Unable to locate subtitle

AWS Well-Architected Framework

AWS Well-Architected Framework: ***Unable to locate subtitle***

Table of Contents

AWS Well-Architected Framework

Veröffentlichungsdatum: 20. Oktober 2022 (Dokumentversionen)

AWS Well-Architected Framework unterstützt Sie dabei, die Vor- und Nachteile der Entscheidungen nachzuvollziehen, die Sie beim Aufbau von Systemen in AWS treffen. Das Framework hilft Ihnen, architektonische Best Practices für die Entwicklung und den Betrieb zuverlässiger, sicherer, effizienter und kosteneffektiver Systeme in der Cloud zu ermitteln.

Einführung

AWS Well-Architected Framework unterstützt Sie dabei, die Vor- und Nachteile der Entscheidungen nachzuvollziehen, die Sie beim Aufbau von Systemen in AWS treffen. Das Framework hilft Ihnen, architektonische Best Practices für den Entwurf und Betrieb zuverlässiger, sicherer, effizienter, kosteneffektiver und nachhaltiger Workloads in der AWS Cloud zu ermitteln. Es bietet Ihnen die Möglichkeit, Ihre Architekturen konsistent hinsichtlich der Einhaltung von Best Practices zu überprüfen und Optimierungsmöglichkeiten zu identifizieren. Die Überprüfung einer Architektur ist kein Audit. Vielmehr ist es eine konstruktive Konversation, in der es um architektonische Entscheidungen geht. Wir sind davon überzeugt, dass architektonisch gute Systeme die Wahrscheinlichkeit des geschäftlichen Erfolgs signifikant beeinflussen.

AWS Solutions Architects entwerfen seit vielen Jahren Architekturen für unterschiedlichste Branchen und Anwendungsfälle. Wir waren am Design und der Überprüfung Tausender Kundenarchitekturen auf AWS beteiligt. Daher kennen wir die bewährten Methoden und Kernstrategien für erfolgreiche Systemarchitekturen in der Cloud.

Das AWS Well-Architected Framework dokumentiert grundlegende Fragen, mit denen Sie klären, ob eine Architektur einwandfrei mit Best Practices für die Cloud vereinbar ist. Über das Framework erhalten Sie eine einheitliche Herangehensweise zur Bewertung der Eigenschaften, die Sie von modernen Cloud-basierten Systemen erwarten, sowie Vorschläge zur Realisierung dieser Eigenschaften. AWS entwickelt sich ständig weiter, und auch wir lernen durch die Arbeit mit unseren Kunden ständig dazu. Und so wie unser Wissen anwächst, können wir immer wieder noch genauer definieren, wodurch sich eine gute architektonische Struktur auszeichnet.

Dieses Framework richtet sich an Technologiefachleute, z. B. Chief Technology Officers (CTO), Architekten, Entwickler und Operations-Mitarbeiter. Die darin enthaltenen Best Practices und Strategien für AWS kommen beim Design und der Nutzung von Cloud Workloads zum Einsatz. Die

Links verweisen auf weitere Implementierungsdetails und Architekturmodelle. Weitere Informationen finden Sie auf der Homepage von AWS Well-Architected-Homepage.

AWS bietet auch an, Ihre Workloads kostenfrei zu überprüfen. Das AWS Well-Architected Tool (AWS WA Tool) ist ein Service in der Cloud, der einen einheitlichen Prozess zum Überprüfen und Messen Ihrer Architektur mit AWS Well-Architected Framework bietet. Vom AWS WA Tool erhalten Sie Empfehlungen, wie Sie Ihre Workloads zuverlässiger, sicherer, effizienter und kostengünstiger machen.

Um Ihnen das Arbeiten nach bewährten Methoden zu erleichtern, haben wir AWS Well-Architected Labsentwickelt. Der Code und die Dokumentation von Labs erlauben Ihnen, eigene Erfahrungen mit der Implementierung bewährter Methoden zu sammeln. Außerdem haben wir uns mit ausgewählten Partner aus dem AWS Partner Network (APN), die Mitglieder des AWS Well-Architected-Partnerprogramms sind, zusammengetan.. Diese AWS-Partner sind bestens mit AWS vertraut und können Sie beim Überprüfen und Verbessern Ihrer Workloads unterstützen.

Definitionen

Die Experten von AWS unterstützen mit bewährten Cloud-Methoden tagtäglich Kunden beim Entwerfen von Systemarchitekturen. Während wir zusammen mit Ihnen die Architektur entwerfen, wägen wir die Anforderungen ab und treffen die richtigen Kompromisse. Wenn Sie die Systeme dann in Live-Umgebungen bereitstellen, beobachten wir, wie gut diese Systeme laufen und welche Auswirkungen die Kompromisse haben.

Unsere bisherigen Erkenntnisse sind die Grundlage von AWS Well-Architected Framework. Das Framework enthält einheitlich zusammengestellte bewährte Methoden, mit denen Kunden und Partner Architekturen bewerten. Anhand verschiedener Fragen können sie beurteilen, wie gut eine Architektur auf die bewährten Methoden von AWS ausgerichtet ist.

Das AWS Well-Architected Framework basiert auf sechs Säulen: betriebliche Exzellenz, Sicherheit, Zuverlässigkeit, Leistungseffizienz, Kostenoptimierung und Nachhaltigkeit.

Tabelle 1. Die Säulen des AWS Well-Architected Framework

Name	Beschreibung
Operational Excellence	Die Fähigkeit, die Entwicklung zu unterstüt zen und Workloads effektiv auszuführen, Einblicke in die Betriebsabläufe zu erhalten

Name	Beschreibung
	und für geschäftlichen Mehrwert unterstüt zende Prozesse und Verfahren fortlaufend zu verbessern.
Sicherheit	In der Säule der Sicherheit wird beschrieben, wie Sie Cloud-Technologien nutzen, um Daten, Systeme und Komponenten so zu schützen, dass sich Ihre Sicherheitslage verbessert.
Zuverlässigkeit	Die Säule „Zuverlässigkeit“ umfasst die Fähigkeit eines Workloads, die beabsicht igte Funktion erwartungsgemäß korrekt und konsistent auszuführen. Dies umfasst die Möglichkeit, den Workload während des gesamten Lebenszyklus zu betreiben und zu testen. Dieses Dokument bietet umfassende Informationen mit bewährte Methoden für die Implementierung zuverlässiger Workloads in AWS.
Leistungseffizienz	Die Fähigkeit, Rechenressourcen effizient entsprechend den Systemanforderungen zu nutzen und diese Effizienz aufrechtzuerhalten , während sich die Nachfrage ändert und die Technologie weiterentwickelt.
Kostenoptimierung	Die Fähigkeit, Systeme so auszuführen, dass sie geschäftlichen Wert bei geringstmöglichen Kosten liefern.

Name	Beschreibung
Nachhaltigkeit	Die Fähigkeit, die Auswirkungen auf die Nachhaltigkeit kontinuierlich zu verbesser n, indem der Energieverbrauch reduziert und die Effizienz aller Komponenten eines Workloads erhöht wird, indem der Nutzen der bereitgestellten Ressourcen maximiert und die insgesamt erforderlichen Ressourcen minimiert werden.

