguration werden Daten asynchron aus einer definierten Primärregion in Replikattabellen in Standby-Regionen repliziert. Replikattabellen in jeder Region können immer Schreibvorgänge akzeptieren, sodass das

Heraufstufen einer Standby-Region zur primären Region auf Anwendungsebene erfolgt. Auch hier muss der Workload irgendwann konsistent sein, weshalb er möglicherweise neu geschrieben werden muss, wenn er nicht von Anfang an darauf ausgelegt war.

Für schreibintensive Workloads sollte eine primäre Region ausgewählt werden, und die Fähigkeit, auf eine Standby-Region umzuschalten, sollte in den Workload integriert werden. Im Vergleich zu einem Active-Active-Ansatz hat ein primärer Standby-Ansatz zusätzliche Kompromisse. Dies liegt daran, dass bei einer Active-Active-Architektur der Workload neu geschrieben werden muss, um intelligentes Routing zu Regionen zu ermöglichen, Sitzungsaffinität herzustellen, idempotente Transaktionen sicherzustellen und potenzielle Konflikte zu bewältigen.

Für die meisten Workloads, die aus Gründen der Resilienz einen Ansatz mit mehreren Regionen verwenden, ist kein aktiv-aktiver Ansatz erforderlich. Sie können eine Sharding-Strategie verwenden, um die Widerstandsfähigkeit zu erhöhen, indem Sie das Ausmaß der Auswirkungen einer Beeinträchtigung auf den gesamten Kundenstamm begrenzen. Wenn Sie einen Kundenstamm effektiv teilen können, können Sie für jeden Shard unterschiedliche Hauptregionen auswählen. Sie können Kunden beispielsweise so teilen, dass die Hälfte der Kunden der Region eins und die andere Hälfte der Region zwei zugeordnet ist. Indem Sie Regionen als Zellen behandeln, können Sie einen Ansatz mit mehreren Regionen schaffen, wodurch der Umfang der Auswirkungen auf Ihre Arbeitslast reduziert wird. Weitere Informationen zu diesem Ansatz finden Sie in der AWS re:Invent-Präsentation.

Sie können den Sharding-Ansatz mit einem Primary-Standby-Ansatz kombinieren, um Failover-Funktionen für die Shards bereitzustellen. Sie müssen einen getesteten Failover-Prozess für die Arbeitslast und auch einen Prozess für den Datenabgleich einrichten, um die Transaktionskonsistenz der Datenspeicher nach dem Failover sicherzustellen. Diese werden später in diesem Handbuch ausführlicher behandelt.

## Wichtige Hinweise

- Es besteht eine hohe Wahrscheinlichkeit, dass Schreibvorgänge, die zur Replikation noch ausstehen, nicht in die Standby-Region übernommen werden, wenn ein Fehler auftritt. Daten sind erst verfügbar, wenn die Replikation wieder aufgenommen wird (unter der Annahme einer asynchronen Replikation).
- Im Rahmen des Failovers ist ein Datenabgleich erforderlich, um sicherzustellen, dass bei Datenspeichern, die asynchrone Replikation verwenden, ein transaktionskonsistenter Zustand aufrechterhalten wird. Dies erfordert eine spezielle Geschäftslogik und wird nicht vom Datenspeicher selbst verwaltet.

- Wenn eine hohe Konsistenz erforderlich ist, müssen die Workloads so angepasst werden, dass sie die erforderliche Latenz eines Datenspeichers, der synchron repliziert, tolerieren.

# Grundlegendes 3 für mehrere Regionen: Verstehen Sie Ihre Workload-Abhängigkeiten

Ein bestimmter Workload kann in einer Region mehrere Abhängigkeiten aufweisen, z. B. AWS-Services verwendete, interne Abhängigkeiten, Abhängigkeiten von Drittanbietern, Netzwerkabhängigkeiten, Zertifikate, Schlüssel, Geheimnisse und Parameter. Um den Betrieb des Workloads in einem Ausfallszenario zu gewährleisten, sollten keine Abhängigkeiten zwischen der primären Region und der Standby-Region bestehen. Jede Region sollte unabhängig voneinander arbeiten können. Um dies zu erreichen, sollten Sie alle Abhängigkeiten im Workload genau prüfen, um sicherzustellen, dass sie in jeder Region verfügbar sind. Dies ist erforderlich, da sich ein Ausfall in der primären Region nicht auf die Standby-Region auswirken sollte. Darüber hinaus müssen Sie wissen, wie der Workload funktioniert, wenn sich eine Abhängigkeit in einem schlechten Zustand befindet oder gar nicht verfügbar ist, damit Sie Lösungen entwickeln können, die diesem Problem angemessen gerecht werden.

## 3.a: AWS-Services

Wenn Sie eine Architektur mit mehreren Regionen entwerfen, ist es wichtig zu verstehen, welche Funktionen diese Dienste für mehrere Regionen bieten und welche Lösungen Sie entwickeln müssen, um Ziele für mehrere Regionen zu erreichen. AWS-Services Beispielsweise können Amazon Aurora und Amazon DynamoDB Daten asynchron in eine Standby-Region replizieren. Alle AWS-Service Abhängigkeiten müssen in allen Regionen verfügbar sein, in denen ein Workload ausgeführt werden soll. Überprüfen Sie die Liste nach Regionen, um zu überprüfen, ob die AWS-Services von Ihnen verwendeten Dienste in den gewünschten Regionen verfügbar sind.

## 3.b: Interne Abhängigkeiten und Abhängigkeiten von Drittanbietern

Stellen Sie sicher, dass die internen Abhängigkeiten aller Workloads in den Regionen verfügbar sind, von denen aus sie ausgeführt werden. Wenn der Workload beispielsweise aus vielen Microservices besteht, identifizieren Sie alle Microservices, die eine Geschäftsfähigkeit ausmachen, und stellen Sie sicher, dass all diese Microservices in jeder Region bereitgestellt werden, von der aus der Workload ausgeführt wird. Definieren Sie alternativ eine Strategie für den ordnungsgemäßen Umgang mit Microservices, die nicht mehr verfügbar sind.

Von regionsübergreifenden Aufrufen zwischen Microservices innerhalb eines Workloads wird abgeraten, und die regionale Isolation sollte beibehalten werden. Dies liegt daran, dass die Schaffung

regionsübergreifender Abhängigkeiten das Risiko eines korrelierten Ausfalls erhöht, wodurch die Vorteile isolierter regionaler Implementierungen des Workloads zunichte gemacht werden. Lokale Abhängigkeiten können ebenfalls Teil der Arbeitslast sein. Daher ist es wichtig zu verstehen, wie sich die Merkmale dieser Integrationen ändern könnten, wenn sich die primäre Region ändern würde. Wenn sich die Standby-Region beispielsweise weiter von der lokalen Umgebung entfernt befindet, kann sich die erhöhte Latenz negativ auswirken.

Wenn Sie Software-as-a-Service (SaaS) -Lösungen, Software Development Kits (SDKs) und andere Abhängigkeiten von Produkten von Drittanbietern verstehen und Szenarien ausführen können, in denen diese Abhängigkeiten entweder beeinträchtigt oder nicht verfügbar sind, erhalten Sie einen besseren Einblick in die Funktionsweise und das Verhalten der Systemkette unter verschiedenen Ausfallmodi. Diese Abhängigkeiten können in Ihrem Anwendungscode enthalten sein, z. B. die externe Verwaltung von Geheimnissen mithilfe AWS Secrets Manager, oder sie können eine Tresorlösung eines Drittanbieters (z. B. HashiCorp) oder Authentifizierungssysteme betreffen, die auf AWS IAM Identity CenterVerbundanmeldungen angewiesen sind.

Redundanz in Bezug auf Abhängigkeiten kann die Widerstandsfähigkeit erhöhen. Wenn eine SaaS-Lösung oder eine Abhängigkeit von einem Drittanbieter denselben primären AWS-Region Workload verwendet, ermitteln Sie gemeinsam mit dem Anbieter, ob dessen Ausfallsicherheit Ihren Anforderungen an den Workload entspricht.

Achten Sie außerdem darauf, dass der Workload und seine Abhängigkeiten, wie z. B. Anwendungen von Drittanbietern, gemeinsam genutzt werden. Wenn die Abhängigkeiten nach einem Failover in (oder von) einer sekundären Region nicht verfügbar sind, wird der Workload möglicherweise nicht vollständig wiederhergestellt.

## 3.c: Failover-Mechanismus

DNS wird häufig als Failover-Mechanismus verwendet, um den Verkehr von der primären Region in eine Standby-Region zu verlagern. Prüfen und überprüfen Sie kritisch alle Abhängigkeiten, die der Failover-Mechanismus mit sich bringt. Wenn Ihr Workload beispielsweise Amazon Route 53 verwendet, `us-east-1` bedeutet das Wissen, dass die Kontrollebene in dieser bestimmten Region gehostet wird, eine Abhängigkeit von der Kontrollebene in dieser bestimmten Region einzugehen. Dies wird nicht als Teil eines Failover-Mechanismus empfohlen, wenn die primäre Region auch `us-east-1` deshalb gilt, weil dadurch ein einziger Fehlerpunkt entsteht. Wenn Sie einen anderen Failover-Mechanismus verwenden, sollten Sie ein tiefes Verständnis von Szenarien haben, in denen ein Failover nicht wie erwartet funktionieren würde, und dann für Notfälle planen oder, falls erforderlich, einen neuen Mechanismus entwickeln. Lesen Sie den Blogbeitrag Creating Disaster

Recovery Mechanisms Using Amazon Route 53, um mehr über Methoden zu erfahren, mit denen Sie ein erfolgreiches Failover durchführen können.

Wie im vorherigen Abschnitt beschrieben, müssen alle Microservices, die Teil einer Geschäftsfunktion sind, in jeder Region verfügbar sein, in der der Workload bereitgestellt wird. Im Rahmen der Failover-Strategie sollten alle Microservices, die Teil der Geschäftsfähigkeit sind, gemeinsam ein Failover durchführen, um die Möglichkeit regionsübergreifender Anrufe auszuschließen. Wenn Microservices unabhängig voneinander ein Failover durchführen, besteht alternativ die Möglichkeit, dass unerwünschtes Verhalten auftritt, z. B. wenn Microservices möglicherweise regionsübergreifende Anrufe tätigen. Dies führt zu Latenz und kann dazu führen, dass der Workload während der Client-Timeouts nicht mehr verfügbar ist.

## 3.d: Abhängigkeiten von der Konfiguration

Zertifikate, Schlüssel, Geheimnisse, Amazon Machine Images (AMIs), Container-Images und Parameter sind Teil der Abhängigkeitsanalyse, die beim Entwurf einer Architektur mit mehreren Regionen erforderlich ist. Wann immer möglich, ist es am besten, diese Komponenten innerhalb der einzelnen Regionen zu lokalisieren, damit sie in Bezug auf diese Abhängigkeiten nicht von mehreren Regionen gemeinsam genutzt werden. Sie sollten beispielsweise die Ablaufdaten von Zertifikaten variieren, um zu verhindern, dass sich ein ablaufendes Zertifikat (bei Alarmen auf „Im Voraus benachrichtigen“) auf mehrere Regionen auswirkt.

Verschlüsselungsschlüssel und -geheimnisse sollten ebenfalls regionsspezifisch sein. Auf diese Weise sind die Auswirkungen auf eine bestimmte Region beschränkt, wenn bei der Rotation eines Schlüssels oder Geheimnisses ein Fehler auftritt.

Schließlich sollten alle Workload-Parameter lokal gespeichert werden, damit der Workload in der jeweiligen Region abgerufen werden kann.

## Wichtige Hinweise

- Eine Architektur mit mehreren Regionen profitiert von der physischen und logischen Trennung zwischen Regionen. Durch die Einführung regionsübergreifender Abhängigkeiten auf Anwendungsebene wird dieser Vorteil zunichte gemacht. Vermeiden Sie solche Abhängigkeiten.
- Die Failover-Steuerung sollte ohne Abhängigkeiten von der primären Region funktionieren.
- Der Failover sollte während der gesamten Benutzererfahrung koordiniert werden, um die Möglichkeit einer erhöhten Latenz und Abhängigkeit regionsübergreifender Anrufe zu vermeiden.

# Grundlage 4 für mehrere Regionen: Einsatzbereitschaft

Der Betrieb eines Workloads mit mehreren Regionen ist eine komplexe Aufgabe, die mit betrieblichen Herausforderungen verbunden ist, die für eine Architektur mit mehreren Regionen spezifisch sind. Dazu gehören die AWS-Konto Verwaltung, die Anpassung der Bereitstellungsprozesse, die Entwicklung einer regionsübergreifenden Beobachtungsstrategie, die Erstellung und Erprobung von Wiederherstellungsprozessen und die anschließende Verwaltung der Kosten. Ein Operational Readiness Review (ORR) kann Teams dabei helfen, einen Workload für die Produktion vorzubereiten, unabhängig davon, ob er in einer einzelnen Region oder in mehreren Regionen ausgeführt wird.

## 4.a: Verwaltung AWS-Konto

Um einen Workload flächenübergreifend bereitzustellen AWS-Regionen, stellen Sie sicher, dass alle AWS-Service Kontingente innerhalb eines Accounts in allen Regionen paritätisch verteilt sind. Identifizieren Sie zunächst alle AWS-Services Komponenten der Architektur, schauen Sie sich die geplante Nutzung in den Standby-Regionen an und vergleichen Sie dann die geplante Nutzung mit der aktuellen Nutzung. In einigen Fällen, wenn die Standby-Region noch nicht genutzt wurde, können Sie sich auf die Standard-Servicekontingente beziehen, um den Ausgangspunkt zu verstehen. Fordern Sie dann für alle Dienste, die verwendet werden, eine Erhöhung des Kontingents an, indem Sie die Konsole für Service Quotas verwenden (Anmeldung erforderlich) oder APIs.

Konfigurieren Sie AWS Identity and Access Management (IAM) -Rollen in jeder Region, um Operatoren, Automatisierungstools und AWS-Services die entsprechenden Berechtigungen für Ressourcen in der Standby-Region zuzuweisen. Um eine regionale Isolierung für Architekturen mit mehreren Regionen zu erreichen, isolieren Sie die Rollen nach Regionen. Stellen Sie sicher, dass die erforderlichen Berechtigungen vorhanden sind, bevor Sie eine Standby-Region live schalten.

## 4.b: Bereitstellungspraktiken

Funktionen für mehrere Regionen können die Bereitstellung eines Workloads in mehreren Regionen erschweren. Sie müssen sicherstellen, dass Sie die Bereitstellung jeweils in einer Region durchführen. Wenn Sie beispielsweise einen Aktiv-Passiv-Ansatz verwenden, sollten Sie die Bereitstellung zuerst in der primären Region und dann in der Standby-Region durchführen. AWS CloudFormationunterstützt Sie bei der Bereitstellung der Infrastruktur in einer oder mehreren Regionen und kann an Ihre Bedürfnisse angepasst werden. AWS CodePipelinehilft

Ihnen beim Aufbau einer integration/continuous delivery (CI/CD (kontinuierlichen) Pipeline mit regionsübergreifenden Aktionen, die den Einsatz in Regionen ermöglichen, die sich von der Region unterscheiden, in der sich die Pipeline befindet. In Kombination mit robusten Bereitstellungsstrategien wie Blau/Grün ermöglicht dies eine Bereitstellung mit minimalen bis gar keinen Ausfallzeiten.

Die Bereitstellung von Stateful-Capabilities kann jedoch komplexer werden, wenn der Status der Anwendung oder der Daten nicht in einen persistenten Speicher ausgelagert wird. Passen Sie in diesen Situationen den Bereitstellungsprozess sorgfältig an Ihre Bedürfnisse an. Entwerfen Sie die Bereitstellungspipeline und den Prozess so, dass die Bereitstellung in jeweils einer Region statt in mehreren Regionen gleichzeitig erfolgt. Dadurch wird die Wahrscheinlichkeit korrelierter Ausfälle zwischen den Regionen verringert. Weitere Informationen zu den Techniken, die Amazon zur Automatisierung von Softwarebereitstellungen verwendet, finden Sie im Artikel Automating safe, hands-off Deployments in der AWS Builders' Library.