In Zusammenhang mit dem AWS Well-Architected Framework verwenden wir diese Bezeichnungen:

- Eine Komponente besteht aus dem Code, der Konfiguration und den AWS-Ressourcen, die für eine Anforderung bereitgestellt werden. Eine Komponente ist häufig die Einheit technischen Eigentums und von anderen Komponenten losgelöst.
- Der Begriff Workload wird für zusammengehörige Komponenten, die geschäftlichen Mehrwert darstellen, verwendet. Ein Workload ist in vielen Fällen der Detaillierungsgrad, von dem Führungskräfte aus Wirtschaft und Technik häufig sprechen.
- Wir betrachten Architektur als das Zusammenwirken von Komponenten in einem Workload. Wie Komponenten kommunizieren und interagieren, ist häufig der Schwerpunkt von Architekturdiagrammen.
- Meilensteine markieren wichtige Änderungen im Laufe der Entwicklung einer Architektur im Produktlebenszyklus (Entwurf, Implementierung, Tests, Inbetriebnahme und Produktionsbetrieb).
- Innerhalb einer Organisation ist das Technologieportfolio die für den Geschäftsbetrieb erforderliche Sammlung an Workloads.
- Der Grad des Aufwands bezeichnet die Zeitspanne, den Aufwand und die Komplexität, die für die Implementierung einer Aufgabe benötigt werden. Jede Organisation muss die Größe und das Fachwissen des Teams sowie die Komplexität des Workloads berücksichtigen, um den Aufwand für die Organisation richtig einzuordnen.
  - Hoch: Die Arbeit dauert möglicherweise mehrere Wochen oder Monate. Dies könnte in mehrere Geschichten, Veröffentlichungen und Aufgaben aufgeteilt werden.
  - Mittel: Die Arbeit dauert möglicherweise mehrere Tage oder Wochen. Dies könnte in mehrere Veröffentlichungen und Aufgaben aufgeteilt werden.

- Niedrig: Die Arbeit dauert möglicherweise mehrere Stunden oder Tage. Dies könnte in mehrere Aufgaben aufgeteilt werden.

Beim Entwerfen von Workloads stellen Sie eine Kosten-Nutzen-Abwägung zwischen Säulen abhängig von Ihrem Geschäftskontext an. Diese Geschäftsentscheidungen können Ihre technischen Prioritäten beeinflussen. Sie können optimieren, um die Nachhaltigkeitswirkung zu verbessern und die Kosten zulasten der Zuverlässigkeit in Entwicklungsumgebungen zu senken. Sie können bei unternehmenskritischen Lösungen auch die Zuverlässigkeit mit höheren Kosten und höherer Nachhaltigkeitswirkung optimieren. Bei E-Commerce-Lösungen kann sich die Leistung auf die Einnahmen und die Kauflust der Kunden auswirken. Sicherheit und betriebliche Exzellenz haben in der Regel keine Wechselwirkung mit den anderen Säulen.

Architektur-Überlegungen

In On-Premise-Umgebungen setzen Kunden oft ein Zentralteam für Technologiearchitektur ein. Dieses ist anderen Produkt- oder Feature-Teams vorgeschaltet, damit diese nach bewährten Methoden arbeiten. Technologiearchitektur-Teams setzen sich üblicherweise aus Fachleuten mit unterschiedlichen Aufgabengebieten zusammen, z. B.: Technical Architect (Infrastruktur), Solutions Architect (Software), Data Architect, Networking Architect und Security Architect. Diese Teams arbeiten oft nach dem TOGAF-Modell oder dem Zachman Framework – als Teil eines Kompetenzbereichs für Enterprise-Architektur.

AWS verteilt Fähigkeiten lieber auf einzelne Teams, anstatt die Kompetenz in einem Zentralteam zu konzentrieren. Wenn die Entscheidungsbefugnis auf mehrere Teams verteilt wird, geht das mit Risiken einher. So muss beispielsweise sichergestellt sein, dass die Teams internen Standards gerecht werden. Um diese Risiken aufzufangen, verwenden wir zwei Methoden. Zum einen arbeiten wir mit Praktiken (Vorgehensweisen, Prozessen, Standards und gemeinhin anerkannte Normen), die darauf abzielen, jedes Team mit dieser Fähigkeit auszustatten. Dazu setzen wir Experten ein, die dafür sorgen, dass die Teams die vorgegebenen Standards übertreffen. Zweitens implementieren wir Mechanismen , die automatisch kontrollieren, ob Standards eingehalten werden.

> „Gut gemeinte Absichten funktionieren nicht. Wer etwas erreichen will, braucht gute Mechanismen“ – Jeff Bezos.

Das bedeutet konkret, dass wir das Bestmögliche, das Menschen leisten können, durch (automatisierte) Mechanismen ersetzen, die kontrollieren, ob Regeln oder Prozesse eingehalten werden. Hinter diesem breit aufgestellten Ansatz stehen die Führungsprinzipien von Amazon. Diese stellen sicher, dass in allen Aufgabenbereichen eine Kultur verankert wird, die vom Kunden aus denkt . Vom Kunden aus denken ist ein grundlegender Bestandteil unseres Innovationsprozesses. Unsere Arbeit richtet sich ganz nach dem Kunden und dessen Wünschen. Kundenfixierte Teams richten die Produktentwicklung auf Kundenwünsche aus.

In Zusammenhang mit Architekturen bedeutet das: Wir erwarten von jedem Team, dass es Architekturen erstellen und nach bewährten Methoden arbeiten kann. Um neuen Teams zu diesen Fähigkeiten zu verhelfen bzw. um bestehende Teams leistungsfähiger zu machen, nehmen wir sie in eine virtuelle Community auf, in der Principal Engineers ihre Entwürfe begutachten und sie an die Best Practices für AWS heranführen. Die Community der Principal Engineers hat die Aufgabe, bewährte Methoden sichtbar und verständlich zu machen. Dies geschieht beispielsweise durch Mittagsvorträge, in denen es um die Anwendung bewährter Methoden an praktischen Beispielen geht. Die Vorträge werden aufgezeichnet und können für das Onboarding neuer Teammitglieder eingesetzt werden.

Wir haben bislang mehrere Tausende Internet-ähnliche Systeme eingerichtet und dabei einen Erfahrungsschatz aufgebaut, aus dem sich die Best Practices für AWS herauskristallisiert haben. Wir bevorzugen, bewährte Methoden mit Hilfe von Daten zu definieren. Wir setzen dafür aber auch Fachexperten (z. B. Principal Engineers) ein. Principal Engineers sind direkt dabei, wenn sich neue bewährte Methoden abzeichnen. Als Community können sie sicherstellen, dass die Teams danach arbeiten. Im Laufe der Zeit werden diese bewährten Methoden in unsere internen Prüfprozesse sowie in Compliance-Mechanismen aufgenommen. Das Well-Architected Framework ist die kundenseitige Implementierung unseres internen Prüfprozesses. Darin ist die Denkweise der Principal Engineers für Zuständigkeitsbereiche vor Ort (z. B. Solutions Architecture, interne Engineering-Teams) festgeschrieben. Das Well-Architected Framework ist ein skalierbarer Mechanismus, mit dem Sie von diesen Erkenntnissen profitieren können.

Wenn so vorgegangen wird wie in einer Community aus Principal Engineers (mit verteilten Architekturzuständigkeiten), kann unserer Ansicht nach eine Well-Architected Enterprise-Architektur zustande kommen, die auf die Kundenwünsche ausgerichtet ist. Technologievordenker (z. B. CTO oder Entwicklungsleiter), die all Ihre Workloads nach den Prinzipien des Well-Architected-Ansatzes prüfen, können die Risiken Ihres Technologieportfolios aufzeigen. Sie identifizieren teamübergreifende Themen, die Ihre Organisation mit Hilfe von Mechanismen, Training oder Mittagsvorträgen angehen könnte. Allesamt Gelegenheiten für Ihre Principal Engineers, ihr Wissen zu bestimmten Themen an mehrere Teams weiterzugeben.

Allgemeine Designprinzipien

Das Well-Architected Framework fasst allgemeine konzeptionelle Grundsätze zusammen, die gutes Design in der Cloud fördern:

- Keine Ungewissheit mehr über die Kapazität: Wenn Sie bei der Bereitstellung eines Workloads eine schlechte Entscheidung zur Kapazität treffen, sitzen Sie anschließend möglicherweise auf nicht genutzten Ressourcen oder haben zu wenig Kapazität und müssen sich mit mangelnder Performance herumschlagen. Beim Cloud Computing gibt es diese Probleme nicht. Sie arbeiten mit so viel oder so wenig Kapazität wie nötig. Das System wird automatisch hoch- oder herunterskaliert.
- Systeme auf Produktionsbetrieb testen: Sie können in der Cloud bei Bedarf eine Testumgebung in Produktionsgröße einrichten, Ihre Tests abschließen und die Ressourcen dann wieder außer Betrieb nehmen. Weil Sie für die Testumgebung nur dann zahlen, wenn sie genutzt wird, können Sie Ihre Live-Umgebung zu einem Bruchteil der Kosten testen, die Sie an einem lokalen Standort hätten.
- Automatisierung vereinfacht Architekturexperimente: Wenn Sie automatisieren, können Sie Ihre Workloads kostengünstig erstellen und replizieren und vermeiden manuellen Aufwand. Sie können an der Automatisierung vorgenommene Änderungen nachverfolgen, die Auswirkungen nachprüfen und ggf. auf die vorherigen Parameter zurücksetzen.
- Voraussetzungen für evolutionäre Architekturen schaffen: In herkömmlichen Umgebungen sind architekturrelevante Entscheidungen oft als statische, einmalig auftretende Ereignisse implementiert. Dementsprechend gibt es während der Lebensdauer des Systems einige wenige große Versionen. Geschäftsvoraussetzungen und ihr Kontext entwickeln sich stetig weiter. Diese anfangs getroffenen Entscheidungen könnten die Fähigkeit des Systems beeinträchtigen, sich auf neue Geschäftsvoraussetzungen einzustellen. In der Cloud können Sie jederzeit automatisieren und testen. Dadurch wird weniger wahrscheinlich, dass sich Änderungen am Design negativ auswirken. Dieses System kann sich im Laufe der Zeit weiterentwickeln. Unternehmen können dann wie selbstverständlich Innovationen für sich nutzen.
- Mit Daten Architekturen weiterentwickeln: Sie können in der Cloud Daten zu der Frage sammeln, wie Ihre architekturrelevanten Entscheidungen auf das Verhalten Ihres Workloads durchschlagen. Sie können also mit faktenbasierten Entscheidungen Ihren Workload verbessern. Ihre Cloud-Infrastruktur ist Code. Das bedeutet, dass Sie diese Daten im Laufe der Zeit in architekturrelevante Entscheidungen und Verbesserungsmaßnahmen einfließen lassen können.
- Verbesserung mit Hilfe von Ernstfallübungen: Simulieren Sie an regelmäßigen Gamedays Vorfälle in der Produktion, um das Verhalten Ihrer Architektur und Prozesse zu simulieren. So können Sie

nachvollziehen, wo nachgebessert werden kann. Zudem üben Sie dabei ein, wie Ihre Organisation mit Ereignissen umgeht.

Die Säulen des Framework

Wenn Sie ein Softwaresystem bauen, gehen Sie ähnlich vor wie beim Hausbau. Wenn das Fundament nicht trägt, können Risse auftreten und das Gebäude unbrauchbar machen. Wenn Sie die Architektur einer Technologielösung planen und die sechs Säulen Operative Exzellenz, Sicherheit, Zuverlässigkeit, Leistungseffizienz, Kostenoptimierung und Nachhaltigkeit vernachlässigen, kann es schwer werden, ein System zu schaffen, das Ihre Erwartungen und Anforderungen erfüllt. Berücksichtigen Sie aber diese Säulen in Ihrer Architektur, steht am Ende ein stabiles, effizientes System. Und das gibt Ihnen Freiraum, um sich auf andere Designaspekte (z. B. funktionale Anforderungen) zu konzentrieren.

Säulen

- Operational Excellence
- Sicherheit
- Zuverlässigkeit
- Leistungseffizienz
- Kostenoptimierung
- Nachhaltigkeit

Operational Excellence

Die Säule für die betriebliche Exzellenz beinhaltet die Fähigkeit, die Entwicklung zu unterstützen und Workloads effektiv auszuführen, Einblicke in die Betriebsabläufe zu erhalten und unterstützende Prozesse und Verfahren fortlaufend zu verbessern, um geschäftlichen Mehrwert zu schaffen.

Die Säule „Betriebliche Exzellenz“ gibt einen Überblick über konzeptionelle Grundsätze, bewährte Methoden und Fragen. Obligatorische Anleitungen zur Implementierung finden Sie im Whitepaper zur Säule für die betriebliche Exzellenz.

Themen

- Designprinzipien
- Definition
- Bewährte Methoden
- Ressourcen

Designprinzipien

Es gibt fünf Designprinzipien für operative Exzellenz in der Cloud:

- Betriebliche Vorgänge als Code ausführen ("Operations-as-Code"): In der Cloud können Sie die gleichen technischen Vorgehensweisen wie beim Anwendungscode in Ihrer gesamten Umgebung anwenden. Sie können sämtliche Workloads (Anwendungen, Infrastruktur) als Code definieren und mit Code aktualisieren. Sie können Ihre betrieblichen Verfahren als Code implementieren und deren Ausführung automatisieren, indem Sie sie von Ereignissen auslösen lassen. Indem der Betrieb als Code ausgeführt wird, werden menschliche Fehler ausgeräumt und einheitliche Reaktionen auf Ereignisse möglich gemacht.
- Vornehmen kleiner, häufiger und umkehrbarer Änderungen: Legen Sie Workloads so aus, dass es möglich ist, Komponenten regelmäßig zu aktualisieren. Nehmen Sie Änderungen in kleinen Schritten vor, die wieder zurückgenommen werden können (ohne dass Kunden dadurch beeinträchtigt werden, sofern möglich).
- Betriebliche Verfahren regelmäßig nachbessern: Suchen Sie beim Einsatz betrieblicher Verfahren nach Möglichkeiten, diese zu verbessern. Entwickeln Sie beim Ausbau Ihrer Workloads auch Ihre Verfahren entsprechend weiter. Legen Sie regelmäßige Termine fest, an denen überprüft wird, ob alle Verfahren effektiv und alle Teams mit den Verfahren vertraut sind.
- Fehlern vorbeugen: Führen Sie vorbeugende Übungen durch, um potenzielle Fehlerquellen zu identifizieren, damit diese behoben oder umgangen werden können. Testen Sie Ihre Ausfallszenarien und stellen Sie sicher, dass Sie deren Auswirkungen kennen. Testen Sie Ihre Reaktionsverfahren, um sicherzustellen, dass diese wirksam sind und dass Ihre Teams mit deren Ausführung vertraut sind. Legen Sie regelmäßige Termine fest, an denen getestet wird, wie Workloads und Teams auf simulierte Ereignisse reagieren.
- Aus allen betrieblichen Ausfällen lernen: Ziehen Sie aus allen betrieblichen Zwischenfällen und Ausfällen entsprechende Lehren und treiben Sie geeignete Verbesserungen voran. Geben Sie Ihre Erkenntnisse an alle Teams in Ihrer gesamten Organisation weiter.

Definition

Die bewährte Methoden für betriebliche Exzellenz in der Cloud lassen sich in vier Bereiche einteilen:

- Organisation
- Vorbereitung
- Betrieb

- Weiterentwicklung

Die Geschäftsleitung Ihres Unternehmens definiert Geschäftsziele. Anforderungen und Prioritäten müssen in Ihrem Unternehmen bekannt sein, damit Aufgaben entsprechend organisiert und durchgeführt und die Geschäftsergebnisse erreicht werden können. Ihr Workload muss die Informationen ausgeben, die für die Unterstützung erforderlich sind. Die Implementierung von Services zur Integration, Bereitstellung und Lieferung Ihres Workloads ermöglicht einen erhöhten Fluss nützlicher Änderungen in die Produktion, indem wiederkehrende Prozesse automatisiert werden.