## 4.c: Beobachtbarkeit

Denken Sie bei der Planung für mehrere Regionen darüber nach, wie Sie den Zustand aller Komponenten in jeder Region überwachen möchten, um einen ganzheitlichen Überblick über den regionalen Zustand zu erhalten. Dies könnte die Überwachung von Metriken im Hinblick auf Replikationsverzögerungen beinhalten, was bei Workloads mit nur einer Region nicht berücksichtigt wird.

Wenn Sie eine Architektur mit mehreren Regionen erstellen, sollten Sie die Leistung des Workloads auch von den Standby-Regionen aus beobachten. Dazu gehören auch Integritätsprüfungen und Canaries (synthetische Tests), die von der Standby-Region aus durchgeführt werden, um einen externen Überblick über den Zustand der primären Region zu erhalten. Darüber hinaus können Sie Amazon CloudWatch Internet Monitor verwenden, um den Status des externen Netzwerks und die Leistung Ihrer Workloads aus der Sicht eines Endbenutzers zu verstehen. In der primären Region sollte dieselbe Beobachtbarkeit für die Überwachung der Standby-Region vorhanden sein.

Die Kanaren aus der Standby-Region sollten die Kennzahlen zum Kundenerlebnis überwachen, um den allgemeinen Zustand der Arbeitslast zu ermitteln. Dies ist erforderlich, da bei einem Problem in der primären Region die Beobachtbarkeit in der primären Region beeinträchtigt werden könnte, was sich negativ auf Ihre Fähigkeit auswirken würde, den Zustand der Arbeitslast einzuschätzen.

In diesem Fall kann eine Beobachtung außerhalb dieser Region Aufschluss geben. Diese Kennzahlen sollten in Dashboards zusammengefasst werden, die in jeder Region verfügbar sind, und in Alarmen, die in jeder Region erstellt werden. Da CloudWatches sich um einen regionalen Dienst

handelt, ist es erforderlich, Alarme in beiden Regionen zu haben. Diese Überwachungsdaten werden verwendet, um den Anruf zum Failover von einer primären Region in eine Standby-Region zu tätigen.

## 4.d: Prozesse und Verfahren

Der beste Zeitpunkt, um die Frage „Wann sollte ich einen Failover durchführen?“ zu beantworten ist lange bevor du es musst. Definieren Sie rechtzeitig vor dem Auftreten eines Problems Wiederherstellungspläne, die Personen, Prozesse und Technologien einbeziehen, und testen Sie sie regelmäßig. Entscheiden Sie sich für einen Rahmen für Sanierungsentscheidungen. Wenn es einen gut geübten Wiederherstellungsprozess gibt und die Dauer bis zur Wiederherstellung genau bekannt ist, können Sie den Wiederherstellungsprozess mit einem Failover starten, das das RTO-Ziel erfüllt. Dieser Zeitpunkt kann unmittelbar nach der Identifizierung eines Problems mit der Anwendung in der primären Region liegen, oder es könnte sich um einen weiteren Zeitpunkt handeln, zu dem die Wiederherstellungsoptionen innerhalb der Anwendung in der Region ausgeschöpft sind.

Die Failover-Aktion selbst sollte zu 100 Prozent automatisiert sein, aber die Entscheidung, den Failover zu aktivieren, sollte von Menschen getroffen werden — in der Regel von einer kleinen Anzahl vorher festgelegter Personen in der Organisation. Diese Personen sollten den Verlust von Daten und Informationen über das Ereignis in Betracht ziehen. Außerdem müssen die Kriterien für einen Failover klar definiert und innerhalb der Organisation allgemein bekannt sein. Um diese Prozesse zu definieren und abzuschließen, können Sie AWS Systems Manager Runbooks verwenden, die eine vollständige end-to-end Automatisierung ermöglichen und die Konsistenz der Prozesse sicherstellen, die während der Tests und beim Failover ausgeführt werden.

Diese Runbooks sollten in den Primär- und Standby-Regionen verfügbar sein, um die Failover- oder Failback-Prozesse zu starten. Sobald diese Automatisierung eingeführt ist, definieren Sie einen regelmäßigen Testrhythmus und halten Sie ihn ein. Auf diese Weise wird sichergestellt, dass bei einem tatsächlichen Ereignis die Reaktion auf einen klar definierten, eingeübten Prozess erfolgt, in den das Unternehmen Vertrauen hat. Es ist auch wichtig, die festgelegten Toleranzen für Datenabgleichsprozesse zu berücksichtigen. Vergewissern Sie sich, dass der vorgeschlagene Prozess die festgelegten RPO/RTO-Anforderungen erfüllt.

## 4.e: Testen

Ein ungetesteter Wiederherstellungsansatz ist gleichbedeutend damit, keinen Wiederherstellungsansatz zu haben. Eine grundlegende Testphase besteht darin, ein Wiederherstellungsverfahren durchzuführen, um die Betriebsregion für Ihre Anwendung zu wechseln.

Manchmal wird dies als Ansatz zur Anwendungsrotation bezeichnet. Wir empfehlen Ihnen, die Funktion zum Umschalten von Regionen in Ihre normale Betriebsposition einzubauen. Dieser Test allein reicht jedoch nicht aus.

Resilienztests sind auch wichtig, um den Wiederherstellungsansatz einer Anwendung zu validieren. Dazu gehören die Implementierung bestimmter Ausfallszenarien, das Verständnis dafür, wie Ihre Anwendung und Ihr Wiederherstellungsprozess reagieren, und dann die Implementierung aller erforderlichen Abhilfemaßnahmen, falls der Test nicht wie geplant verlaufen sollte. Wenn Sie Ihr Wiederherstellungsverfahren ohne Fehler testen, erfahren Sie nicht, wie sich Ihre Anwendung insgesamt verhält, wenn Fehler auftreten. Sie müssen einen Plan entwickeln, um Ihre Wiederherstellung anhand erwarteter Ausfallszenarien zu testen. AWS Fault Injection Servicebietet eine ständig wachsende Liste von Szenarien, um Ihnen den Einstieg zu erleichtern.

Dies ist besonders wichtig für Hochverfügbarkeitsanwendungen, bei denen strenge Tests erforderlich sind, um sicherzustellen, dass die Geschäftskontinuitätsziele erreicht werden. Durch proaktives Testen der Wiederherstellungsfunktionen wird das Risiko von Produktionsausfällen reduziert, wodurch das Vertrauen gestärkt wird, dass die Anwendung die gewünschte begrenzte Wiederherstellungszeit erreichen kann. Durch regelmäßige Tests wird auch betriebliches Fachwissen aufgebaut, sodass das Team bei auftretenden Ausfällen schnell und zuverlässig die Wiederherstellung durchführen kann. Die Berücksichtigung der menschlichen Komponente oder des Prozesses Ihres Wiederherstellungsansatzes ist genauso wichtig wie die technischen Aspekte.

## 4.f: Kosten und Komplexität

Die Auswirkungen einer Architektur mit mehreren Regionen auf die Kosten sind auf eine höhere Infrastrukturnutzung, höhere Betriebskosten und mehr Ressourcenaufwand zurückzuführen. Wie bereits erwähnt, entsprechen die Infrastrukturkosten in einer Standby-Region den Infrastrukturkosten in einer primären Region bei der Vorbereitstellung, sodass sich Ihre Gesamtkosten verdoppeln. Stellen Sie Kapazität so bereit, dass sie für den täglichen Betrieb ausreicht, aber dennoch genügend Pufferkapazität reserviert ist, um Nachfragespitzen zu verkraften. Konfigurieren Sie dann in jeder Region dieselben Grenzwerte.

Wenn Sie eine Active-Active-Architektur verwenden, müssen Sie außerdem möglicherweise Änderungen auf Anwendungsebene vornehmen, um Ihre Anwendung in einer Architektur mit mehreren Regionen erfolgreich auszuführen. Die Entwicklung und der Betrieb dieser Änderungen können zeit- und ressourcenintensiv sein. Unternehmen müssen zumindest Zeit damit verbringen, die technischen und geschäftlichen Abhängigkeiten in den einzelnen Regionen zu verstehen und Failover- und Failback-Prozesse zu entwerfen.

Die Teams sollten außerdem normale Failover- und Failback-Übungen durchführen, um sich mit Runbooks, die während einer Veranstaltung verwendet werden, vertraut zu machen. Diese Übungen sind zwar entscheidend, um das erwartete Ergebnis einer Investition in mehreren Regionen zu erzielen, sie bedeuten jedoch auch Opportunitätskosten und nehmen Zeit und Ressourcen für andere Aktivitäten in Anspruch.

## 4.g: Organisatorische Failover-Strategie für mehrere Regionen

AWS-Regionen bietet Grenzen zur Fehlerisolierung, um korrelierte Ausfälle zu verhindern und die Auswirkungen von etwaigen AWS-Service Beeinträchtigungen auf eine einzelne Region einzudämmen. Sie können diese Fehlergrenzen verwenden, um regionsübergreifende Anwendungen zu erstellen, die aus unabhängigen, fehlerisolierten Replikaten in jeder Region bestehen, um Szenarien mit gemeinsamem Schicksal zu begrenzen. Auf diese Weise können Sie regionsübergreifende Anwendungen erstellen und eine Reihe von Ansätzen — von Backup und Wiederherstellung über Pilotbetrieb bis hin zu Active-Active — zur Implementierung Ihrer multiregionalen Architektur verwenden. Anwendungen werden jedoch in der Regel nicht isoliert betrieben. Berücksichtigen Sie daher sowohl die Komponenten, die Sie verwenden werden, als auch deren Abhängigkeiten als Teil Ihrer Failover-Strategie. Im Allgemeinen arbeiten mehrere Anwendungen zusammen, um eine User Story zu unterstützen. Dabei handelt es sich um eine spezielle Funktion, die einem Endbenutzer angeboten wird, z. B. das Posten eines Bilds und einer Bildunterschrift in einer Social-Media-App oder das Auschecken auf einer E-Commerce-Website. Aus diesem Grund sollten Sie eine unternehmensweite Failover-Strategie für mehrere Regionen entwickeln, die für die nötige Koordination und Konsistenz sorgt, damit Ihr Ansatz erfolgreich ist.

Es gibt vier übergeordnete Strategien, aus denen Unternehmen als Leitfaden für einen regionenübergreifenden Ansatz wählen können. Diese sind vom detailliertesten bis zum umfassendsten Ansatz aufgeführt:

- Failover auf Komponentenebene
- Failover einzelner Anwendungen
- Failover mit Abhängigkeitsdiagramm
- Failover für das gesamte Anwendungsportfolio

Jede Strategie hat Kompromisse und geht auf unterschiedliche Herausforderungen ein. Dazu gehören Flexibilität bei der Entscheidungsfindung, die Fähigkeit, die Failover-Kombinationen zu testen, modales Verhalten und organisatorische Investitionen in Planung und Implementierung.

Weitere Informationen zu den einzelnen Strategien finden Sie im AWS Blogbeitrag Creating an Organizational Multi-Regions Failover-Strategie.

## Wichtige Hinweise

- Überprüfen Sie alle AWS-Service Kontingente, um sicherzustellen, dass sie in allen Regionen, in denen die Arbeitslast ausgeführt wird, einheitlich sind.
- Der Bereitstellungsprozess sollte jeweils auf eine Region abzielen, anstatt mehrere Regionen gleichzeitig einzubeziehen.
- Zusätzliche Messwerte wie die Verzögerung bei der Replikation sind spezifisch für Szenarien mit mehreren Regionen und sollten überwacht werden.
- Erweitern Sie die Überwachung der Arbeitslast über die primäre Region hinaus. Überwachen Sie die Kennzahlen zum Kundenerlebnis für jede Region und messen Sie diese Daten außerhalb jeder Region, in der ein Workload ausgeführt wird.
- Testen Sie Failover und Failback regelmäßig. Implementieren Sie ein einziges Runbook für Failover- und Failback-Prozesse und verwenden Sie es sowohl für Tests als auch für Live-Events. Runbooks für Tests und Live-Events sollten sich nicht unterscheiden.
- Machen Sie sich mit den Kompromissen der Failover-Strategien vertraut. Implementieren Sie ein Abhängigkeitsdiagramm oder eine Strategie für das gesamte Anwendungsportfolio.

# Fazit und Ressourcen

In diesem Leitfaden wurden allgemeine Anwendungsfälle für Architekturen mit mehreren Regionen, die Grundlagen der Implementierung dieser Architekturen und die Auswirkungen dieses Ansatzes behandelt. Sie können diese Grundlagen auf jeden Workload anwenden und anhand der Informationen entscheiden, ob eine Architektur mit mehreren Regionen der richtige Ansatz für Ihr Unternehmen ist.

Weitere Informationen finden Sie in den folgenden Ressourcen:

- AWS Zentrum für Architektur
- AWS Well-Architected Framework
- AWS Well-Architected Tool
- Erstellung einer organisatorischen Failover-Strategie für mehrere Regionen (Blogbeitrag)AWS
- AWS Funktionen für mehrere Regionen (re:POST-Artikel)AWS

# Dokumentverlauf

In der folgenden Tabelle werden wichtige Änderungen in diesem Leitfaden beschrieben. Um Benachrichtigungen über zukünftige Aktualisierungen zu erhalten, können Sie einen RSS-Feed abonnieren.

| Änderung | Beschreibung | Datum |
| --- | --- | --- |
| Aktualisierungen | Aktualisierungen im gesamten Leitfaden. | 27. Dezember 2024 |
| Erste Veröffentlichung | — | 20. Dezember 2022 |

# AWS Glossar zu präskriptiven Leitlinien

Die folgenden Begriffe werden häufig in Strategien, Leitfäden und Mustern von AWS Prescriptive Guidance verwendet. Um Einträge vorzuschlagen, verwenden Sie bitte den Link Feedback geben am Ende des Glossars.

## Zahlen

7 Rs

Sieben gängige Migrationsstrategien für die Verlagerung von Anwendungen in die Cloud. Diese Strategien bauen auf den 5 Rs auf, die Gartner 2011 identifiziert hat, und bestehen aus folgenden Elementen:

- Faktorwechsel/Architekturwechsel – Verschieben Sie eine Anwendung und ändern Sie ihre Architektur, indem Sie alle Vorteile cloudnativer Feature nutzen, um Agilität, Leistung und Skalierbarkeit zu verbessern. Dies beinhaltet in der Regel die Portierung des Betriebssystems und der Datenbank. Beispiel: Migrieren Sie Ihre lokale Oracle-Datenbank auf die Amazon Aurora PostgreSQL-kompatible Edition.
- Plattformwechsel (Lift and Reshape) – Verschieben Sie eine Anwendung in die Cloud und führen Sie ein gewisses Maß an Optimierung ein, um die Cloud-Funktionen zu nutzen. Beispiel: Migrieren Sie Ihre lokale Oracle-Datenbank zu Amazon Relational Database Service (Amazon RDS) für Oracle in der. AWS Cloud
- Neukauf (Drop and Shop) – Wechseln Sie zu einem anderen Produkt, indem Sie typischerweise von einer herkömmlichen Lizenz zu einem SaaS-Modell wechseln. Beispiel: Migrieren Sie Ihr CRM-System (Customer Relationship Management) zu Salesforce.com.
- Hostwechsel (Lift and Shift) – Verschieben Sie eine Anwendung in die Cloud, ohne Änderungen vorzunehmen, um die Cloud-Funktionen zu nutzen. Beispiel: Migrieren Sie Ihre lokale Oracle-Datenbank zu Oracle auf einer EC2 Instanz in der. AWS Cloud
- Verschieben (Lift and Shift auf Hypervisor-Ebene) – Verlagern Sie die Infrastruktur in die Cloud, ohne neue Hardware kaufen, Anwendungen umschreiben oder Ihre bestehenden Abläufe ändern zu müssen. Sie migrieren Server von einer lokalen Plattform zu einem Cloud-Dienst für dieselbe Plattform. Beispiel: Migrieren Sie eine Microsoft Hyper-V Anwendung zu AWS.
- Beibehaltung (Wiederaufgreifen) – Bewahren Sie Anwendungen in Ihrer Quellumgebung auf. Dazu können Anwendungen gehören, die einen umfangreichen Faktorwechsel erfordern und

die Sie auf einen späteren Zeitpunkt verschieben möchten, sowie ältere Anwendungen, die Sie beibehalten möchten, da es keine geschäftliche Rechtfertigung für ihre Migration gibt.