Es kann Risiken im Zusammenhang mit dem Betrieb Ihres Workloads geben. Sie müssen diese Risiken verstehen und eine fundierte Entscheidung dazu treffen, ob der Übergang in die Produktion vollzogen werden sollte. Ihre Teams müssen in der Lage sein, den Workload zu unterstützen. Geschäfts- und Betriebsmetriken, die von den gewünschten Geschäftsergebnissen abgeleitet werden, ermöglichen Ihnen, den Zustand Ihres Workloads und Ihrer Betriebsaktivitäten nachzuvollziehen und auf Vorfälle zu reagieren. Ihre Prioritäten ändern sich, wenn sich Ihre geschäftlichen Anforderungen und die geschäftliche Umgebung ändern. Verwenden Sie diese als Feedback-Schleife, um Ihr Unternehmen und den Betrieb Ihres Workloads kontinuierlich zu verbessern.

Bewährte Methoden

Themen

Organisation

Um die Prioritäten festlegen zu können, die den geschäftlichen Erfolg ermöglichen, müssen Ihre Teams gemeinsam in Erfahrung bringen, wie sämtliche Workloads aussehen, welche Rolle die einzelnen Teams dabei spielen und was für geschäftliche Ziele damit erreicht werden sollen. Mit gut definierten Prioritäten erzielen Ihre Bemühungen den größtmöglichen Nutzen. Bewerten Sie die Bedürfnisse interner und externer Kunden. Binden Sie dabei alle wichtigen Beteiligten ein, einschließlich der Geschäfts-, Entwicklungs- und Betriebsteams, um zu bestimmen, auf

welche Bereiche die Anstrengungen konzentriert werden sollten. Durch das Bewerten von Kundenbedürfnissen wird sichergestellt, dass Sie den Support, der für die Erzielung der gewünschten geschäftlichen Ergebnisse erforderlich ist, genau kennen und verstehen. Stellen Sie sicher, dass Sie sich der Richtlinien oder Verpflichtungen bewusst sind, die von der Führung Ihres Unternehmens definiert wurden. Bewerten Sie externe Faktoren, z. B. gesetzliche Compliance-Anforderungen und Branchenstandards, die einen bestimmten Fokus erfordern oder verstärken können. Überprüfen Sie, ob Sie Mechanismen haben, um Änderungen an internen Governance- und externen Compliance-Anforderungen zu identifizieren. Wenn keine Anforderungen festgestellt werden, stellen Sie sicher, dass diese Prüfung sorgfältig durchgeführt wurde. Überprüfen Sie Ihre Prioritäten regelmäßig, damit sie bei Bedarf aktualisiert werden können.

Bewerten Sie Bedrohungen für das Unternehmen (z. B. Geschäftsrisiken und -verpflichtungen und Bedrohungen der Informationssicherheit) und pflegen Sie diese Informationen in einem Risikoregister. Bewerten Sie die Auswirkungen von Risiken und Kompromissen zwischen konkurrierenden Interessen oder alternativen Ansätzen. Beispielsweise kann eine beschleunigte Markteinführung neuer Funktionen vor der Kostenoptimierung Vorrang haben, oder Sie können eine relationale Datenbank für nicht relationale Daten wählen, um die Migration eines Systems ohne Refactoring zu vereinfachen. Wägen Sie die Vorteile und Risiken ab, um fundierte Entscheidungen zu treffen, wenn es darum geht, auf welche Bereiche die Anstrengungen konzentriert werden sollen. Einige Risiken oder Entscheidungen können eine bestimmte Zeit lang akzeptabel sein. Es gibt ggf. die Möglichkeit, die damit verbundenen Risiken zu minimieren, oder es ist zu einem bestimmten Zeitpunkt nicht mehr akzeptabel, dass ein Risiko weiterhin bestehen bleibt. In diesem Fall ergreifen Sie Maßnahmen, um das Risiko zu beheben.

Ihre Teams müssen ihre Rolle beim Erreichen von Geschäftsergebnissen verstehen. Teams müssen ihre Rollen beim Erfolg anderer Teams verstehen, die Rolle anderer Teams bei ihrem eigenen Erfolg und sie müssen gemeinsame Ziele haben. Wenn sie Verantwortlichkeit, Zuständigkeit und Entscheidungsfindung verstehen und wissen, wer zum Treffen von Entscheidungen berechtigt ist, können sie die Anstrengungen fokussieren und Ihren Teams zu maximalen Vorteilen verhelfen. Die Anforderungen eines Teams werden durch den unterstützten Kunden, das Unternehmen, die Zusammensetzung des Teams und die Merkmale der jeweiligen Workloads beeinflusst. Es ist nicht sinnvoll, davon auszugehen, dass ein einziges Betriebsmodell alle Teams und Workloads in Ihrem Unternehmen unterstützen kann.

Stellen Sie sicher, dass für jede Anwendung, jeden Workload, jede Plattform und jede Infrastrukturkomponente zuständige Besitzer vorhanden sind und dass jeder Prozess und jedes Verfahren einen festen Besitzer hat, der für die Definition verantwortlich ist, und Besitzer, die für die Leistung verantwortlich sind.

Durch das Verständnis für den geschäftlichen Nutzen der einzelnen Komponenten, Prozesse und Verfahren sowie dafür, weshalb diese Ressourcen vorhanden sind oder Aktivitäten ausgeführt werden und warum diese Zuständigkeit besteht, basieren die Aktionen Ihrer Teammitglieder auf fundierten Informationen. Definieren Sie eindeutig die Verantwortlichkeiten der Teammitglieder, damit sie entsprechend handeln und Mechanismen zur Identifizierung von Verantwortlichkeit und Zuständigkeit besitzen. Nutzen Sie entsprechende Mechanismen zum Anfordern von Ergänzungen, Änderungen und Ausnahmen, damit Sie die Innovation nicht einschränken. Definieren Sie Vereinbarungen zwischen Teams, die beschreiben, wie sie für die gegenseitige und die Unterstützung der Geschäftsergebnisse zusammenarbeiten.

Unterstützen Sie Ihre Teammitglieder, damit sie effektiver handeln und positiv zu Ihrem Geschäftsergebnis beitragen können. Die beteiligten Führungskräfte sollten Erwartungen festlegen und den Erfolg messen. Sie sollten als Sponsor, Fürsprecher und treibende Kraft für die Übernahme bewährter Methoden und die Weiterentwicklung des Unternehmens auftreten. Die Teammitglieder müssen Maßnahmen ergreifen können, wenn Ergebnisse gefährdet sind, um Auswirkungen zu minimieren. Sie müssen dazu ermutigt werden, Entscheidungsträger und Interessenvertreter über ermittelte Risiken zu informieren, damit diese angegangen und Vorfälle vermieden werden können. Kommunizieren Sie bekannte Risiken und geplante Ereignisse zeitnah, klar und umsetzbar, damit Teammitglieder rechtzeitig entsprechende Maßnahmen ergreifen können.

Ermutigen Sie das Ausprobieren neuer Ansätze, damit schneller Erkenntnisse erreicht werden und sorgen Sie dafür, dass Teammitglieder interessiert und motiviert bleiben. Teams müssen ihre Fähigkeiten erweitern, um neue Technologien einzuführen und Änderungen bei Bedarf und Zuständigkeiten zu unterstützen. Dies sollten sie durch spezielle, strukturierte Lernzeiten unterstützen und ermutigen. Stellen Sie sicher, dass Ihre Teams über die nötigen Ressourcen verfügen (Tools und Teammitglieder), um positiv zu Ihren Geschäftsergebnissen beitragen zu können. Profitieren Sie von der Diversität im gesamten Unternehmen, um verschiedene einzigartige Standpunkte zu erfahren. Nutzen Sie diese Perspektive, um Innovation zu fördern, Ihre Annahmen in Frage zu stellen und das Risiko einer Verzerrung durch automatische Bestätigung zu reduzieren. Erweitern Sie Inklusion, Diversität und Offenheit innerhalb Ihrer Teams, um nützliche Perspektiven zu gewinnen.

Wenn es externe behördliche oder Compliance-Anforderungen gibt, die für Ihre Organisation gelten, sollten Sie Ihre Teams mithilfe der von AWS Cloud-Compliance bereitgestellten Ressourcen darin schulen, welche Auswirkungen es bei Ihren Prioritäten zu berücksichtigen gilt. Das Well-Architected Framework legt den Schwerpunkt auf Lernen, Messen und Verbessern. Es bietet einen konsistenten Ansatz, mit dem Sie Architekturen bewerten und Designs implementieren können, die sich im Laufe der Zeit skalieren lassen. AWS stellt das AWS Well-Architected Tool bereit, mit dem Sie Ihren Ansatz vor der Entwicklung, den Status Ihrer Workloads vor der Produktion und den Status Ihrer Workloads

in der Produktion überprüfen können. Sie können Workloads mit den neuesten bewährten Methoden für die AWS-Architektur vergleichen, ihren Gesamtstatus überwachen und Einblicke in potenzielle Risiken erhalten. AWS Trusted Advisor bietet als Tool Zugriff auf verschiedene wichtige Prüfungen, die Optimierungsempfehlungen ausgeben. Diese Informationen können Ihnen beim Festlegen Ihrer Prioritäten helfen. Kunden mit Business und Enterprise Support erhalten Zugriff auf weitere Prüfungen in den Bereichen Sicherheit, Zuverlässigkeit, Leistung und Kostenoptimierung, die beim Festlegen von Prioritäten noch hilfreicher sind.

AWS kann Ihnen helfen, Ihre Teams über AWS und die verfügbaren Services zu schulen, sodass alle Mitarbeiter wissen, welche Auswirkungen ihre Entscheidungen auf Ihren Workload haben können. Bei der Schulung Ihrer Teams sollten Sie die vom AWS Support (AWS Knowledge Center, AWS Discussion Forums und AWS Support Center) bereitgestellten Ressourcen und AWS-Dokumente nutzen. Wenn Sie eine Frage zu AWS haben, können Sie sich über das AWS Support Center an den AWS Support wenden. AWS stellt in der Amazon Builders' Library auch bewährte Methoden und Muster vor, die wir durch den Betrieb von AWS gelernt haben. Eine Vielzahl weiterer nützlicher Informationen finden Sie im AWS-Blog und im offiziellen AWS-Podcast. AWS Training and Certification bietet einige kostenlose Schulungen durch digitale Kurse im Selbststudium zu den Grundlagen von AWS. Sie können sich auch für eine Schulung registrieren, die von Dozenten geleitet wird, um die AWS-Fähigkeiten und -Fertigkeiten Ihres Teams auszubauen.

Sie sollten Tools oder Services verwenden, mit denen Sie Ihre Umgebungen kontenübergreifend verwalten können, z. B. AWS Organizations. Das unterstützt Sie bei der Verwaltung Ihrer Betriebsmodelle. Services wie AWS Control Tower erweitern diese Verwaltungsfunktion, sodass Sie Pläne (die Ihre Betriebsmodelle unterstützen) für die Einrichtung von Konten definieren, laufende Governance mit AWS Organizations anwenden und die Bereitstellung neuer Konten automatisieren können. Anbieter von verwalteten Services wie AWS Managed Services, AWS Managed Services-Partner oder Anbieter von verwalteten Services im AWS-Partnernetzwerk stellen Fachwissen zur Implementierung von Cloud-Umgebungen bereit und unterstützen Ihre Sicherheits- und Compliance-Anforderungen und Geschäftsziele. Durch die Erweiterung Ihres Betriebsmodells um Managed Services können Sie Zeit und Ressourcen sparen, Ihre internen Teams klein halten und sich auf strategische Ergebnisse konzentrieren, die Ihr Unternehmen auszeichnen, anstatt neue Fähigkeiten und Kompetenzen zu entwickeln.

In den folgenden Fragen geht es um Überlegungen zur operativen Exzellenz. (Eine Liste der Fragen und bewährten Methoden zur operativen Exzellenz finden Sie im Anhang).

OPS 1: Wie können Sie Ihre Prioritäten bestimmen?

Alle Beteiligten müssen verstehen, welchen Anteil sie am geschäftlichen Erfolg haben. Setzen Sie sich gemeinsame Ziele, damit Sie die Prioritäten für Ressourcen festlegen können. Dadurch erzielen Ihre Bemühungen den größtmöglichen Nutzen.

OPS 2: Wie strukturieren Sie Ihr Unternehmen, um die gewünschten Geschäftsergebnisse zu erzielen?

Ihre Teams müssen ihre Rolle beim Erreichen von Geschäftsergebnissen verstehen. Teams müssen ihre Rollen beim Erfolg anderer Teams verstehen, die Rolle anderer Teams bei ihrem eigenen Erfolg und sie müssen gemeinsame Ziele haben. Wenn sie Verantwortlichkeit, Zuständig keit und Entscheidungsfindung nachvollziehen können und wissen, wer dazu berechtigt ist, Entscheidungen zu treffen, können ihre Anstrengungen fokussiert und der Nutzen Ihrer Teams maximiert werden.

OPS 3: Wie unterstützt Ihre Unternehmenskultur Ihre Geschäftsergebnisse?

Stellen Sie Ihren Teammitgliedern Unterstützung bereit, damit sie effektiver handeln und Ihr Geschäftsergebnis unterstützen können.

Manchmal kann es vorkommen, dass man zu viel Augenmerk auf eine kleine Auswahl von operativen Prioritäten richtet. Gehen Sie langfristig gut ausgewogen vor, um sicherzustellen, dass erforderliche Fähigkeiten entwickelt und Risiken verwaltet werden. Überprüfen Sie die Prioritäten regelmäßig und passen Sie sie an geänderte Anforderungen an. Wenn Verantwortlichkeit und Zuständigkeit undefiniert oder unbekannt sind, besteht das Risiko, dass erforderliche Aktionen nicht rechtzeitig ausgeführt werden und redundante und potenziell widersprüchliche Anstrengungen unternommen werden, um diese Anforderungen zu erfüllen. Die Unternehmenskultur wirkt sich direkt auf die Zufriedenheit und Bindung der Teammitglieder aus. Ermöglichen Sie die Interaktion und aktivieren Sie die Fähigkeiten Ihrer Teammitglieder für den Erfolg Ihres Unternehmens. Durch Experimente werden Innovationen möglich und Ideen zu Ergebnissen. Sie sollten anerkennen, dass unerwünschte Ergebnisse erfolgreiche Experimente sein können, durch die ein Pfad aufgezeigt wurde, der nicht zum Erfolg führt.

Vorbereitung

Zur Vorbereitung auf Operational Excellence müssen Sie in Erfahrung bringen, mit welchen Workloads zu rechnen ist und wie diese wahrscheinlich ausfallen werden. Dann können Sie diese so gestalten, dass Sie Einblick in deren Status erhalten und entsprechende Verfahren zu deren Unterstützung entwerfen.

Gestalten Sie Ihren Workload so, dass er die Informationen bereitstellt, die Sie benötigen, um den internen Status (z. B. Metriken, Protokolle, Ereignisse und Ablaufverfolgungen) über alle Komponenten hinweg zu verstehen. Dies erhöht die Transparenz und erleichtert die Untersuchung von Problemen. Iterieren Sie zur Entwicklung der erforderlichen Telemetrie, um den Zustand Ihres Workloads zu überwachen, festzustellen, wann Ergebnisse gefährdet sind, und effektiv zu reagieren. Erfassen Sie beim Instrumentieren Ihres Workloads möglichst viele situationsbezogene Informationen (z. B. Statusänderungen, Benutzeraktivitäten, Zugriffe mit einer Berechtigung, Verwendungszähler) – in dem Wissen, dass Sie die wirklich nützlichen Informationen später herausfiltern können.

Verwenden Sie Strategien, die die Übertragung von Änderungen auf die Produktionsumgebung verbessern und Refactoring, schnelles Feedback zur Qualität sowie eine schnelle Fehlerbehebung ermöglichen. Dadurch fließen nützliche Änderungen schneller in die Produktion ein und es treten bei der Bereitstellung weniger Probleme auf. Zudem können Probleme, die durch Bereitstellungsaktivitäten verursacht oder in Ihren Umgebungen erkannt werden, schnell aufgespürt und gelöst werden.