- Außerbetriebnahme – Dekommissionierung oder Entfernung von Anwendungen, die in Ihrer Quellumgebung nicht mehr benötigt werden.

# A

ABAC

Siehe attributbasierte Zugriffskontrolle.

abstrahierte Dienste

Weitere Informationen finden Sie unter Managed Services.

ACID

Siehe Atomarität, Konsistenz, Isolierung und Haltbarkeit.

Aktiv-Aktiv-Migration

Eine Datenbankmigrationsmethode, bei der die Quell- und Zieldatenbanken synchron gehalten werden (mithilfe eines bidirektionalen Replikationstools oder dualer Schreibvorgänge) und beide Datenbanken Transaktionen von miteinander verbundenen Anwendungen während der Migration verarbeiten. Diese Methode unterstützt die Migration in kleinen, kontrollierten Batches, anstatt einen einmaligen Cutover zu erfordern. Es ist flexibler, erfordert aber mehr Arbeit als eine aktiv-passive Migration.

Aktiv-Passiv-Migration

Eine Datenbankmigrationsmethode, bei der die Quell- und Zieldatenbanken synchron gehalten werden, aber nur die Quelldatenbank Transaktionen von verbindenden Anwendungen verarbeitet, während Daten in die Zieldatenbank repliziert werden. Die Zieldatenbank akzeptiert während der Migration keine Transaktionen.

Aggregatfunktion

Eine SQL-Funktion, die mit einer Gruppe von Zeilen arbeitet und einen einzelnen Rückgabewert für die Gruppe berechnet. Beispiele für Aggregatfunktionen sind `SUM` und`MAX`.

AI

Siehe künstliche Intelligenz.

AIOps

Siehe Operationen im Bereich künstliche Intelligenz.

Anonymisierung

Der Prozess des dauerhaften Löschens personenbezogener Daten in einem Datensatz. Anonymisierung kann zum Schutz der Privatsphäre beitragen. Anonymisierte Daten gelten nicht mehr als personenbezogene Daten.

Anti-Muster

Eine häufig verwendete Lösung für ein wiederkehrendes Problem, bei dem die Lösung kontraproduktiv, ineffektiv oder weniger wirksam als eine Alternative ist.

Anwendungssteuerung

Ein Sicherheitsansatz, bei dem nur zugelassene Anwendungen verwendet werden können, um ein System vor Schadsoftware zu schützen.

Anwendungsportfolio

Eine Sammlung detaillierter Informationen zu jeder Anwendung, die von einer Organisation verwendet wird, einschließlich der Kosten für die Erstellung und Wartung der Anwendung und ihres Geschäftswerts. Diese Informationen sind entscheidend für den Prozess der Portfoliofindung und -analyse und hilft bei der Identifizierung und Priorisierung der Anwendungen, die migriert, modernisiert und optimiert werden sollen.

künstliche Intelligenz (KI)

Das Gebiet der Datenverarbeitungswissenschaft, das sich der Nutzung von Computertechnologien zur Ausführung kognitiver Funktionen widmet, die typischerweise mit Menschen in Verbindung gebracht werden, wie Lernen, Problemlösen und Erkennen von Mustern. Weitere Informationen finden Sie unter Was ist künstliche Intelligenz?

Operationen mit künstlicher Intelligenz (AIOps)

Der Prozess des Einsatzes von Techniken des Machine Learning zur Lösung betrieblicher Probleme, zur Reduzierung betrieblicher Zwischenfälle und menschlicher Eingriffe sowie zur Steigerung der Servicequalität. Weitere Informationen zur Verwendung in der AWS Migrationsstrategie finden Sie im Operations Integration Guide. AIOps

Asymmetrische Verschlüsselung

Ein Verschlüsselungsalgorithmus, der ein Schlüsselpaar, einen öffentlichen Schlüssel für die Verschlüsselung und einen privaten Schlüssel für die Entschlüsselung verwendet. Sie können den

öffentlichen Schlüssel teilen, da er nicht für die Entschlüsselung verwendet wird. Der Zugriff auf den privaten Schlüssel sollte jedoch stark eingeschränkt sein.

Atomizität, Konsistenz, Isolierung, Haltbarkeit (ACID)

Eine Reihe von Softwareeigenschaften, die die Datenvalidität und betriebliche Zuverlässigkeit einer Datenbank auch bei Fehlern, Stromausfällen oder anderen Problemen gewährleisten.

Attributbasierte Zugriffskontrolle (ABAC)

Die Praxis, detaillierte Berechtigungen auf der Grundlage von Benutzerattributen wie Abteilung, Aufgabenrolle und Teamname zu erstellen. Weitere Informationen finden Sie unter ABAC AWS in der AWS Identity and Access Management (IAM-) Dokumentation.

autoritative Datenquelle

Ein Ort, an dem Sie die primäre Version der Daten speichern, die als die zuverlässigste Informationsquelle angesehen wird. Sie können Daten aus der maßgeblichen Datenquelle an andere Speicherorte kopieren, um die Daten zu verarbeiten oder zu ändern, z. B. zu anonymisieren, zu redigieren oder zu pseudonymisieren.

Availability Zone

Ein bestimmter Standort innerhalb einer AWS-Region , der vor Ausfällen in anderen Availability Zones geschützt ist und kostengünstige Netzwerkkonnektivität mit niedriger Latenz zu anderen Availability Zones in derselben Region bietet.

AWS Framework für die Cloud-Einführung (AWS CAF)

Ein Framework mit Richtlinien und bewährten Verfahren, das Unternehmen bei der Entwicklung eines effizienten und effektiven Plans für den erfolgreichen Umstieg auf die Cloud unterstützt. AWS AWS CAF unterteilt die Leitlinien in sechs Schwerpunktbereiche, die als Perspektiven bezeichnet werden: Unternehmen, Mitarbeiter, Unternehmensführung, Plattform, Sicherheit und Betrieb. Die Perspektiven Geschäft, Mitarbeiter und Unternehmensführung konzentrieren sich auf Geschäftskompetenzen und -prozesse, während sich die Perspektiven Plattform, Sicherheit und Betriebsabläufe auf technische Fähigkeiten und Prozesse konzentrieren. Die Personalperspektive zielt beispielsweise auf Stakeholder ab, die sich mit Personalwesen (HR), Personalfunktionen und Personalmanagement befassen. Aus dieser Perspektive bietet AWS CAF Leitlinien für Personalentwicklung, Schulung und Kommunikation, um das Unternehmen auf eine erfolgreiche Cloud-Einführung vorzubereiten. Weitere Informationen finden Sie auf der AWS -CAF-Webseite und dem AWS -CAF-Whitepaper.

AWS Workload-Qualifizierungsrahmen (AWS WQF)

Ein Tool, das Workloads bei der Datenbankmigration bewertet, Migrationsstrategien empfiehlt und Arbeitsschätzungen bereitstellt. AWS WQF ist in () enthalten. AWS Schema Conversion Tool AWS SCT Es analysiert Datenbankschemas und Codeobjekte, Anwendungscode, Abhängigkeiten und Leistungsmerkmale und stellt Bewertungsberichte bereit.

# B

schlechter Bot

Ein Bot, der Einzelpersonen oder Organisationen stören oder ihnen Schaden zufügen soll.

BCP

Siehe Planung der Geschäftskontinuität.

Verhaltensdiagramm

Eine einheitliche, interaktive Ansicht des Ressourcenverhaltens und der Interaktionen im Laufe der Zeit. Sie können ein Verhaltensdiagramm mit Amazon Detective verwenden, um fehlgeschlagene Anmeldeversuche, verdächtige API-Aufrufe und ähnliche Vorgänge zu untersuchen. Weitere Informationen finden Sie unter Daten in einem Verhaltensdiagramm in der Detective-Dokumentation.

Big-Endian-System

Ein System, welches das höchstwertige Byte zuerst speichert. Siehe auch Endianness.

Binäre Klassifikation

Ein Prozess, der ein binäres Ergebnis vorhersagt (eine von zwei möglichen Klassen). Beispielsweise könnte Ihr ML-Modell möglicherweise Probleme wie „Handelt es sich bei dieser E-Mail um Spam oder nicht?“ vorhersagen müssen oder „Ist dieses Produkt ein Buch oder ein Auto?“

Bloom-Filter

Eine probabilistische, speichereffiziente Datenstruktur, mit der getestet wird, ob ein Element Teil einer Menge ist.

Blau/Grün-Bereitstellung

Eine Bereitstellungsstrategie, bei der Sie zwei separate, aber identische Umgebungen erstellen. Sie führen die aktuelle Anwendungsversion in einer Umgebung (blau) und die neue

Anwendungsversion in der anderen Umgebung (grün) aus. Mit dieser Strategie können Sie schnell und mit minimalen Auswirkungen ein Rollback durchführen.

Bot

Eine Softwareanwendung, die automatisierte Aufgaben über das Internet ausführt und menschliche Aktivitäten oder Interaktionen simuliert. Manche Bots sind nützlich oder nützlich, wie z. B. Webcrawler, die Informationen im Internet indexieren. Einige andere Bots, sogenannte bösartige Bots, sollen Einzelpersonen oder Organisationen stören oder ihnen Schaden zufügen.

Botnetz

Netzwerke von Bots, die mit Malware infiziert sind und unter der Kontrolle einer einzigen Partei stehen, die als Bot-Herder oder Bot-Operator bezeichnet wird. Botnetze sind der bekannteste Mechanismus zur Skalierung von Bots und ihrer Wirkung.

branch

Ein containerisierter Bereich eines Code-Repositorys. Der erste Zweig, der in einem Repository erstellt wurde, ist der Hauptzweig. Sie können einen neuen Zweig aus einem vorhandenen Zweig erstellen und dann Feature entwickeln oder Fehler in dem neuen Zweig beheben. Ein Zweig, den Sie erstellen, um ein Feature zu erstellen, wird allgemein als Feature-Zweig bezeichnet. Wenn das Feature zur Veröffentlichung bereit ist, führen Sie den Feature-Zweig wieder mit dem Hauptzweig zusammen. Weitere Informationen finden Sie unter Über Branches (GitHub Dokumentation).

Zugang durch Glasbruch

Unter außergewöhnlichen Umständen und im Rahmen eines genehmigten Verfahrens ist dies eine schnelle Methode für einen Benutzer, auf einen Bereich zuzugreifen AWS-Konto , für den er in der Regel keine Zugriffsrechte besitzt. Weitere Informationen finden Sie unter dem Indikator Implementation break-glass procedures in den AWS Well-Architected-Leitlinien.

Brownfield-Strategie

Die bestehende Infrastruktur in Ihrer Umgebung. Wenn Sie eine Brownfield-Strategie für eine Systemarchitektur anwenden, richten Sie sich bei der Gestaltung der Architektur nach den Einschränkungen der aktuellen Systeme und Infrastruktur. Wenn Sie die bestehende Infrastruktur erweitern, könnten Sie Brownfield- und Greenfield-Strategien mischen.

Puffer-Cache

Der Speicherbereich, in dem die am häufigsten abgerufenen Daten gespeichert werden.

Geschäftsfähigkeit

Was ein Unternehmen tut, um Wert zu generieren (z. B. Vertrieb, Kundenservice oder Marketing). Microservices-Architekturen und Entwicklungsentscheidungen können von den Geschäftskapazitäten beeinflusst werden. Weitere Informationen finden Sie im Abschnitt Organisiert nach Geschäftskapazitäten des Whitepapers Ausführen von containerisierten Microservices in AWS.

Planung der Geschäftskontinuität (BCP)

Ein Plan, der die potenziellen Auswirkungen eines störenden Ereignisses, wie z. B. einer groß angelegten Migration, auf den Betrieb berücksichtigt und es einem Unternehmen ermöglicht, den Betrieb schnell wieder aufzunehmen.

# C

CAF

Weitere Informationen finden Sie unter Framework für die AWS Cloud-Einführung.

Bereitstellung auf Kanaren

Die langsame und schrittweise Veröffentlichung einer Version für Endbenutzer. Wenn Sie sich sicher sind, stellen Sie die neue Version bereit und ersetzen die aktuelle Version vollständig.

CCoE

Weitere Informationen finden Sie im Cloud Center of Excellence.

CDC

Siehe Erfassung von Änderungsdaten.

Erfassung von Datenänderungen (CDC)

Der Prozess der Nachverfolgung von Änderungen an einer Datenquelle, z. B. einer Datenbanktabelle, und der Aufzeichnung von Metadaten zu der Änderung. Sie können CDC für verschiedene Zwecke verwenden, z. B. für die Prüfung oder Replikation von Änderungen in einem Zielsystem, um die Synchronisation aufrechtzuerhalten.

Chaos-Technik

Absichtliches Einführen von Ausfällen oder Störungsereignissen, um die Widerstandsfähigkeit eines Systems zu testen. Sie können AWS Fault Injection Service (AWS FIS) verwenden, um Experimente durchzuführen, die Ihre AWS Workloads stress, und deren Reaktion zu bewerten.

CI/CD

Siehe Continuous Integration und Continuous Delivery.

Klassifizierung

Ein Kategorisierungsprozess, der bei der Erstellung von Vorhersagen hilft. ML-Modelle für Klassifikationsprobleme sagen einen diskreten Wert voraus. Diskrete Werte unterscheiden sich immer voneinander. Beispielsweise muss ein Modell möglicherweise auswerten, ob auf einem Bild ein Auto zu sehen ist oder nicht.

clientseitige Verschlüsselung

Lokale Verschlüsselung von Daten, bevor das Ziel sie AWS-Service empfängt.

Cloud-Exzellenzzentrum (CCoE)

Ein multidisziplinäres Team, das die Cloud-Einführung in der gesamten Organisation vorantreibt, einschließlich der Entwicklung bewährter Cloud-Methoden, der Mobilisierung von Ressourcen, der Festlegung von Migrationszeitplänen und der Begleitung der Organisation durch groß angelegte Transformationen. Weitere Informationen finden Sie in den CCoE-Beiträgen im AWS Cloud Enterprise Strategy Blog.

Cloud Computing

Die Cloud-Technologie, die typischerweise für die Ferndatenspeicherung und das IoT-Gerätemanagement verwendet wird. Cloud Computing ist häufig mit Edge-Computing-Technologie verbunden.

Cloud-Betriebsmodell

In einer IT-Organisation das Betriebsmodell, das zum Aufbau, zur Weiterentwicklung und Optimierung einer oder mehrerer Cloud-Umgebungen verwendet wird. Weitere Informationen finden Sie unter Aufbau Ihres Cloud-Betriebsmodells.

Phasen der Einführung der Cloud

Die vier Phasen, die Unternehmen bei der Migration in der Regel durchlaufen AWS Cloud:

- Projekt – Durchführung einiger Cloud-bezogener Projekte zu Machbarkeitsnachweisen und zu Lernzwecken
- Fundament — Tätigen Sie grundlegende Investitionen, um Ihre Cloud-Einführung zu skalieren (z. B. Einrichtung einer landing zone, Definition eines CCo E, Einrichtung eines Betriebsmodells)

- Migration – Migrieren einzelner Anwendungen
- Neuentwicklung – Optimierung von Produkten und Services und Innovation in der Cloud

Diese Phasen wurden von Stephen Orban im Blogbeitrag The Journey Toward Cloud-First & the Stages of Adoption im AWS Cloud Enterprise Strategy-Blog definiert. Informationen darüber, wie sie mit der AWS Migrationsstrategie zusammenhängen, finden Sie im Leitfaden zur Vorbereitung der Migration.