Verwenden Sie Ansätze, die ein schnelles Feedback zur Qualität liefern und eine umgehende Wiederherstellung des vorherigen Zustands nach Änderungen ermöglichen, die nicht zu den gewünschten Ergebnissen führen. Mit diesen Verfahren können Sie die Auswirkung von Problemen eindämmen, die durch die Bereitstellung von Änderungen entstehen. Kalkulieren Sie nicht erfolgreiche Änderungen ein, damit Sie bei Bedarf schneller reagieren und die vorgenommenen Änderungen testen und validieren können. Achten Sie auf geplante Aktivitäten in Ihren Umgebungen, damit Sie mit dem Risiko von Änderungen umgehen können, die sich auf geplante Aktivitäten auswirken. Nehmen Sie häufige, kleine und umkehrbare Änderungen vor, um den Umfang der Änderungen einzuschränken. Dies erleichtert die Fehlersuche und ermöglicht eine schnellere Korrektur, da die Möglichkeit besteht, eine Änderung zurückzusetzen. Dies bedeutet auch, dass Sie häufiger von den Vorteilen wertvoller Änderungen profitieren.

Bewerten Sie die operative Bereitschaft Ihres Workloads, der Prozesse und Verfahren sowie Ihrer Mitarbeiter, damit Sie die operativen Risiken im Zusammenhang mit Ihrem Workload genau kennen. Sie sollten einen konsistenten Prozess (inklusive manueller und automatisierter Checklisten)

anwenden, damit Sie wissen, wann Sie bereit sind, Ihren Workload oder eine Änderung live zu schalten. Auf diese Weise können Sie auch alle Bereiche finden, die Sie für die Planung benötigen. Ihre routinemäßigen Aktivitäten sollten in Runbooks notiert werden, und Playbooks helfen Ihnen bei der Lösung von Problemen. Machen Sie sich mit den Vorteilen und Risiken vertraut, um fundierte Entscheidungen treffen und Änderungen für die Produktion ermöglichen zu können.

Mit AWS können Sie sämtliche Workloads (Anwendungen, Infrastruktur, Richtlinien, Governance und Betrieb) als Code aufrufen. Das bedeutet, dass Sie für jedes Element Ihres Stacks dieselbe technische Vorgehensweise anwenden können, die Sie für Anwendungscode nutzen. Diese können Sie über Teams oder Organisationen hinweg teilen und damit die Auswirkung der Entwicklungsbemühungen verstärken. Verwenden Sie Operations-as-Code in der Cloud und nutzen Sie die Möglichkeit, sicher zu experimentieren, Ihren Workload und betriebliche Verfahren zu entwickeln und Ausfälle zu üben. Durch den Einsatz von AWS CloudFormation verfügen Sie über konsistente, auf Vorlagen basierende und in einer Sandbox befindliche Entwicklungs-, Test- und Produktionsumgebungen mit steigender betrieblicher Kontrolle.

In den folgenden Fragen geht es um Überlegungen zur operativen Exzellenz.

OPS 4: Wie können Sie Ihren Workload so konzipieren, dass sein jeweiliger Zustand klar ersichtlich ist?

Gestalten Sie Ihren Workload so, dass er die Informationen liefert, die Sie benötigen, um seinen internen Zustand über alle Komponenten (z. B. Metriken, Protokolle und Tracing) hinweg zu verstehen. Auf diese Weise können Sie im Bedarfsfall effektiv reagieren.

OPS 5: Wie können Sie Fehler reduzieren, die Fehlerbehebung erleichtern und den Ablauf bis zur Produktion verbessern?

Verwenden Sie Strategien, die die Übertragung von Änderungen auf die Produktionsumgebung verbessern und Refactoring, schnelles Feedback zur Qualität sowie eine schnelle Fehlerbehebung ermöglichen. Dadurch fließen nützliche Änderungen schneller in die Produktion ein und es treten bei der Bereitstellung weniger Probleme auf. Zudem können Probleme, die durch Bereitstellungsaktivitäten verursacht werden, schnell aufgespürt und gelöst werden.

OPS 6: Wie können Sie Bereitstellungsrisiken eindämmen?

Verwenden Sie Ansätze, die ein schnelles Feedback zur Qualität liefern und eine umgehende Wiederherstellung des vorherigen Zustands nach Änderungen ermöglichen, die nicht zu den gewünschten Ergebnissen führen. Mit diesen Verfahren können Sie die Auswirkung von Problemen eindämmen, die durch die Bereitstellung von Änderungen entstehen.

OPS 7: Wie bringen Sie in Erfahrung, ob Sie für die Unterstützung eines Workloads bereit sind?

Bewerten Sie die betriebliche Bereitschaft Ihres Workloads, Prozesse und Verfahren sowie Ihrer Mitarbeiter, damit Sie die betrieblichen Risiken im Zusammenhang mit Ihrer Workload genau kennen.

Investieren Sie in die Implementierung betrieblicher Aktivitäten als Code, um die Produktivität von Betriebsmitarbeitern zu maximieren, Fehlerraten zu minimieren und automatisierte Reaktionen zu ermöglichen. Beugen Sie wo möglich Fehlern vor und stellen Sie entsprechende Abläufe auf. Wenden Sie Metadaten mithilfe von Ressourcen-Tags und AWS Resource Groups nach einer konsistenten Markierungsstrategie an, um die Identifizierung Ihrer Ressourcen zu ermöglichen. Versehen Sie Ihre Ressourcen mit Tags für Organisation, Kostenkalkulation, Zugriffssteuerung und Zielrichtung der Ausführung von automatisierten Betriebsaktivitäten. Übernehmen Sie Bereitstellungsmethoden, die die Elastizität der Cloud ausnutzen, um Entwicklungsaktivitäten, die Vorabbereitstellung von Systemen und damit schnellere Implementierungen zu ermöglichen. Wenn Sie an Checklisten, mit denen Sie Ihre Workloads beurteilen, Änderungen vornehmen, bedenken Sie auch, was mit live geschalteten Systemen geschehen soll, die mit den Änderungen nicht mehr kompatibel sind.

Betrieb

Der erfolgreiche Betrieb eines Workloads wird daran gemessen, ob geschäftliche Ergebnisse erreicht und Kundenanforderungen erfüllt werden. Definieren Sie zu erwartende Ergebnisse, legen Sie fest, wie der Erfolg gemessen wird, und geben Sie an, welche Metriken in Berechnungen verwendet werden sollen, mit denen festgestellt wird, ob Workload und Betrieb erfolgreich sind. Der betriebliche Status beinhaltet sowohl den Status des Workloads als auch den Status und Erfolg der betrieblichen Vorgänge, die zur Unterstützung des Workloads ausgeführt werden (z. B. Bereitstellung und Vorfallreaktion). Legen Sie Metrikausgangswerte für die Verbesserung, Untersuchung und

Intervention fest. Erfassen und analysieren Sie Ihre Metriken und prüfen Sie dann nach, wie weit diese mit ihrem Verständnis von betrieblichen Erfolgen übereinstimmen und welche Änderungen es im zeitlichen Verlauf gibt. Finden Sie anhand gesammelter Metriken heraus, ob kundenseitige und geschäftliche Anforderungen erfüllt werden, und stellen Sie fest, wo noch etwas verbessert werden kann.