CMDB

Siehe Datenbank für das Konfigurationsmanagement.

Code-Repository

Ein Ort, an dem Quellcode und andere Komponenten wie Dokumentation, Beispiele und Skripts gespeichert und im Rahmen von Versionskontrollprozessen aktualisiert werden. Zu den gängigen Cloud-Repositorys gehören GitHub oderBitbucket Cloud. Jede Version des Codes wird Zweig genannt. In einer Microservice-Struktur ist jedes Repository einer einzelnen Funktionalität gewidmet. Eine einzelne CI/CD-Pipeline kann mehrere Repositorien verwenden.

Kalter Cache

Ein Puffer-Cache, der leer oder nicht gut gefüllt ist oder veraltete oder irrelevante Daten enthält. Dies beeinträchtigt die Leistung, da die Datenbank-Instance aus dem Hauptspeicher oder der Festplatte lesen muss, was langsamer ist als das Lesen aus dem Puffercache.

Kalte Daten

Daten, auf die selten zugegriffen wird und die in der Regel historisch sind. Bei der Abfrage dieser Art von Daten sind langsame Abfragen in der Regel akzeptabel. Durch die Verlagerung dieser Daten auf leistungsschwächere und kostengünstigere Speicherstufen oder -klassen können Kosten gesenkt werden.

Computer Vision (CV)

Ein Bereich der KI, der maschinelles Lernen nutzt, um Informationen aus visuellen Formaten wie digitalen Bildern und Videos zu analysieren und zu extrahieren. Amazon SageMaker AI bietet beispielsweise Bildverarbeitungsalgorithmen für CV.

Drift in der Konfiguration

Bei einer Arbeitslast eine Änderung der Konfiguration gegenüber dem erwarteten Zustand. Dies kann dazu führen, dass der Workload nicht mehr richtlinienkonform wird, und zwar in der Regel schrittweise und unbeabsichtigt.

Verwaltung der Datenbankkonfiguration (CMDB)

Ein Repository, das Informationen über eine Datenbank und ihre IT-Umgebung speichert und verwaltet, inklusive Hardware- und Softwarekomponenten und deren Konfigurationen. In der Regel verwenden Sie Daten aus einer CMDB in der Phase der Portfolioerkennung und -analyse der Migration.

Konformitätspaket

Eine Sammlung von AWS Config Regeln und Abhilfemaßnahmen, die Sie zusammenstellen können, um Ihre Konformitäts- und Sicherheitsprüfungen individuell anzupassen. Mithilfe einer YAML-Vorlage können Sie ein Conformance Pack als einzelne Entität in einer AWS-Konto AND-Region oder unternehmensweit bereitstellen. Weitere Informationen finden Sie in der Dokumentation unter Conformance Packs. AWS Config

Kontinuierliche Bereitstellung und kontinuierliche Integration (CI/CD)

Der Prozess der Automatisierung der Quell-, Build-, Test-, Staging- und Produktionsphasen des Softwareveröffentlichungsprozesses. CI/CD is commonly described as a pipeline. CI/CDkann Ihnen helfen, Prozesse zu automatisieren, die Produktivität zu steigern, die Codequalität zu verbessern und schneller zu liefern. Weitere Informationen finden Sie unter Vorteile der kontinuierlichen Auslieferung. CD kann auch für kontinuierliche Bereitstellung stehen. Weitere Informationen finden Sie unterKontinuierliche Auslieferung im Vergleich zu kontinuierlicher Bereitstellung.

CV

Siehe Computer Vision.

# D

Daten im Ruhezustand

Daten, die in Ihrem Netzwerk stationär sind, z. B. Daten, die sich im Speicher befinden.

Datenklassifizierung

Ein Prozess zur Identifizierung und Kategorisierung der Daten in Ihrem Netzwerk auf der Grundlage ihrer Kritikalität und Sensitivität. Sie ist eine wichtige Komponente jeder Strategie für das Management von Cybersecurity-Risiken, da sie Ihnen hilft, die geeigneten Schutz- und Aufbewahrungskontrollen für die Daten zu bestimmen. Die Datenklassifizierung ist ein Bestandteil

der Sicherheitssäule im AWS Well-Architected Framework. Weitere Informationen finden Sie unter Datenklassifizierung.

Datendrift

Eine signifikante Variation zwischen den Produktionsdaten und den Daten, die zum Trainieren eines ML-Modells verwendet wurden, oder eine signifikante Änderung der Eingabedaten im Laufe der Zeit. Datendrift kann die Gesamtqualität, Genauigkeit und Fairness von ML-Modellvorhersagen beeinträchtigen.

Daten während der Übertragung

Daten, die sich aktiv durch Ihr Netzwerk bewegen, z. B. zwischen Netzwerkressourcen.

Datennetz

Ein architektonisches Framework, das verteilte, dezentrale Dateneigentum mit zentraler Verwaltung und Steuerung ermöglicht.

Datenminimierung

Das Prinzip, nur die Daten zu sammeln und zu verarbeiten, die unbedingt erforderlich sind. Durch Datenminimierung im AWS Cloud können Datenschutzrisiken, Kosten und der CO2-Fußabdruck Ihrer Analysen reduziert werden.

Datenperimeter

Eine Reihe präventiver Schutzmaßnahmen in Ihrer AWS Umgebung, die sicherstellen, dass nur vertrauenswürdige Identitäten auf vertrauenswürdige Ressourcen von erwarteten Netzwerken zugreifen. Weitere Informationen finden Sie unter Aufbau eines Datenperimeters auf. AWS

Vorverarbeitung der Daten

Rohdaten in ein Format umzuwandeln, das von Ihrem ML-Modell problemlos verarbeitet werden kann. Die Vorverarbeitung von Daten kann bedeuten, dass bestimmte Spalten oder Zeilen entfernt und fehlende, inkonsistente oder doppelte Werte behoben werden.

Herkunft der Daten

Der Prozess der Nachverfolgung des Ursprungs und der Geschichte von Daten während ihres gesamten Lebenszyklus, z. B. wie die Daten generiert, übertragen und gespeichert wurden.

betroffene Person

Eine Person, deren Daten gesammelt und verarbeitet werden.

Data Warehouse

Ein Datenverwaltungssystem, das Business Intelligence wie Analysen unterstützt. Data Warehouses enthalten in der Regel große Mengen historischer Daten und werden in der Regel für Abfragen und Analysen verwendet.

Datenbankdefinitionssprache (DDL)

Anweisungen oder Befehle zum Erstellen oder Ändern der Struktur von Tabellen und Objekten in einer Datenbank.

Datenbankmanipulationssprache (DML)

Anweisungen oder Befehle zum Ändern (Einfügen, Aktualisieren und Löschen) von Informationen in einer Datenbank.

DDL

Siehe Datenbankdefinitionssprache.

Deep-Ensemble

Mehrere Deep-Learning-Modelle zur Vorhersage kombinieren. Sie können Deep-Ensembles verwenden, um eine genauere Vorhersage zu erhalten oder um die Unsicherheit von Vorhersagen abzuschätzen.

Deep Learning

Ein ML-Teilbereich, der mehrere Schichten künstlicher neuronaler Netzwerke verwendet, um die Zuordnung zwischen Eingabedaten und Zielvariablen von Interesse zu ermitteln.

defense-in-depth

Ein Ansatz zur Informationssicherheit, bei dem eine Reihe von Sicherheitsmechanismen und -kontrollen sorgfältig in einem Computernetzwerk verteilt werden, um die Vertraulichkeit, Integrität und Verfügbarkeit des Netzwerks und der darin enthaltenen Daten zu schützen. Wenn Sie diese Strategie anwenden AWS, fügen Sie mehrere Steuerelemente auf verschiedenen Ebenen der AWS Organizations Struktur hinzu, um die Ressourcen zu schützen. Ein defense-in-depth Ansatz könnte beispielsweise Multi-Faktor-Authentifizierung, Netzwerksegmentierung und Verschlüsselung kombinieren.

delegierter Administrator

In AWS Organizations kann ein kompatibler Dienst ein AWS Mitgliedskonto registrieren, um die Konten der Organisation und die Berechtigungen für diesen Dienst zu verwalten. Dieses Konto

wird als delegierter Administrator für diesen Service bezeichnet. Weitere Informationen und eine Liste kompatibler Services finden Sie unter Services, die mit AWS Organizations funktionieren in der AWS Organizations -Dokumentation.

Bereitstellung

Der Prozess, bei dem eine Anwendung, neue Feature oder Codekorrekturen in der Zielumgebung verfügbar gemacht werden. Die Bereitstellung umfasst das Implementieren von Änderungen an einer Codebasis und das anschließende Erstellen und Ausführen dieser Codebasis in den Anwendungsumgebungen.

Entwicklungsumgebung

Siehe Umgebung.

Detektivische Kontrolle

Eine Sicherheitskontrolle, die darauf ausgelegt ist, ein Ereignis zu erkennen, zu protokollieren und zu warnen, nachdem ein Ereignis eingetreten ist. Diese Kontrollen stellen eine zweite Verteidigungslinie dar und warnen Sie vor Sicherheitsereignissen, bei denen die vorhandenen präventiven Kontrollen umgangen wurden. Weitere Informationen finden Sie unter Detektivische Kontrolle in Implementierung von Sicherheitskontrollen in AWS.

Abbildung des Wertstroms in der Entwicklung (DVSM)

Ein Prozess zur Identifizierung und Priorisierung von Einschränkungen, die sich negativ auf Geschwindigkeit und Qualität im Lebenszyklus der Softwareentwicklung auswirken. DVSM erweitert den Prozess der Wertstromanalyse, der ursprünglich für Lean-Manufacturing-Praktiken konzipiert wurde. Es konzentriert sich auf die Schritte und Teams, die erforderlich sind, um durch den Softwareentwicklungsprozess Mehrwert zu schaffen und zu steigern.

digitaler Zwilling

Eine virtuelle Darstellung eines realen Systems, z. B. eines Gebäudes, einer Fabrik, einer Industrieanlage oder einer Produktionslinie. Digitale Zwillinge unterstützen vorausschauende Wartung, Fernüberwachung und Produktionsoptimierung.

Maßtabelle

In einem Sternschema eine kleinere Tabelle, die Datenattribute zu quantitativen Daten in einer Faktentabelle enthält. Bei Attributen von Dimensionstabellen handelt es sich in der Regel um Textfelder oder diskrete Zahlen, die sich wie Text verhalten. Diese Attribute werden häufig zum Einschränken von Abfragen, zum Filtern und zur Kennzeichnung von Ergebnismengen verwendet.

Katastrophe

Ein Ereignis, das verhindert, dass ein Workload oder ein System seine Geschäftsziele an seinem primären Einsatzort erfüllt. Diese Ereignisse können Naturkatastrophen, technische Ausfälle oder das Ergebnis menschlichen Handelns sein, z. B. unbeabsichtigte Fehlkonfigurationen oder ein Malware-Angriff.

Disaster Recovery (DR)

Die Strategie und der Prozess, die Sie verwenden, um Ausfallzeiten und Datenverluste aufgrund einer Katastrophe zu minimieren. Weitere Informationen finden Sie unter Disaster Recovery von Workloads unter AWS: Wiederherstellung in der Cloud im AWS Well-Architected Framework.

DML

Siehe Sprache zur Datenbankmanipulation.

Domainorientiertes Design

Ein Ansatz zur Entwicklung eines komplexen Softwaresystems, bei dem seine Komponenten mit sich entwickelnden Domains oder Kerngeschäftszielen verknüpft werden, denen jede Komponente dient. Dieses Konzept wurde von Eric Evans in seinem Buch Domaingesteuertes Design: Bewältigen der Komplexität im Herzen der Software (Boston: Addison-Wesley Professional, 2003) vorgestellt. Informationen darüber, wie Sie domaingesteuertes Design mit dem Strangler-Fig-Muster verwenden können, finden Sie unter Schrittweises Modernisieren älterer Microsoft ASP.NET (ASMX)-Webservices mithilfe von Containern und Amazon API Gateway.

DR

Siehe Disaster Recovery.

Erkennung von Driften

Verfolgung von Abweichungen von einer Basiskonfiguration. Sie können es beispielsweise verwenden, AWS CloudFormation um Abweichungen bei den Systemressourcen zu erkennen, oder Sie können AWS Control Tower damit Änderungen in Ihrer landing zone erkennen, die sich auf die Einhaltung von Governance-Anforderungen auswirken könnten.

DVSM

Siehe Abbildung des Wertstroms in der Entwicklung.

# E

EDA

Siehe explorative Datenanalyse.

EDI

Siehe elektronischer Datenaustausch.

Edge-Computing

Die Technologie, die die Rechenleistung für intelligente Geräte an den Rändern eines IoT-Netzwerks erhöht. Im Vergleich zu Cloud Computing kann Edge Computing die Kommunikationslatenz reduzieren und die Reaktionszeit verbessern.

elektronischer Datenaustausch (EDI)

Der automatisierte Austausch von Geschäftsdokumenten zwischen Organisationen. Weitere Informationen finden Sie unter Was ist elektronischer Datenaustausch.

Verschlüsselung

Ein Rechenprozess, der Klartextdaten, die für Menschen lesbar sind, in Chiffretext umwandelt.

Verschlüsselungsschlüssel

Eine kryptografische Zeichenfolge aus zufälligen Bits, die von einem Verschlüsselungsalgorithmus generiert wird. Schlüssel können unterschiedlich lang sein, und jeder Schlüssel ist so konzipiert, dass er unvorhersehbar und einzigartig ist.

Endianismus

Die Reihenfolge, in der Bytes im Computerspeicher gespeichert werden. Big-Endian-Systeme speichern das höchstwertige Byte zuerst. Little-Endian-Systeme speichern das niedrigwertigste Byte zuerst.

Endpunkt

Siehe Service-Endpunkt.

Endpunkt-Services

Ein Service, den Sie in einer Virtual Private Cloud (VPC) hosten können, um ihn mit anderen Benutzern zu teilen. Sie können einen Endpunktdienst mit anderen AWS-Konten oder AWS Identity and Access Management (IAM AWS PrivateLink -) Prinzipalen erstellen und diesen

Berechtigungen gewähren. Diese Konten oder Prinzipale können sich privat mit Ihrem Endpunktservice verbinden, indem sie Schnittstellen-VPC-Endpunkte erstellen. Weitere Informationen finden Sie unter Einen Endpunkt-Service erstellen in der Amazon Virtual Private Cloud (Amazon VPC)-Dokumentation.

Unternehmensressourcenplanung (ERP)

Ein System, das wichtige Geschäftsprozesse (wie Buchhaltung, MES und Projektmanagement) für ein Unternehmen automatisiert und verwaltet.

Envelope-Verschlüsselung

Der Prozess der Verschlüsselung eines Verschlüsselungsschlüssels mit einem anderen Verschlüsselungsschlüssel. Weitere Informationen finden Sie unter Envelope-Verschlüsselung in der AWS Key Management Service (AWS KMS) -Dokumentation.

Umgebung

Eine Instance einer laufenden Anwendung. Die folgenden Arten von Umgebungen sind beim Cloud-Computing üblich:

- Entwicklungsumgebung – Eine Instance einer laufenden Anwendung, die nur dem Kernteam zur Verfügung steht, das für die Wartung der Anwendung verantwortlich ist. Entwicklungsumgebungen werden verwendet, um Änderungen zu testen, bevor sie in höhere Umgebungen übertragen werden. Diese Art von Umgebung wird manchmal als Testumgebung bezeichnet.
- Niedrigere Umgebungen – Alle Entwicklungsumgebungen für eine Anwendung, z. B. solche, die für erste Builds und Tests verwendet wurden.
- Produktionsumgebung – Eine Instance einer laufenden Anwendung, auf die Endbenutzer zugreifen können. In einer CI/CD-Pipeline ist die Produktionsumgebung die letzte Bereitstellungsumgebung.
- Höhere Umgebungen – Alle Umgebungen, auf die auch andere Benutzer als das Kernentwicklungsteam zugreifen können. Dies kann eine Produktionsumgebung, Vorproduktionsumgebungen und Umgebungen für Benutzerakzeptanztests umfassen.

Epics

In der agilen Methodik sind dies funktionale Kategorien, die Ihnen helfen, Ihre Arbeit zu organisieren und zu priorisieren. Epics bieten eine allgemeine Beschreibung der Anforderungen und Implementierungsaufgaben. Zu den Sicherheitsthemen AWS von CAF gehören beispielsweise Identitäts- und Zugriffsmanagement, Detektivkontrollen, Infrastruktursicherheit,

Datenschutz und Reaktion auf Vorfälle. Weitere Informationen zu Epics in der AWS - Migrationsstrategie finden Sie im Leitfaden zur Programm-Implementierung.

ERP

Siehe Enterprise Resource Planning.

Explorative Datenanalyse (EDA)

Der Prozess der Analyse eines Datensatzes, um seine Hauptmerkmale zu verstehen. Sie sammeln oder aggregieren Daten und führen dann erste Untersuchungen durch, um Muster zu finden, Anomalien zu erkennen und Annahmen zu überprüfen. EDA wird durchgeführt, indem zusammenfassende Statistiken berechnet und Datenvisualisierungen erstellt werden.

# F

Faktentabelle

Die zentrale Tabelle in einem Sternschema. Sie speichert quantitative Daten über den Geschäftsbetrieb. In der Regel enthält eine Faktentabelle zwei Arten von Spalten: Spalten, die Kennzahlen enthalten, und Spalten, die einen Fremdschlüssel für eine Dimensionstabelle enthalten.

schnell scheitern

Eine Philosophie, die häufige und inkrementelle Tests verwendet, um den Entwicklungslebenszyklus zu verkürzen. Dies ist ein wichtiger Bestandteil eines agilen Ansatzes.

Grenze zur Fehlerisolierung

Dabei handelt es sich um eine Grenze AWS Cloud, z. B. eine Availability Zone AWS-Region, eine Steuerungsebene oder eine Datenebene, die die Auswirkungen eines Fehlers begrenzt und die Widerstandsfähigkeit von Workloads verbessert. Weitere Informationen finden Sie unter Grenzen zur AWS Fehlerisolierung.

Feature-Zweig

Siehe Zweig.

Features

Die Eingabedaten, die Sie verwenden, um eine Vorhersage zu treffen. In einem Fertigungskontext könnten Feature beispielsweise Bilder sein, die regelmäßig von der Fertigungslinie aus aufgenommen werden.

Bedeutung der Feature

Wie wichtig ein Feature für die Vorhersagen eines Modells ist. Dies wird in der Regel als numerischer Wert ausgedrückt, der mit verschiedenen Techniken wie Shapley Additive Explanations (SHAP) und integrierten Gradienten berechnet werden kann. Weitere Informationen finden Sie unter Interpretierbarkeit von Modellen für maschinelles Lernen mit AWS.

Featuretransformation

Daten für den ML-Prozess optimieren, einschließlich der Anreicherung von Daten mit zusätzlichen Quellen, der Skalierung von Werten oder der Extraktion mehrerer Informationssätze aus einem einzigen Datenfeld. Das ermöglicht dem ML-Modell, von den Daten profitieren. Wenn Sie beispielsweise das Datum „27.05.2021 00:15:37“ in „2021“, „Mai“, „Donnerstag“ und „15“ aufschlüsseln, können Sie dem Lernalgorithmus helfen, nuancierte Muster zu erlernen, die mit verschiedenen Datenkomponenten verknüpft sind.

Eingabeaufforderung mit wenigen Klicks

Bereitstellung einer kleinen Anzahl von Beispielen, die die Aufgabe und das gewünschte Ergebnis veranschaulichen, bevor das LLM aufgefordert wird, eine ähnliche Aufgabe auszuführen. Bei dieser Technik handelt es sich um eine Anwendung des kontextbezogenen Lernens, bei der Modelle anhand von Beispielen (Aufnahmen) lernen, die in Eingabeaufforderungen eingebettet sind. Bei Aufgaben, die spezifische Formatierungs-, Argumentations- oder Fachkenntnisse erfordern, kann die Eingabeaufforderung mit wenigen Handgriffen effektiv sein. Siehe auch Zero-Shot Prompting.

FGAC

Siehe detaillierte Zugriffskontrolle.

Feinkörnige Zugriffskontrolle (FGAC)

Die Verwendung mehrerer Bedingungen, um eine Zugriffsanfrage zuzulassen oder abzulehnen.

Flash-Cut-Migration

Eine Datenbankmigrationsmethode, bei der eine kontinuierliche Datenreplikation durch Erfassung von Änderungsdaten verwendet wird, um Daten in kürzester Zeit zu migrieren, anstatt einen schrittweisen Ansatz zu verwenden. Ziel ist es, Ausfallzeiten auf ein Minimum zu beschränken.

FM

Siehe Fundamentmodell.

Fundamentmodell (FM)

Ein großes neuronales Deep-Learning-Netzwerk, das mit riesigen Datensätzen generalisierter und unbeschrifteter Daten trainiert wurde. FMs sind in der Lage, eine Vielzahl allgemeiner Aufgaben zu erfüllen, z. B. Sprache zu verstehen, Text und Bilder zu generieren und Konversationen in natürlicher Sprache zu führen. Weitere Informationen finden Sie unter Was sind Foundation-Modelle.

# G

generative KI

Eine Untergruppe von KI-Modellen, die mit großen Datenmengen trainiert wurden und mit einer einfachen Textaufforderung neue Inhalte und Artefakte wie Bilder, Videos, Text und Audio erstellen können. Weitere Informationen finden Sie unter Was ist Generative KI.

Geoblocking

Siehe geografische Einschränkungen.

Geografische Einschränkungen (Geoblocking)

Bei Amazon eine Option CloudFront, um zu verhindern, dass Benutzer in bestimmten Ländern auf Inhaltsverteilungen zugreifen. Sie können eine Zulassungsliste oder eine Sperrliste verwenden, um zugelassene und gesperrte Länder anzugeben. Weitere Informationen finden Sie in der Dokumentation unter Beschränkung der geografischen Verteilung Ihrer Inhalte. CloudFront

Gitflow-Workflow

Ein Ansatz, bei dem niedrigere und höhere Umgebungen unterschiedliche Zweige in einem Quellcode-Repository verwenden. Der Gitflow-Workflow gilt als veraltet, und der Trunk-basierte Workflow ist der moderne, bevorzugte Ansatz.

goldenes Bild

Ein Snapshot eines Systems oder einer Software, der als Vorlage für die Bereitstellung neuer Instanzen dieses Systems oder dieser Software verwendet wird. In der Fertigung kann ein Golden Image beispielsweise zur Bereitstellung von Software auf mehreren Geräten verwendet werden und trägt zur Verbesserung der Geschwindigkeit, Skalierbarkeit und Produktivität bei der Geräteherstellung bei.

Greenfield-Strategie

Das Fehlen vorhandener Infrastruktur in einer neuen Umgebung. Bei der Einführung einer Neuausrichtung einer Systemarchitektur können Sie alle neuen Technologien ohne Einschränkung der Kompatibilität mit der vorhandenen Infrastruktur auswählen, auch bekannt als Brownfield. Wenn Sie die bestehende Infrastruktur erweitern, könnten Sie Brownfield- und Greenfield-Strategien mischen.

Integritätsschutz

Eine allgemeine Regel, die dazu beiträgt, Ressourcen, Richtlinien und die Einhaltung von Vorschriften in allen Unternehmenseinheiten zu regeln (OUs). Präventiver Integritätsschutz setzt Richtlinien durch, um die Einhaltung von Standards zu gewährleisten. Sie werden mithilfe von Service-Kontrollrichtlinien und IAM-Berechtigungsgrenzen implementiert. Detektivischer Integritätsschutz erkennt Richtlinienverstöße und Compliance-Probleme und generiert Warnmeldungen zur Abhilfe. Sie werden mithilfe von AWS Config, AWS Security Hub, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector und benutzerdefinierten AWS Lambda Prüfungen implementiert.

# H

HEKTAR

Siehe Hochverfügbarkeit.

Heterogene Datenbankmigration

Migrieren Sie Ihre Quelldatenbank in eine Zieldatenbank, die eine andere Datenbank-Engine verwendet (z. B. Oracle zu Amazon Aurora). Eine heterogene Migration ist in der Regel Teil einer Neuarchitektur, und die Konvertierung des Schemas kann eine komplexe Aufgabe sein. AWS bietet AWS SCT, welches bei Schemakonvertierungen hilft.

hohe Verfügbarkeit (HA)

Die Fähigkeit eines Workloads, im Falle von Herausforderungen oder Katastrophen kontinuierlich und ohne Eingreifen zu arbeiten. HA-Systeme sind so konzipiert, dass sie automatisch ein Failover durchführen, gleichbleibend hohe Leistung bieten und unterschiedliche Lasten und Ausfälle mit minimalen Leistungseinbußen bewältigen.

historische Modernisierung

Ein Ansatz zur Modernisierung und Aufrüstung von Betriebstechnologiesystemen (OT), um den Bedürfnissen der Fertigungsindustrie besser gerecht zu werden. Ein Historian ist eine Art von Datenbank, die verwendet wird, um Daten aus verschiedenen Quellen in einer Fabrik zu sammeln und zu speichern.

Holdout-Daten

Ein Teil historischer, beschrifteter Daten, der aus einem Datensatz zurückgehalten wird, der zum Trainieren eines Modells für maschinelles Lernen verwendet wird. Sie können Holdout-Daten verwenden, um die Modellleistung zu bewerten, indem Sie die Modellvorhersagen mit den Holdout-Daten vergleichen.

Homogene Datenbankmigration

Migrieren Sie Ihre Quelldatenbank zu einer Zieldatenbank, die dieselbe Datenbank-Engine verwendet (z. B. Microsoft SQL Server zu Amazon RDS für SQL Server). Eine homogene Migration ist in der Regel Teil eines Hostwechsels oder eines Plattformwechsels. Sie können native Datenbankserviceprogramme verwenden, um das Schema zu migrieren.

heiße Daten

Daten, auf die häufig zugegriffen wird, z. B. Echtzeitdaten oder aktuelle Translationsdaten. Für diese Daten ist in der Regel eine leistungsstarke Speicherebene oder -klasse erforderlich, um schnelle Abfrageantworten zu ermöglichen.

Hotfix

Eine dringende Lösung für ein kritisches Problem in einer Produktionsumgebung. Aufgrund seiner Dringlichkeit wird ein Hotfix normalerweise außerhalb des typischen DevOps Release-Workflows erstellt.

Hypercare-Phase

Unmittelbar nach dem Cutover, der Zeitraum, in dem ein Migrationsteam die migrierten Anwendungen in der Cloud verwaltet und überwacht, um etwaige Probleme zu beheben. In der Regel dauert dieser Zeitraum 1–4 Tage. Am Ende der Hypercare-Phase überträgt das Migrationsteam in der Regel die Verantwortung für die Anwendungen an das Cloud-Betriebsteam.

# I

IaC

Sehen Sie sich Infrastruktur als Code an.

Identitätsbasierte Richtlinie

Eine Richtlinie, die einem oder mehreren IAM-Prinzipalen zugeordnet ist und deren Berechtigungen innerhalb der AWS Cloud Umgebung definiert.

Leerlaufanwendung

Eine Anwendung mit einer durchschnittlichen CPU- und Arbeitsspeicherauslastung zwischen 5 und 20 Prozent über einen Zeitraum von 90 Tagen. In einem Migrationsprojekt ist es üblich, diese Anwendungen außer Betrieb zu nehmen oder sie On-Premises beizubehalten.

IIoT

Siehe Industrielles Internet der Dinge.

unveränderliche Infrastruktur

Ein Modell, das eine neue Infrastruktur für Produktionsworkloads bereitstellt, anstatt die bestehende Infrastruktur zu aktualisieren, zu patchen oder zu modifizieren. Unveränderliche Infrastrukturen sind von Natur aus konsistenter, zuverlässiger und vorhersehbarer als veränderliche Infrastrukturen. Weitere Informationen finden Sie in der Best Practice Deploy using immutable infrastructure im AWS Well-Architected Framework.

Eingehende (ingress) VPC

In einer Architektur AWS mit mehreren Konten ist dies eine VPC, die Netzwerkverbindungen von außerhalb einer Anwendung akzeptiert, überprüft und weiterleitet. Die AWS Security Reference Architecture empfiehlt, Ihr Netzwerkkonto mit eingehendem und ausgehendem Datenverkehr und Inspektion einzurichten, VPCs um die bidirektionale Schnittstelle zwischen Ihrer Anwendung und dem Internet im weiteren Sinne zu schützen.

Inkrementelle Migration

Eine Cutover-Strategie, bei der Sie Ihre Anwendung in kleinen Teilen migrieren, anstatt eine einziges vollständiges Cutover durchzuführen. Beispielsweise könnten Sie zunächst nur einige Microservices oder Benutzer auf das neue System umstellen. Nachdem Sie sich vergewissert haben, dass alles ordnungsgemäß funktioniert, können Sie weitere Microservices oder Benutzer

schrittweise verschieben, bis Sie Ihr Legacy-System außer Betrieb nehmen können. Diese Strategie reduziert die mit großen Migrationen verbundenen Risiken.

Industrie 4.0

Ein Begriff, der 2016 von Klaus Schwab eingeführt wurde und sich auf die Modernisierung von Fertigungsprozessen durch Fortschritte in den Bereichen Konnektivität, Echtzeitdaten, Automatisierung, Analytik und KI/ML bezieht.

Infrastruktur

Alle Ressourcen und Komponenten, die in der Umgebung einer Anwendung enthalten sind.

Infrastructure as Code (IaC)

Der Prozess der Bereitstellung und Verwaltung der Infrastruktur einer Anwendung mithilfe einer Reihe von Konfigurationsdateien. IaC soll Ihnen helfen, das Infrastrukturmanagement zu zentralisieren, Ressourcen zu standardisieren und schnell zu skalieren, sodass neue Umgebungen wiederholbar, zuverlässig und konsistent sind.

industrielles Internet der Dinge (T) IIo

Einsatz von mit dem Internet verbundenen Sensoren und Geräten in Industriesektoren wie Fertigung, Energie, Automobilindustrie, Gesundheitswesen, Biowissenschaften und Landwirtschaft. Weitere Informationen finden Sie unter Aufbau einer digitalen Transformationsstrategie für das industrielle Internet der Dinge (IIoT).

Inspektions-VPC

In einer Architektur AWS mit mehreren Konten eine zentralisierte VPC, die Inspektionen des Netzwerkverkehrs zwischen VPCs (in demselben oder unterschiedlichen AWS-Regionen), dem Internet und lokalen Netzwerken verwaltet. In der AWS Security Reference Architecture wird empfohlen, Ihr Netzwerkkonto mit eingehendem und ausgehendem Datenverkehr sowie Inspektionen einzurichten, VPCs um die bidirektionale Schnittstelle zwischen Ihrer Anwendung und dem Internet im weiteren Sinne zu schützen.

Internet of Things (IoT)

Das Netzwerk verbundener physischer Objekte mit eingebetteten Sensoren oder Prozessoren, das über das Internet oder über ein lokales Kommunikationsnetzwerk mit anderen Geräten und Systemen kommuniziert. Weitere Informationen finden Sie unter Was ist IoT?

Interpretierbarkeit

Ein Merkmal eines Modells für Machine Learning, das beschreibt, inwieweit ein Mensch verstehen kann, wie die Vorhersagen des Modells von seinen Eingaben abhängen. Weitere Informationen finden Sie unter Interpretierbarkeit von Modellen für maschinelles Lernen mit. AWS

IoT

Siehe Internet der Dinge.