Um betriebliche Exzellenz zu erreichen, ist eine effiziente und effektive Verwaltung betrieblicher Ereignisse erforderlich. Dies gilt sowohl für geplante als auch für ungeplante betriebliche Ereignisse. Greifen Sie bei bekannten Ereignissen auf vorab aufgestellte Runbooks zurück. Lassen Sie sich bei der Untersuchung und Behebung von Problemen von Playbooks helfen. Priorisieren Sie Ihre Reaktionen auf Ereignisse anhand der Beeinträchtigungen, die das jeweilige Ereignis für den Geschäftsbetrieb und die Kunden mit sich bringt. Stellen Sie sicher, dass für einen Alarm, der bei einem bestimmten Ereignis ausgelöst werden soll, auch ein auszuführendes Verfahren inklusive eines zuständigen Besitzers festgelegt ist. Legen Sie vorab fest, welche Mitarbeiter für die Behebung eines Ereignisses zuständig sein sollen. Dazu gehören auch Auslöser für einen Eskalationsprozess, über den im Notfall auf der Grundlage der Dringlichkeit und Auswirkungen weitere Mitarbeiter herangezogen werden sollen. Für den Fall, dass eine nicht vorab festgelegte Vorfallreaktion erforderlich ist, die möglicherweise den geschäftlichen Betrieb beeinträchtigen kann, legen Sie Personen fest, die über die nötige Autorität für Entscheidungen verfügen.

Geben Sie Informationen zum betrieblichen Status von Workloads über Dashboards und Mitteilungen weiter, die auf die Zielgruppe (z. B. Kunde, Unternehmen, Entwickler, Betriebsteam) zugeschnitten sind, damit die jeweiligen Personen geeignete Maßnahmen durchführen können und wissen, wann der normale Betrieb wieder weitergeht.

In AWS können Sie Dashboard-Ansichten Ihrer Metriken generieren, die aus Workloads erfasst wurden oder nativ aus AWS stammen. Sie können CloudWatch oder Anwendungen von Drittanbietern verwenden, um Ansichten von betrieblichen Aktivitäten auf geschäftlicher, Workload-bezogener und betrieblicher Ebene zusammenzustellen und anzuzeigen. AWS stellt über seine Protokollierungsfähigkeiten (wie AWS X-Ray, CloudWatch, CloudTrail und VPC Flow Logs) Einblicke in Workloads bereit. So können Workload-Probleme identifiziert werden, was bei der Ursachenanalyse und Behebung von Fehlern hilft.

In den folgenden Fragen geht es um Überlegungen zur operativen Exzellenz.

OPS 8: Wie können Sie den Zustand Ihres Workloads beurteilen?

Definieren, erfassen und analysieren Sie Workload-Metriken, um einen Einblick in Workload-Ereignisse zu erhalten. Dies ist wichtig, damit Sie bei Bedarf entsprechende Maßnahmen ergreifen können.

OPS 9: Wie können Sie den Zustand Ihrer Operationen beurteilen?

Definieren, erfassen und analysieren Sie Metriken für Operationen, um einen Einblick in Ereigniss e rund um Ihre operativen Abläufe zu erhalten. Dies ist wichtig, damit Sie bei Bedarf entsprech ende Maßnahmen ergreifen können.

OPS 10: Wie bewältigen Sie Workload- und operationsspezifische Ereignisse?

Erarbeiten und prüfen Sie Verfahren für die Reaktion auf Ereignisse, um Beeinträchtigungen für Ihren Workload zu minimieren.

Alle von Ihnen erfassten Metriken sollten an die geschäftlichen Anforderungen und Ergebnisse angepasst werden, die sie unterstützen. Entwickeln Sie skriptbasierte Antworten auf bekannte Ereignisse und automatisieren Sie deren Leistung als Reaktion auf die Ereigniserkennung.

Weiterentwicklung

Sie müssen für anhaltende Operational Excellence dazulernen, Erkenntnisse weitergeben und kontinuierliche Verbesserungen anstreben. Planen Sie Arbeitszyklen ein, um kontinuierlich kleinere Verbesserungen vorzunehmen. Analysieren Sie nach einem Vorfall alle Ereignisse, die sich auf den Kunden auswirken. Identifizieren Sie die beitragenden Faktoren und Präventivmaßnahmen, um Wiederholungen zu begrenzen oder zu verhindern. Teilen Sie den betroffenen Communitys die beitragenden Faktoren nach Bedarf mit. Beurteilen und priorisieren Sie in regelmäßigen Abständen Möglichkeiten für Verbesserungen (z. B. Anfragen nach Features, Behebung von Problemen, Compliance-Anforderungen), inklusive Workload- und Betriebsverfahren.

Nehmen Sie Feedback-Schleifen in Ihre Verfahren auf, um Verbesserungsmöglichkeiten schnell zu erfahren und Rückmeldungen aus dem Praxisbetrieb zu dokumentieren.

Geben Sie die Dinge, die Sie erfahren, an andere Teams weiter, damit alle davon profitieren. Untersuchen Sie, ob Ihre neuen Erkenntnisse vielleicht Trends aufzeigen, und führen Sie nachträglich teamübergreifende Analysen von operativen Metriken durch, um Verbesserungsmöglichkeiten und -methoden festzustellen. Implementieren Sie Änderungen, die zu Verbesserungen führen sollen, und beurteilen Sie deren Ergebnisse.

In AWS können Sie Ihre Protokolldaten nach Amazon S3 exportieren oder Protokolle zur langfristigen Speicherung direkt an Amazon S3 senden. Mit AWS Glue können Sie Ihre Protokolldaten in Amazon S3 für Analysen erkunden und vorbereiten und zugehörige Metadaten in AWS AWS Glue Data Catalog speichern. Amazon Athena kann durch seine native Integration mit AWS Glue dann zum Analysieren Ihrer Protokolldaten durch Abfragen mit Standard-SQL verwendet werden. Mit einem Business Intelligence-Tool wie Amazon QuickSight können Sie Ihre Daten visualisieren, untersuchen und analysieren. Erkennen von Trends und Ereignissen, die zu einer Verbesserung führen können.

In der folgenden Frage geht es um Überlegungen zur operativen Exzellenz.

OPS 11: Wie können Sie Arbeitsvorgänge weiterentwickeln?

Kalkulieren Sie Zeit und Ressourcen für kontinuierliche schrittweise Verbesserungen ein, damit sich die Effektivität und Effizienz Ihrer Operationen ständig weiterentwickeln.

Das Fundament für eine erfolgreiche Weiterentwicklung des Betriebs sind ständige kleinere Verbesserungen, das Bereitstellen sicherer Umgebungen und Zeitrahmen zum Experimentieren, Entwickeln und Testen von Verbesserungen sowie das Schaffen eines Umfeldes, in dem alle ermutigt werden, aus Fehlern zu lernen. Die operative Unterstützung für Sandbox-, Entwicklungs-, Test- und Produktionsumgebungen, mit steigenden Leveln von operativer Kontrolle erleichtert die Entwicklung und steigert die Kalkulierbarkeit, dass Änderungen zu erfolgreichen Ergebnissen führen.

Ressourcen

Weitere Informationen zu bewährten Methoden für betriebliche Exzellenz finden Sie in den folgenden Ressourcen.

Dokumentation

- DevOps und AWS

Whitepaper

- Säule „Betriebliche Exzellenz“

Video

- DevOps bei Amazon

Sicherheit

In der Säule der Sicherheit wird beschrieben, wie Sie Daten, Systeme und Komponenten so schützen, dass Sie Cloud-Technologien nutzen können, um Ihre Sicherheitslage zu verbessern.

Die Säule für Sicherheit bietet einen Überblick über konzeptionelle Grundsätze, Best Practices und Fragen. Verbindliche Leitlinien zur Implementierung finden Sie im Whitepaper der Säule für Sicherheit.

Themen

- Designprinzipien
- Definition
- Bewährte Methoden
- Ressourcen

Designprinzipien

Es gibt sieben Designprinzipien für die Sicherheit in der Cloud:

- Implementieren einer starken IdentitätsgrundlageImplementieren Sie das Prinzip der geringsten Berechtigung und erzwingen Sie die Trennung von Pflichten durch eine entsprechende Autorisierung für jede Interaktion mit Ihren AWS-Ressourcen. Zentralisieren Sie die Identitätsverwaltung und vermeiden Sie die Abhängigkeit von langfristigen statischen Anmeldeinformationen.
- Nachverfolgbarkeit: Überwachen, melden und prüfen Sie Aktionen und Änderungen in Ihrer Umgebung in Echtzeit. Integrieren Sie die Protokoll- und Metrikerfassung in Systeme, um automatisch zu untersuchen und Maßnahmen zu ergreifen.