IT information library (ITIL, IT-Informationsbibliothek)

Eine Reihe von bewährten Methoden für die Bereitstellung von IT-Services und die Abstimmung dieser Services auf die Geschäftsanforderungen. ITIL bietet die Grundlage für ITSM.

T service management (ITSM, IT-Servicemanagement)

Aktivitäten im Zusammenhang mit der Gestaltung, Implementierung, Verwaltung und Unterstützung von IT-Services für eine Organisation. Informationen zur Integration von Cloud-Vorgängen mit ITSM-Tools finden Sie im Leitfaden zur Betriebsintegration.

BIS

Weitere Informationen finden Sie in der IT-Informationsbibliothek.

ITSM

Siehe IT-Servicemanagement.

# L

Labelbasierte Zugangskontrolle (LBAC)

Eine Implementierung der Mandatory Access Control (MAC), bei der den Benutzern und den Daten selbst jeweils explizit ein Sicherheitslabelwert zugewiesen wird. Die Schnittmenge zwischen der Benutzersicherheitsbeschriftung und der Datensicherheitsbeschriftung bestimmt, welche Zeilen und Spalten für den Benutzer sichtbar sind.

Landing Zone

Eine landing zone ist eine gut strukturierte AWS Umgebung mit mehreren Konten, die skalierbar und sicher ist. Dies ist ein Ausgangspunkt, von dem aus Ihre Organisationen Workloads und Anwendungen schnell und mit Vertrauen in ihre Sicherheits- und Infrastrukturumgebung starten

und bereitstellen können. Weitere Informationen zu Landing Zones finden Sie unter Einrichtung einer sicheren und skalierbaren AWS -Umgebung mit mehreren Konten..

großes Sprachmodell (LLM)

Ein Deep-Learning-KI-Modell, das anhand einer riesigen Datenmenge vorab trainiert wurde. Ein LLM kann mehrere Aufgaben ausführen, z. B. Fragen beantworten, Dokumente zusammenfassen, Text in andere Sprachen übersetzen und Sätze vervollständigen. Weitere Informationen finden Sie unter Was sind. LLMs

Große Migration

Eine Migration von 300 oder mehr Servern.

SCHWARZ

Weitere Informationen finden Sie unter Label-basierte Zugriffskontrolle.

Geringste Berechtigung

Die bewährte Sicherheitsmethode, bei der nur die für die Durchführung einer Aufgabe erforderlichen Mindestberechtigungen erteilt werden. Weitere Informationen finden Sie unter Geringste Berechtigungen anwenden in der IAM-Dokumentation.

Lift and Shift

Siehe 7 Rs.

Little-Endian-System

Ein System, welches das niedrigwertigste Byte zuerst speichert. Siehe auch Endianness.

LLM

Siehe großes Sprachmodell.

Niedrigere Umgebungen

Siehe Umgebung.

# M

Machine Learning (ML)

Eine Art künstlicher Intelligenz, die Algorithmen und Techniken zur Mustererkennung und zum Lernen verwendet. ML analysiert aufgezeichnete Daten, wie z. B. Daten aus dem Internet der

Dinge (IoT), und lernt daraus, um ein statistisches Modell auf der Grundlage von Mustern zu erstellen. Weitere Informationen finden Sie unter Machine Learning.

Hauptzweig

Siehe Filiale.

Malware

Software, die entwickelt wurde, um die Computersicherheit oder den Datenschutz zu gefährden. Malware kann Computersysteme stören, vertrauliche Informationen durchsickern lassen oder sich unbefugten Zugriff verschaffen. Beispiele für Malware sind Viren, Würmer, Ransomware, Trojaner, Spyware und Keylogger.

verwaltete Dienste

AWS-Services für die die Infrastrukturebene, das Betriebssystem und die Plattformen AWS betrieben werden, und Sie greifen auf die Endgeräte zu, um Daten zu speichern und abzurufen. Amazon Simple Storage Service (Amazon S3) und Amazon DynamoDB sind Beispiele für Managed Services. Diese werden auch als abstrakte Dienste bezeichnet.

Manufacturing Execution System (MES)

Ein Softwaresystem zur Nachverfolgung, Überwachung, Dokumentation und Steuerung von Produktionsprozessen, bei denen Rohstoffe in der Fertigung zu fertigen Produkten umgewandelt werden.

MAP

Siehe Migration Acceleration Program.

Mechanismus

Ein vollständiger Prozess, bei dem Sie ein Tool erstellen, die Akzeptanz des Tools vorantreiben und anschließend die Ergebnisse überprüfen, um Anpassungen vorzunehmen. Ein Mechanismus ist ein Zyklus, der sich im Laufe seiner Tätigkeit selbst verstärkt und verbessert. Weitere Informationen finden Sie unter Aufbau von Mechanismen im AWS Well-Architected Framework.

Mitgliedskonto

Alle AWS-Konten außer dem Verwaltungskonto, die Teil einer Organisation in sind. AWS Organizations Ein Konto kann jeweils nur einer Organisation angehören.

DURCHEINANDER

Siehe Manufacturing Execution System.

Message Queuing-Telemetrietransport (MQTT)

Ein leichtes machine-to-machine (M2M) -Kommunikationsprotokoll, das auf dem Publish/Subscribe-Muster für IoT-Geräte mit beschränkten Ressourcen basiert.

Microservice

Ein kleiner, unabhängiger Dienst, der über genau definierte Kanäle kommuniziert APIs und in der Regel kleinen, eigenständigen Teams gehört. Ein Versicherungssystem kann beispielsweise Microservices beinhalten, die Geschäftsfunktionen wie Vertrieb oder Marketing oder Subdomains wie Einkauf, Schadenersatz oder Analytik zugeordnet sind. Zu den Vorteilen von Microservices gehören Agilität, flexible Skalierung, einfache Bereitstellung, wiederverwendbarer Code und Ausfallsicherheit. Weitere Informationen finden Sie unter Integration von Microservices mithilfe serverloser Dienste. AWS

Microservices-Architekturen

Ein Ansatz zur Erstellung einer Anwendung mit unabhängigen Komponenten, die jeden Anwendungsprozess als Microservice ausführen. Diese Microservices kommunizieren mithilfe von Lightweight über eine klar definierte Schnittstelle. APIs Jeder Microservice in dieser Architektur kann aktualisiert, bereitgestellt und skaliert werden, um den Bedarf an bestimmten Funktionen einer Anwendung zu decken. Weitere Informationen finden Sie unter Implementierung von Microservices auf. AWS

Migration Acceleration Program (MAP)

Ein AWS Programm, das Beratung, Unterstützung, Schulungen und Services bietet, um Unternehmen dabei zu unterstützen, eine solide betriebliche Grundlage für die Umstellung auf die Cloud zu schaffen und die anfänglichen Kosten von Migrationen auszugleichen. MAP umfasst eine Migrationsmethode für die methodische Durchführung von Legacy-Migrationen sowie eine Reihe von Tools zur Automatisierung und Beschleunigung gängiger Migrationsszenarien.

Migration in großem Maßstab

Der Prozess, bei dem der Großteil des Anwendungsportfolios in Wellen in die Cloud verlagert wird, wobei in jeder Welle mehr Anwendungen schneller migriert werden. In dieser Phase werden die bewährten Verfahren und Erkenntnisse aus den früheren Phasen zur Implementierung einer Migrationsfabrik von Teams, Tools und Prozessen zur Optimierung der Migration von Workloads durch Automatisierung und agile Bereitstellung verwendet. Dies ist die dritte Phase der AWS -Migrationsstrategie.

Migrationsfabrik

Funktionsübergreifende Teams, die die Migration von Workloads durch automatisierte, agile Ansätze optimieren. Zu den Teams in der Migrationsabteilung gehören in der Regel Betriebsabläufe, Geschäftsanalysten und Eigentümer, Migrationsingenieure, Entwickler und DevOps Experten, die in Sprints arbeiten. Zwischen 20 und 50 Prozent eines Unternehmensanwendungsportfolios bestehen aus sich wiederholenden Mustern, die durch einen Fabrik-Ansatz optimiert werden können. Weitere Informationen finden Sie in Diskussion über Migrationsfabriken und den Leitfaden zur Cloud-Migration-Fabrik in diesem Inhaltssatz.

Migrationsmetadaten

Die Informationen über die Anwendung und den Server, die für den Abschluss der Migration benötigt werden. Für jedes Migrationsmuster ist ein anderer Satz von Migrationsmetadaten erforderlich. Beispiele für Migrationsmetadaten sind das Zielsubnetz, die Sicherheitsgruppe und AWS das Konto.

Migrationsmuster

Eine wiederholbare Migrationsaufgabe, in der die Migrationsstrategie, das Migrationsziel und die verwendete Migrationsanwendung oder der verwendete Migrationsservice detailliert beschrieben werden. Beispiel: Rehost-Migration zu Amazon EC2 mit AWS Application Migration Service.

Migration Portfolio Assessment (MPA)

Ein Online-Tool, das Informationen zur Validierung des Geschäftsszenarios für die Migration auf das bereitstellt. AWS Cloud MPA bietet eine detaillierte Portfoliobewertung (richtige Servergröße, Preisgestaltung, Gesamtbetriebskostenanalyse, Migrationskostenanalyse) sowie Migrationsplanung (Anwendungsdatenanalyse und Datenerfassung, Anwendungsgruppierung, Migrationspriorisierung und Wellenplanung). Das MPA-Tool (Anmeldung erforderlich) steht allen AWS Beratern und APN-Partnerberatern kostenlos zur Verfügung.

Migration Readiness Assessment (MRA)

Der Prozess, bei dem mithilfe des AWS CAF Erkenntnisse über den Cloud-Bereitschaftsstatus eines Unternehmens gewonnen, Stärken und Schwächen identifiziert und ein Aktionsplan zur Schließung festgestellter Lücken erstellt wird. Weitere Informationen finden Sie im Benutzerhandbuch für Migration Readiness. MRA ist die erste Phase der AWS - Migrationsstrategie.

Migrationsstrategie

Der Ansatz, der verwendet wurde, um einen Workload auf den AWS Cloud zu migrieren. Weitere Informationen finden Sie im Eintrag 7 Rs in diesem Glossar und unter Mobilisieren Sie Ihr Unternehmen, um umfangreiche Migrationen zu beschleunigen.

ML

Siehe maschinelles Lernen.

Modernisierung

Umwandlung einer veralteten (veralteten oder monolithischen) Anwendung und ihrer Infrastruktur in ein agiles, elastisches und hochverfügbares System in der Cloud, um Kosten zu senken, die Effizienz zu steigern und Innovationen zu nutzen. Weitere Informationen finden Sie unter Strategie zur Modernisierung von Anwendungen in der AWS Cloud.

Bewertung der Modernisierungsfähigkeit

Eine Bewertung, anhand derer festgestellt werden kann, ob die Anwendungen einer Organisation für die Modernisierung bereit sind, Vorteile, Risiken und Abhängigkeiten identifiziert und ermittelt wird, wie gut die Organisation den zukünftigen Status dieser Anwendungen unterstützen kann. Das Ergebnis der Bewertung ist eine Vorlage der Zielarchitektur, eine Roadmap, in der die Entwicklungsphasen und Meilensteine des Modernisierungsprozesses detailliert beschrieben werden, sowie ein Aktionsplan zur Behebung festgestellter Lücken. Weitere Informationen finden Sie unter Evaluierung der Modernisierungsbereitschaft von Anwendungen in der AWS Cloud.

Monolithische Anwendungen (Monolithen)

Anwendungen, die als ein einziger Service mit eng gekoppelten Prozessen ausgeführt werden. Monolithische Anwendungen haben verschiedene Nachteile. Wenn ein Anwendungs-Feature stark nachgefragt wird, muss die gesamte Architektur skaliert werden. Das Hinzufügen oder Verbessern der Feature einer monolithischen Anwendung wird ebenfalls komplexer, wenn die Codebasis wächst. Um diese Probleme zu beheben, können Sie eine Microservices-Architektur verwenden. Weitere Informationen finden Sie unter Zerlegen von Monolithen in Microservices.

MPA

Siehe Bewertung des Migrationsportfolios.

MQTT

Siehe Message Queuing-Telemetrietransport.

Mehrklassen-Klassifizierung

Ein Prozess, der dabei hilft, Vorhersagen für mehrere Klassen zu generieren (wobei eines von mehr als zwei Ergebnissen vorhergesagt wird). Ein ML-Modell könnte beispielsweise fragen: „Ist dieses Produkt ein Buch, ein Auto oder ein Telefon?“ oder „Welche Kategorie von Produkten ist für diesen Kunden am interessantesten?“

veränderbare Infrastruktur

Ein Modell, das die bestehende Infrastruktur für Produktionsworkloads aktualisiert und modifiziert. Für eine verbesserte Konsistenz, Zuverlässigkeit und Vorhersagbarkeit empfiehlt das AWS Well-Architected Framework die Verwendung einer unveränderlichen Infrastruktur als bewährte Methode.

# O

OAC

Siehe Origin Access Control.

OAI

Siehe Zugriffsidentität von Origin.

COM

Siehe organisatorisches Change-Management.

Offline-Migration

Eine Migrationsmethode, bei der der Quell-Workload während des Migrationsprozesses heruntergefahren wird. Diese Methode ist mit längeren Ausfallzeiten verbunden und wird in der Regel für kleine, unkritische Workloads verwendet.

OI

Siehe Betriebsintegration.

OLA

Siehe Vereinbarung auf operativer Ebene.

Online-Migration

Eine Migrationsmethode, bei der der Quell-Workload auf das Zielsystem kopiert wird, ohne offline genommen zu werden. Anwendungen, die mit dem Workload verbunden sind, können während

der Migration weiterhin funktionieren. Diese Methode beinhaltet keine bis minimale Ausfallzeit und wird in der Regel für kritische Produktionsworkloads verwendet.

OPC-UA

Siehe Open Process Communications — Unified Architecture.

Offene Prozesskommunikation — Einheitliche Architektur (OPC-UA)

Ein machine-to-machine (M2M) -Kommunikationsprotokoll für die industrielle Automatisierung. OPC-UA bietet einen Interoperabilitätsstandard mit Datenverschlüsselungs-, Authentifizierungs- und Autorisierungsschemata.

Vereinbarung auf Betriebsebene (OLA)

Eine Vereinbarung, in der klargestellt wird, welche funktionalen IT-Gruppen sich gegenseitig versprechen zu liefern, um ein Service Level Agreement (SLA) zu unterstützen.

Überprüfung der Betriebsbereitschaft (ORR)

Eine Checkliste mit Fragen und zugehörigen bewährten Methoden, die Ihnen helfen, Vorfälle und mögliche Ausfälle zu verstehen, zu bewerten, zu verhindern oder deren Umfang zu reduzieren. Weitere Informationen finden Sie unter Operational Readiness Reviews (ORR) im AWS Well-Architected Framework.

Betriebstechnologie (OT)

Hardware- und Softwaresysteme, die mit der physischen Umgebung zusammenarbeiten, um industrielle Abläufe, Ausrüstung und Infrastruktur zu steuern. In der Fertigung ist die Integration von OT- und Informationstechnologie (IT) -Systemen ein zentraler Schwerpunkt der Industrie 4.0-Transformationen.

Betriebsintegration (OI)

Der Prozess der Modernisierung von Abläufen in der Cloud, der Bereitschaftsplanung, Automatisierung und Integration umfasst. Weitere Informationen finden Sie im Leitfaden zur Betriebsintegration.

Organisationspfad

Ein Pfad, der von erstellt wird und in AWS CloudTrail dem alle Ereignisse für alle AWS-Konten in einer Organisation protokolliert werden. AWS Organizations Diese Spur wird in jedem AWS-Konto , der Teil der Organisation ist, erstellt und verfolgt die Aktivität in jedem Konto. Weitere Informationen finden Sie in der CloudTrail Dokumentation unter Einen Trail für eine Organisation erstellen.

Organisatorisches Veränderungsmanagement (OCM)

Ein Framework für das Management wichtiger, disruptiver Geschäftstransformationen aus Sicht der Mitarbeiter, der Kultur und der Führung. OCM hilft Organisationen dabei, sich auf neue Systeme und Strategien vorzubereiten und auf diese umzustellen, indem es die Akzeptanz von Veränderungen beschleunigt, Übergangsprobleme angeht und kulturelle und organisatorische Veränderungen vorantreibt. In der AWS Migrationsstrategie wird dieses Framework aufgrund der Geschwindigkeit des Wandels, der bei Projekten zur Cloud-Einführung erforderlich ist, als Mitarbeiterbeschleunigung bezeichnet. Weitere Informationen finden Sie im OCM-Handbuch.

Ursprungszugriffskontrolle (OAC)

In CloudFront, eine erweiterte Option zur Zugriffsbeschränkung, um Ihre Amazon Simple Storage Service (Amazon S3) -Inhalte zu sichern. OAC unterstützt alle S3-Buckets insgesamt AWS-Regionen, serverseitige Verschlüsselung mit AWS KMS (SSE-KMS) sowie dynamische `PUT` und `DELETE` Anfragen an den S3-Bucket.

Ursprungszugriffsidentität (OAI)

In CloudFront, eine Option zur Zugriffsbeschränkung, um Ihre Amazon S3 S3-Inhalte zu sichern. Wenn Sie OAI verwenden, CloudFront erstellt es einen Principal, mit dem sich Amazon S3 authentifizieren kann. Authentifizierte Principals können nur über eine bestimmte Distribution auf Inhalte in einem S3-Bucket zugreifen. CloudFront Siehe auch OAC, das eine detailliertere und verbesserte Zugriffskontrolle bietet.

ORR

Weitere Informationen finden Sie unter Überprüfung der Betriebsbereitschaft.

NICHT

Siehe Betriebstechnologie.

Ausgehende (egress) VPC

In einer Architektur AWS mit mehreren Konten eine VPC, die Netzwerkverbindungen verarbeitet, die von einer Anwendung aus initiiert werden. Die AWS Security Reference Architecture empfiehlt die Einrichtung Ihres Netzwerkkontos mit eingehendem und ausgehendem Datenverkehr sowie Inspektion, VPCs um die bidirektionale Schnittstelle zwischen Ihrer Anwendung und dem Internet im weiteren Sinne zu schützen.

# P

Berechtigungsgrenze

Eine IAM-Verwaltungsrichtlinie, die den IAM-Prinzipalen zugeordnet ist, um die maximalen Berechtigungen festzulegen, die der Benutzer oder die Rolle haben kann. Weitere Informationen finden Sie unter Berechtigungsgrenzen für IAM-Entitys in der IAM-Dokumentation.

persönlich identifizierbare Informationen (PII)

Informationen, die, wenn sie direkt betrachtet oder mit anderen verwandten Daten kombiniert werden, verwendet werden können, um vernünftige Rückschlüsse auf die Identität einer Person zu ziehen. Beispiele für personenbezogene Daten sind Namen, Adressen und Kontaktinformationen.

Personenbezogene Daten

Siehe persönlich identifizierbare Informationen.

Playbook

Eine Reihe vordefinierter Schritte, die die mit Migrationen verbundenen Aufgaben erfassen, z. B. die Bereitstellung zentraler Betriebsfunktionen in der Cloud. Ein Playbook kann die Form von Skripten, automatisierten Runbooks oder einer Zusammenfassung der Prozesse oder Schritte annehmen, die für den Betrieb Ihrer modernisierten Umgebung erforderlich sind.

PLC

Siehe programmierbare Logiksteuerung.

PLM

Siehe Produktlebenszyklusmanagement.

policy

Ein Objekt, das Berechtigungen definieren (siehe identitätsbasierte Richtlinie), Zugriffsbedingungen spezifizieren (siehe ressourcenbasierte Richtlinie) oder die maximalen Berechtigungen für alle Konten in einer Organisation definieren kann AWS Organizations (siehe Dienststeuerungsrichtlinie).

Polyglotte Beharrlichkeit

Unabhängige Auswahl der Datenspeichertechnologie eines Microservices auf der Grundlage von Datenzugriffsmustern und anderen Anforderungen. Wenn Ihre Microservices über dieselbe

Datenspeichertechnologie verfügen, kann dies zu Implementierungsproblemen oder zu Leistungseinbußen führen. Microservices lassen sich leichter implementieren und erzielen eine bessere Leistung und Skalierbarkeit, wenn sie den Datenspeicher verwenden, der ihren Anforderungen am besten entspricht. Weitere Informationen finden Sie unter Datenpersistenz in Microservices aktivieren.

Portfoliobewertung

Ein Prozess, bei dem das Anwendungsportfolio ermittelt, analysiert und priorisiert wird, um die Migration zu planen. Weitere Informationen finden Sie in Bewerten der Migrationsbereitschaft.

predicate

Eine Abfragebedingung, die `true` oder zurückgibt`false`, was üblicherweise in einer Klausel vorkommt. `WHERE`

Prädikat Pushdown

Eine Technik zur Optimierung von Datenbankabfragen, bei der die Daten in der Abfrage vor der Übertragung gefiltert werden. Dadurch wird die Datenmenge reduziert, die aus der relationalen Datenbank abgerufen und verarbeitet werden muss, und die Abfrageleistung wird verbessert.

Präventive Kontrolle

Eine Sicherheitskontrolle, die verhindern soll, dass ein Ereignis eintritt. Diese Kontrollen stellen eine erste Verteidigungslinie dar, um unbefugten Zugriff oder unerwünschte Änderungen an Ihrem Netzwerk zu verhindern. Weitere Informationen finden Sie unter Präventive Kontrolle in Implementierung von Sicherheitskontrollen in AWS.

Prinzipal

Eine Entität AWS , die Aktionen ausführen und auf Ressourcen zugreifen kann. Bei dieser Entität handelt es sich in der Regel um einen Root-Benutzer für eine AWS-Konto, eine IAM-Rolle oder einen Benutzer. Weitere Informationen finden Sie unter Prinzipal in Rollenbegriffe und -konzepte in der IAM-Dokumentation.

Datenschutz von Natur aus

Ein systemtechnischer Ansatz, der den Datenschutz während des gesamten Entwicklungsprozesses berücksichtigt.

Privat gehostete Zonen

Ein Container, der Informationen darüber enthält, wie Amazon Route 53 auf DNS-Abfragen für eine Domain und deren Subdomains innerhalb einer oder mehrerer VPCs Domains antworten

soll. Weitere Informationen finden Sie unter Arbeiten mit privat gehosteten Zonen in der Route-53-Dokumentation.

proaktive Steuerung

Eine Sicherheitskontrolle, die den Einsatz nicht richtlinienkonformer Ressourcen verhindern soll. Diese Steuerelemente scannen Ressourcen, bevor sie bereitgestellt werden. Wenn die Ressource nicht mit der Steuerung konform ist, wird sie nicht bereitgestellt. Weitere Informationen finden Sie im Referenzhandbuch zu Kontrollen in der AWS Control Tower Dokumentation und unter Proaktive Kontrollen unter Implementierung von Sicherheitskontrollen am AWS.

Produktlebenszyklusmanagement (PLM)

Das Management von Daten und Prozessen für ein Produkt während seines gesamten Lebenszyklus, vom Design, der Entwicklung und Markteinführung über Wachstum und Reife bis hin zur Markteinführung und Markteinführung.

Produktionsumgebung

Siehe Umgebung.

Speicherprogrammierbare Steuerung (SPS)

In der Fertigung ein äußerst zuverlässiger, anpassungsfähiger Computer, der Maschinen überwacht und Fertigungsprozesse automatisiert.

schnelle Verkettung

Verwendung der Ausgabe einer LLM-Eingabeaufforderung als Eingabe für die nächste Aufforderung, um bessere Antworten zu generieren. Diese Technik wird verwendet, um eine komplexe Aufgabe in Unteraufgaben zu unterteilen oder um eine vorläufige Antwort iterativ zu verfeinern oder zu erweitern. Sie trägt dazu bei, die Genauigkeit und Relevanz der Antworten eines Modells zu verbessern und ermöglicht detailliertere, personalisierte Ergebnisse.

Pseudonymisierung

Der Prozess, bei dem persönliche Identifikatoren in einem Datensatz durch Platzhalterwerte ersetzt werden. Pseudonymisierung kann zum Schutz der Privatsphäre beitragen. Pseudonymisierte Daten gelten weiterhin als personenbezogene Daten.

publish/subscribe (pub/sub)

Ein Muster, das asynchrone Kommunikation zwischen Microservices ermöglicht, um die Skalierbarkeit und Reaktionsfähigkeit zu verbessern. In einem auf Microservices basierenden MES kann ein Microservice beispielsweise Ereignismeldungen in einem Kanal veröffentlichen,

den andere Microservices abonnieren können. Das System kann neue Microservices hinzufügen, ohne den Veröffentlichungsservice zu ändern.

## Q

Abfrageplan

Eine Reihe von Schritten, wie Anweisungen, die für den Zugriff auf die Daten in einem relationalen SQL-Datenbanksystem verwendet werden.

Abfrageplanregression

Wenn ein Datenbankserviceoptimierer einen weniger optimalen Plan wählt als vor einer bestimmten Änderung der Datenbankumgebung. Dies kann durch Änderungen an Statistiken, Beschränkungen, Umgebungseinstellungen, Abfrageparameter-Bindungen und Aktualisierungen der Datenbank-Engine verursacht werden.

## R

RACI-Matrix

Siehe verantwortlich, rechenschaftspflichtig, konsultiert, informiert (RACI).

LAPPEN

Siehe Erweiterte Generierung beim Abrufen.

Ransomware

Eine bösartige Software, die entwickelt wurde, um den Zugriff auf ein Computersystem oder Daten zu blockieren, bis eine Zahlung erfolgt ist.

RASCI-Matrix

Siehe verantwortlich, rechenschaftspflichtig, konsultiert, informiert (RACI).

RCAC

Siehe Zugriffskontrolle für Zeilen und Spalten.

Read Replica

Eine Kopie einer Datenbank, die nur für Lesezwecke verwendet wird. Sie können Abfragen an das Lesereplikat weiterleiten, um die Belastung auf Ihrer Primärdatenbank zu reduzieren.

neu strukturieren

Siehe 7 Rs.

Recovery Point Objective (RPO)

Die maximal zulässige Zeitspanne seit dem letzten Datenwiederherstellungspunkt. Damit wird festgelegt, was als akzeptabler Datenverlust zwischen dem letzten Wiederherstellungspunkt und der Serviceunterbrechung gilt.

Wiederherstellungszeitziel (RTO)

Die maximal zulässige Verzögerung zwischen der Betriebsunterbrechung und der Wiederherstellung des Dienstes.

Refaktorierung

Siehe 7 Rs.

Region

Eine Sammlung von AWS Ressourcen in einem geografischen Gebiet. Jeder AWS-Region ist isoliert und unabhängig von den anderen, um Fehlertoleranz, Stabilität und Belastbarkeit zu gewährleisten. Weitere Informationen finden Sie unter Geben Sie an, was AWS-Regionen Ihr Konto verwenden kann.

Regression

Eine ML-Technik, die einen numerischen Wert vorhersagt. Zum Beispiel, um das Problem „Zu welchem Preis wird dieses Haus verkauft werden?“ zu lösen Ein ML-Modell könnte ein lineares Regressionsmodell verwenden, um den Verkaufspreis eines Hauses auf der Grundlage bekannter Fakten über das Haus (z. B. die Quadratmeterzahl) vorherzusagen.

rehosten

Siehe 7 Rs.

Veröffentlichung

In einem Bereitstellungsprozess der Akt der Förderung von Änderungen an einer Produktionsumgebung.

umziehen

Siehe 7 Rs.

neue Plattform

Siehe 7 Rs.

Rückkauf

Siehe 7 Rs.

Ausfallsicherheit

Die Fähigkeit einer Anwendung, Störungen zu widerstehen oder sich von ihnen zu erholen. Hochverfügbarkeit und Notfallwiederherstellung sind häufig Überlegungen bei der Planung der Ausfallsicherheit in der. AWS Cloud Weitere Informationen finden Sie unter AWS Cloud Resilienz.

Ressourcenbasierte Richtlinie

Eine mit einer Ressource verknüpfte Richtlinie, z. B. ein Amazon-S3-Bucket, ein Endpunkt oder ein Verschlüsselungsschlüssel. Diese Art von Richtlinie legt fest, welchen Prinzipalen der Zugriff gewährt wird, welche Aktionen unterstützt werden und welche anderen Bedingungen erfüllt sein müssen.

RACI-Matrix (verantwortlich, rechenschaftspflichtig, konsultiert, informiert)

Eine Matrix, die die Rollen und Verantwortlichkeiten aller an Migrationsaktivitäten und Cloud-Operationen beteiligten Parteien definiert. Der Matrixname leitet sich von den in der Matrix definierten Zuständigkeitstypen ab: verantwortlich (R), rechenschaftspflichtig (A), konsultiert (C) und informiert (I). Der Unterstützungstyp (S) ist optional. Wenn Sie Unterstützung einbeziehen, wird die Matrix als RASCI-Matrix bezeichnet, und wenn Sie sie ausschließen, wird sie als RACI-Matrix bezeichnet.

Reaktive Kontrolle

Eine Sicherheitskontrolle, die darauf ausgelegt ist, die Behebung unerwünschter Ereignisse oder Abweichungen von Ihren Sicherheitsstandards voranzutreiben. Weitere Informationen finden Sie unter Reaktive Kontrolle in Implementieren von Sicherheitskontrollen in AWS.

Beibehaltung

Siehe 7 Rs.

zurückziehen

Siehe 7 Rs.

Retrieval Augmented Generation (RAG)

Eine generative KI-Technologie, bei der ein LLM auf eine maßgebliche Datenquelle verweist, die sich außerhalb seiner Trainingsdatenquellen befindet, bevor eine Antwort generiert wird. Ein RAG-Modell könnte beispielsweise eine semantische Suche in der Wissensdatenbank oder in benutzerdefinierten Daten einer Organisation durchführen. Weitere Informationen finden Sie unter Was ist RAG.

Drehung

Der Vorgang, bei dem ein Geheimnis regelmäßig aktualisiert wird, um es einem Angreifer zu erschweren, auf die Anmeldeinformationen zuzugreifen.

Zugriffskontrolle für Zeilen und Spalten (RCAC)

Die Verwendung einfacher, flexibler SQL-Ausdrücke mit definierten Zugriffsregeln. RCAC besteht aus Zeilenberechtigungen und Spaltenmasken.

RPO

Siehe Recovery Point Objective.

RTO

Siehe Ziel der Wiederherstellungszeit.

Runbook

Eine Reihe manueller oder automatisierter Verfahren, die zur Ausführung einer bestimmten Aufgabe erforderlich sind. Diese sind in der Regel darauf ausgelegt, sich wiederholende Operationen oder Verfahren mit hohen Fehlerquoten zu rationalisieren.

# S

SAML 2.0

Ein offener Standard, den viele Identitätsanbieter (IdPs) verwenden. Diese Funktion ermöglicht föderiertes Single Sign-On (SSO), sodass sich Benutzer bei den API-Vorgängen anmelden AWS Management Console oder die AWS API-Operationen aufrufen können, ohne dass Sie einen Benutzer in IAM für alle in Ihrer Organisation erstellen müssen. Weitere Informationen zum SAML-2.0.-basierten Verbund finden Sie unter Über den SAML-2.0-basierten Verbund in der IAM-Dokumentation.

SCADA

Siehe Aufsichtskontrolle und Datenerfassung.

SCP

Siehe Richtlinie zur Dienstkontrolle.

Secret

Interne AWS Secrets Manager, vertrauliche oder eingeschränkte Informationen, wie z. B. ein Passwort oder Benutzeranmeldedaten, die Sie in verschlüsselter Form speichern. Es besteht aus dem geheimen Wert und seinen Metadaten. Der geheime Wert kann binär, eine einzelne Zeichenfolge oder mehrere Zeichenketten sein. Weitere Informationen finden Sie unter Was ist in einem Secrets Manager Manager-Geheimnis? in der Secrets Manager Manager-Dokumentation.

Sicherheit durch Design

Ein systemtechnischer Ansatz, der die Sicherheit während des gesamten Entwicklungsprozesses berücksichtigt.

Sicherheitskontrolle

Ein technischer oder administrativer Integritätsschutz, der die Fähigkeit eines Bedrohungsakteurs, eine Schwachstelle auszunutzen, verhindert, erkennt oder einschränkt. Es gibt vier Haupttypen von Sicherheitskontrollen: präventiv, detektiv, reaktionsschnell und proaktiv.

Härtung der Sicherheit

Der Prozess, bei dem die Angriffsfläche reduziert wird, um sie widerstandsfähiger gegen Angriffe zu machen. Dies kann Aktionen wie das Entfernen von Ressourcen, die nicht mehr benötigt werden, die Implementierung der bewährten Sicherheitsmethode der Gewährung geringster Berechtigungen oder die Deaktivierung unnötiger Feature in Konfigurationsdateien umfassen.

System zur Verwaltung von Sicherheitsinformationen und Ereignissen (security information and event management – SIEM)

Tools und Services, die Systeme für das Sicherheitsinformationsmanagement (SIM) und das Management von Sicherheitsereignissen (SEM) kombinieren. Ein SIEM-System sammelt, überwacht und analysiert Daten von Servern, Netzwerken, Geräten und anderen Quellen, um Bedrohungen und Sicherheitsverletzungen zu erkennen und Warnmeldungen zu generieren.

Automatisierung von Sicherheitsreaktionen

Eine vordefinierte und programmierte Aktion, die darauf ausgelegt ist, automatisch auf ein Sicherheitsereignis zu reagieren oder es zu beheben. Diese Automatisierungen dienen als

detektive oder reaktionsschnelle Sicherheitskontrollen, die Sie bei der Implementierung bewährter AWS Sicherheitsmethoden unterstützen. Beispiele für automatisierte Antwortaktionen sind das Ändern einer VPC-Sicherheitsgruppe, das Patchen einer EC2 Amazon-Instance oder das Rotieren von Anmeldeinformationen.

Serverseitige Verschlüsselung

Verschlüsselung von Daten am Zielort durch denjenigen AWS-Service , der sie empfängt.

Service-Kontrollrichtlinie (SCP)

Eine Richtlinie, die eine zentrale Steuerung der Berechtigungen für alle Konten in einer Organisation in ermöglicht AWS Organizations. SCPs Definieren Sie Leitplanken oder legen Sie Grenzwerte für Aktionen fest, die ein Administrator an Benutzer oder Rollen delegieren kann. Sie können sie SCPs als Zulassungs- oder Ablehnungslisten verwenden, um festzulegen, welche Dienste oder Aktionen zulässig oder verboten sind. Weitere Informationen finden Sie in der AWS Organizations Dokumentation unter Richtlinien zur Dienststeuerung.

Service-Endpunkt

Die URL des Einstiegspunkts für einen AWS-Service. Sie können den Endpunkt verwenden, um programmgesteuert eine Verbindung zum Zielservice herzustellen. Weitere Informationen finden Sie unter AWS-Service -Endpunkte in der Allgemeine AWS-Referenz.

Service Level Agreement (SLA)

Eine Vereinbarung, in der klargestellt wird, was ein IT-Team seinen Kunden zu bieten verspricht, z. B. in Bezug auf Verfügbarkeit und Leistung der Services.

Service-Level-Indikator (SLI)

Eine Messung eines Leistungsaspekts eines Dienstes, z. B. seiner Fehlerrate, Verfügbarkeit oder Durchsatz.

Service-Level-Ziel (SLO)

Eine Zielkennzahl, die den Zustand eines Dienstes darstellt, gemessen anhand eines Service-Level-Indikators.

Modell der geteilten Verantwortung

Ein Modell, das die Verantwortung beschreibt, mit der Sie gemeinsam AWS für Cloud-Sicherheit und Compliance verantwortlich sind. AWS ist für die Sicherheit der Cloud verantwortlich, während Sie für die Sicherheit in der Cloud verantwortlich sind. Weitere Informationen finden Sie unter Modell der geteilten Verantwortung.

SIEM

Siehe Sicherheitsinformations- und Event-Management-System.

Single Point of Failure (SPOF)

Ein Fehler in einer einzelnen, kritischen Komponente einer Anwendung, der das System stören kann.

SLA

Siehe Service Level Agreement.

SLI

Siehe Service-Level-Indikator.

ALSO

Siehe Service-Level-Ziel.

split-and-seed Modell

Ein Muster für die Skalierung und Beschleunigung von Modernisierungsprojekten. Sobald neue Features und Produktversionen definiert werden, teilt sich das Kernteam auf, um neue Produktteams zu bilden. Dies trägt zur Skalierung der Fähigkeiten und Services Ihrer Organisation bei, verbessert die Produktivität der Entwickler und unterstützt schnelle Innovationen. Weitere Informationen finden Sie unter Schrittweiser Ansatz zur Modernisierung von Anwendungen in der AWS Cloud.

SPOTTEN

Siehe Single Point of Failure.

Sternschema

Eine Datenbank-Organisationsstruktur, die eine große Faktentabelle zum Speichern von Transaktions- oder Messdaten und eine oder mehrere kleinere dimensionale Tabellen zum Speichern von Datenattributen verwendet. Diese Struktur ist für die Verwendung in einem Data Warehouse oder für Business Intelligence-Zwecke konzipiert.

Strangler-Fig-Muster

Ein Ansatz zur Modernisierung monolithischer Systeme, bei dem die Systemfunktionen schrittweise umgeschrieben und ersetzt werden, bis das Legacy-System außer Betrieb

genommen werden kann. Dieses Muster verwendet die Analogie einer Feigenrebe, die zu einem etablierten Baum heranwächst und schließlich ihren Wirt überwindet und ersetzt. Das Muster wurde eingeführt von Martin Fowler als Möglichkeit, Risiken beim Umschreiben monolithischer Systeme zu managen. Ein Beispiel für die Anwendung dieses Musters finden Sie unter Schrittweises Modernisieren älterer Microsoft ASP.NET (ASMX)-Webservices mithilfe von Containern und Amazon API Gateway.

Subnetz

Ein Bereich von IP-Adressen in Ihrer VPC. Ein Subnetz muss sich in einer einzigen Availability Zone befinden.

Aufsichtskontrolle und Datenerfassung (SCADA)

In der Fertigung ein System, das Hardware und Software zur Überwachung von Sachanlagen und Produktionsabläufen verwendet.

Symmetrische Verschlüsselung

Ein Verschlüsselungsalgorithmus, der denselben Schlüssel zum Verschlüsseln und Entschlüsseln der Daten verwendet.

synthetisches Testen

Testen eines Systems auf eine Weise, die Benutzerinteraktionen simuliert, um potenzielle Probleme zu erkennen oder die Leistung zu überwachen. Sie können Amazon CloudWatch Synthetics verwenden, um diese Tests zu erstellen.

Systemaufforderung

Eine Technik, mit der einem LLM Kontext, Anweisungen oder Richtlinien zur Verfügung gestellt werden, um sein Verhalten zu steuern. Systemaufforderungen helfen dabei, den Kontext festzulegen und Regeln für Interaktionen mit Benutzern festzulegen.

# T

tags

Schlüssel-Wert-Paare, die als Metadaten für die Organisation Ihrer Ressourcen dienen. AWS Mit Tags können Sie Ressourcen verwalten, identifizieren, organisieren, suchen und filtern. Weitere Informationen finden Sie unter Markieren Ihrer AWS -Ressourcen.

Zielvariable

Der Wert, den Sie in überwachtem ML vorhersagen möchten. Dies wird auch als Ergebnisvariable bezeichnet. In einer Fertigungsumgebung könnte die Zielvariable beispielsweise ein Produktfehler sein.

Aufgabenliste

Ein Tool, das verwendet wird, um den Fortschritt anhand eines Runbooks zu verfolgen. Eine Aufgabenliste enthält eine Übersicht über das Runbook und eine Liste mit allgemeinen Aufgaben, die erledigt werden müssen. Für jede allgemeine Aufgabe werden der geschätzte Zeitaufwand, der Eigentümer und der Fortschritt angegeben.

Testumgebungen

Siehe Umgebung.

Training

Daten für Ihr ML-Modell bereitstellen, aus denen es lernen kann. Die Trainingsdaten müssen die richtige Antwort enthalten. Der Lernalgorithmus findet Muster in den Trainingsdaten, die die Attribute der Input-Daten dem Ziel (die Antwort, die Sie voraussagen möchten) zuordnen. Es gibt ein ML-Modell aus, das diese Muster erfasst. Sie können dann das ML-Modell verwenden, um Voraussagen für neue Daten zu erhalten, bei denen Sie das Ziel nicht kennen.

Transit-Gateway

Ein Netzwerk-Transit-Hub, über den Sie Ihre Netzwerke VPCs und Ihre lokalen Netzwerke miteinander verbinden können. Weitere Informationen finden Sie in der Dokumentation unter Was ist ein Transit-Gateway. AWS Transit Gateway

Stammbasierter Workflow

Ein Ansatz, bei dem Entwickler Feature lokal in einem Feature-Zweig erstellen und testen und diese Änderungen dann im Hauptzweig zusammenführen. Der Hauptzweig wird dann sequentiell für die Entwicklungs-, Vorproduktions- und Produktionsumgebungen erstellt.

Vertrauenswürdiger Zugriff

Gewährung von Berechtigungen für einen Dienst, den Sie angeben, um Aufgaben in Ihrer Organisation AWS Organizations und in deren Konten in Ihrem Namen auszuführen. Der vertrauenswürdige Service erstellt in jedem Konto eine mit dem Service verknüpfte Rolle, wenn diese Rolle benötigt wird, um Verwaltungsaufgaben für Sie auszuführen. Weitere Informationen

finden Sie in der AWS Organizations Dokumentation unter Verwendung AWS Organizations mit anderen AWS Diensten.

Optimieren

Aspekte Ihres Trainingsprozesses ändern, um die Genauigkeit des ML-Modells zu verbessern. Sie können das ML-Modell z. B. trainieren, indem Sie einen Beschriftungssatz generieren, Beschriftungen hinzufügen und diese Schritte dann mehrmals unter verschiedenen Einstellungen wiederholen, um das Modell zu optimieren.

Zwei-Pizzen-Team

Ein kleines DevOps Team, das Sie mit zwei Pizzen ernähren können. Eine Teamgröße von zwei Pizzen gewährleistet die bestmögliche Gelegenheit zur Zusammenarbeit bei der Softwareentwicklung.

# U

Unsicherheit

Ein Konzept, das sich auf ungenaue, unvollständige oder unbekannte Informationen bezieht, die die Zuverlässigkeit von prädiktiven ML-Modellen untergraben können. Es gibt zwei Arten von Unsicherheit: Epistemische Unsicherheit wird durch begrenzte, unvollständige Daten verursacht, wohingegen aleatorische Unsicherheit durch Rauschen und Randomisierung verursacht wird, die in den Daten liegt. Weitere Informationen finden Sie im Leitfaden Quantifizieren der Unsicherheit in Deep-Learning-Systemen.

undifferenzierte Aufgaben

Diese Arbeit wird auch als Schwerstarbeit bezeichnet. Dabei handelt es sich um Arbeiten, die zwar für die Erstellung und den Betrieb einer Anwendung erforderlich sind, aber dem Endbenutzer keinen direkten Mehrwert bieten oder keinen Wettbewerbsvorteil bieten. Beispiele für undifferenzierte Aufgaben sind Beschaffung, Wartung und Kapazitätsplanung.

höhere Umgebungen

Siehe Umgebung.

# V

Vacuuming

Ein Vorgang zur Datenbankwartung, bei dem die Datenbank nach inkrementellen Aktualisierungen bereinigt wird, um Speicherplatz zurückzugewinnen und die Leistung zu verbessern.

Versionskontrolle

Prozesse und Tools zur Nachverfolgung von Änderungen, z. B. Änderungen am Quellcode in einem Repository.

VPC-Peering

Eine Verbindung zwischen zwei VPCs , die es Ihnen ermöglicht, den Verkehr mithilfe privater IP-Adressen weiterzuleiten. Weitere Informationen finden Sie unter Was ist VPC-Peering? in der Amazon-VPC-Dokumentation.

Schwachstelle

Ein Software- oder Hardwarefehler, der die Sicherheit des Systems beeinträchtigt.

# W

Warmer Cache

Ein Puffer-Cache, der aktuelle, relevante Daten enthält, auf die häufig zugegriffen wird. Die Datenbank-Instance kann aus dem Puffer-Cache lesen, was schneller ist als das Lesen aus dem Hauptspeicher oder von der Festplatte.

warme Daten

Daten, auf die selten zugegriffen wird. Bei der Abfrage dieser Art von Daten sind mäßig langsame Abfragen in der Regel akzeptabel.

Fensterfunktion

Eine SQL-Funktion, die eine Berechnung für eine Gruppe von Zeilen durchführt, die sich in irgendeiner Weise auf den aktuellen Datensatz beziehen. Fensterfunktionen sind nützlich für die Verarbeitung von Aufgaben wie die Berechnung eines gleitenden Durchschnitts oder für den Zugriff auf den Wert von Zeilen auf der Grundlage der relativen Position der aktuellen Zeile.

Workload

Ein Workload ist eine Sammlung von Ressourcen und Code, die einen Unternehmenswert bietet, wie z. B. eine kundenorientierte Anwendung oder ein Backend-Prozess.

Workstream

Funktionsgruppen in einem Migrationsprojekt, die für eine bestimmte Reihe von Aufgaben verantwortlich sind. Jeder Workstream ist unabhängig, unterstützt aber die anderen Workstreams im Projekt. Der Portfolio-Workstream ist beispielsweise für die Priorisierung von Anwendungen, die Wellenplanung und die Erfassung von Migrationsmetadaten verantwortlich. Der Portfolio-Workstream liefert diese Komponenten an den Migrations-Workstream, der dann die Server und Anwendungen migriert.

WURM

Sehen Sie einmal schreiben, viele lesen.

WQF

Siehe AWS Workload-Qualifizierungsrahmen.

einmal schreiben, viele lesen (WORM)

Ein Speichermodell, das Daten ein einziges Mal schreibt und verhindert, dass die Daten gelöscht oder geändert werden. Autorisierte Benutzer können die Daten so oft wie nötig lesen, aber sie können sie nicht ändern. Diese Datenspeicherinfrastruktur gilt als unveränderlich.

# Z

Zero-Day-Exploit

Ein Angriff, in der Regel Malware, der eine Zero-Day-Sicherheitslücke ausnutzt.

Zero-Day-Sicherheitslücke

Ein unfehlbarer Fehler oder eine Sicherheitslücke in einem Produktionssystem. Bedrohungsakteure können diese Art von Sicherheitslücke nutzen, um das System anzugreifen. Entwickler werden aufgrund des Angriffs häufig auf die Sicherheitsanfälligkeit aufmerksam.

Zero-Shot-Aufforderung

Bereitstellung von Anweisungen für die Ausführung einer Aufgabe an einen LLM, jedoch ohne Beispiele (Schnappschüsse), die ihm als Orientierungshilfe dienen könnten. Der LLM muss sein

vortrainiertes Wissen einsetzen, um die Aufgabe zu bewältigen. Die Effektivität von Zero-Shot Prompting hängt von der Komplexität der Aufgabe und der Qualität der Aufforderung ab. Siehe auch Few-Shot-Prompting.

Zombie-Anwendung

Eine Anwendung, deren durchschnittliche CPU- und Arbeitsspeichernutzung unter 5 Prozent liegt. In einem Migrationsprojekt ist es üblich, diese Anwendungen außer Betrieb zu nehmen.

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.