- Sicherheit auf allen Ebenen: Wenden Sie einen umfassenden Verteidigungsansatz mit mehreren Sicherheitskontrollen an. Wenden Sie diesen auf allen Ebenen an (z. B. Netzwerkgrenzen, VPC, Lastverteilung, alle Instances und Datenverarbeitungsservices, Betriebssystem, Anwendung und Code).
- Automatisieren bewährter Sicherheitsverfahren: Mithilfe automatisierter softwarebasierter Sicherheitsmechanismen können Sie Ihr System sicher, schnell und kosteneffektiv skalieren. Erstellen Sie sichere Architekturen, einschließlich implementierter Kontrollen, die als Code in versionsgesteuerten Vorlagen definiert und verwaltet werden.
- Schutz von Daten während der Übertragung und im Ruhezustand: Klassifizieren Sie Daten nach Sensibilität und Nutzungsmechanismen wie Verschlüsselung, Tokenisierung und Zugriff, sofern zutreffend.
- Trennen von Benutzern und Daten: Verwenden Sie Mechanismen und Tools, um den direkten Zugriff oder die manuelle Verarbeitung von Daten zu reduzieren oder gänzlich zu eliminieren. Sie reduzieren dadurch das Risiko, dass sensible Daten verloren gehen, geändert werden oder anderweitigen Benutzerfehlern unterliegen.
- Vorbereitung auf Sicherheitsereignisse: Seien Sie auf Vorfälle vorbereitet. Richten Sie entsprechend Ihren organisatorischen Anforderungen ein Verfahren zur Vorfallverwaltung sowie Richtlinien für die Überprüfung ein. Simulieren Sie Vorfallreaktionen und nutzen Sie automatisierbare Tools, um die Erkennung, Untersuchung und Wiederherstellung zu beschleunigen.

Definition

Es gibt sechs bewährte Methoden für die Sicherheit in der Cloud:

- Sicherheit
- Identity and Access Management
- Erkennung
- Schutz der Infrastruktur
- Datenschutz
- Vorfallbehandlung

Vor der Entwicklung von Workloads ist es wichtig, geeignete Sicherheitsverfahren festzulegen. Sie müssen die einzelnen Prozesse steuern können. Wichtig ist auch, dass Sie Sicherheitsvorfälle

erkennen, Ihre Systeme und Services schützen und die Vertraulichkeit und Integrität von Daten durch entsprechende Schutzmaßnahmen wahren können. Richten Sie ein gut definiertes und geübtes Verfahren ein, das es Ihnen ermöglicht, auf Sicherheitsvorfälle zu reagieren. Derartige Tools und Techniken sind unabdinglich, um Ihr Unternehmen vor finanziellen Verlusten zu schützen und gesetzliche Vorgaben zu erfüllen.

Das AWS-Modell der geteilten Verantwortung ermöglicht Unternehmen, durch die Migration zur Cloud ihre Sicherheits- und Compliance-Ziele zu erfüllen. Dadurch, dass sich AWS um den physischen Schutz der Infrastruktur unserer Cloud-Services kümmert, können Sie sich als AWS-Kunde darauf konzentrieren, mithilfe unserer Services Ihre Ziele zu erreichen. Sie haben in der AWS Cloud auch einen verbesserten Zugriff auf Sicherheitsdaten und können automatisch auf Sicherheitsereignisse reagieren.

Bewährte Methoden

Themen

Sicherheit

Um Ihre Workload sicher zu betreiben, müssen Sie auf jeden Sicherheitsbereich übergreifende bewährte Methoden anwenden. Nutzen Sie Anforderungen und Prozesse, die Sie in Operational Excellence definiert haben, auf Organisations- und Workload-Ebene, und wenden Sie sie auf alle Bereiche an.

Bleiben Sie auf dem Laufenden mit AWS- und Branchenempfehlungen sowie Bedrohungsinformationen, um Ihr Bedrohungsmodell und Ihre Kontrollziele weiterzuentwickeln. Durch die Automatisierung von Sicherheitsprozessen, Tests und Validierung können Sie Ihre Sicherheitsvorgänge skalieren.

In der folgenden Frage geht es um Überlegungen zur Sicherheit. (Eine Liste der Fragen und bewährten Methoden zur Sicherheit finden Sie im Anhang).

SICH 1: Wie können Sie Ihre Workload sicher betreiben?

Um Ihre Workload sicher zu betreiben, müssen Sie auf jeden Sicherheitsbereich übergreifende bewährte Methoden anwenden. Nutzen Sie Anforderungen und Prozesse, die Sie in Operational Excellence definiert haben, auf Organisations- und Workload-Ebene, und wenden Sie sie auf alle Bereiche an. Bleiben Sie auf dem Laufenden mit Empfehlungen von AWS, branchenspezifischen Quellen sowie Informationsquellen zu Bedrohungen, um Ihr Bedrohungsmodell und Ihre Kontrollz iele weiterzuentwickeln. Durch die Automatisierung von Sicherheitsprozessen, Tests und Validieru ng können Sie Ihre Sicherheitsvorgänge skalieren.

In AWS empfehlen wir die Trennung verschiedener Workloads nach Konto, basierend auf ihrer Funktion und den Anforderungen an die Compliance oder Datensensibilität.

Identity and Access Management

Das Identity and Access Management ist ein wichtiger Bestandteil eines Informationssicherheitsprogramms. Es stellt sicher, dass nur autorisierte und authentifizierte Benutzer in dem von Ihnen gewünschten Umfang auf Ihre Ressourcen zugreifen können. Definieren Sie beispielsweise Prinzipien (d. h. Konten, Benutzer, Rollen und Services, die Aktionen in Ihrem Konto durchführen), erstellen Sie entsprechende Richtlinien, und implementieren Sie eine strenge Verwaltung von Anmeldeinformationen. Diese Elemente der Rechteverwaltung bilden die Grundlage der Authentifizierung und Autorisierung.

In AWS erfolgt die Rechteverwaltung primär durch den AWS Identity and Access Management (IAM)-Service. Damit können Sie sowohl den Benutzerzugriff als auch den programmgesteuerten Zugriff auf AWS-Services und -Ressourcen steuern. Wenden Sie detaillierte Richtlinien an, um Benutzern, Gruppen, Rollen oder Ressourcen Berechtigungen zuzuweisen. Darüber hinaus können Sie die Verwendung starker Kennwörter erzwingen. Sie können deren Komplexität vorgeben, Wiederverwendungen vermeiden und Multi-Factor Authentication (MFA) nutzen. Sie haben die Möglichkeit, die Rechteverwaltung mit Ihrem Verzeichnisdienst zu verbinden. Wenn Sie Workloads haben, die Zugriff auf AWS erfordern, ermöglicht IAM diesen auf sichere Weise durch Rollen, Instance-Profile, Identitätsverbund und temporäre Anmeldeinformationen.

In den folgenden Fragen geht es um Überlegungen zur Sicherheit.

SICH 2: Wie verwalten Sie Identitäten für Personen und Maschinen?

Es gibt zwei Arten von Identitäten, die Sie beim Betrieb sicherer AWS-Workloads verwalten müssen. Wenn Sie wissen, welche Art von Identität Sie verwalten und wie Sie Zugriff gewähren müssen, können Sie sicherstellen, dass die richtigen Identitäten unter den richtigen Bedingungen Zugriff auf die richtigen Ressourcen haben.

Menschliche Identitäten: Ihre Administratoren, Entwickler, Bediener und Endbenutzer benötigen eine Identität für den Zugriff auf Ihre AWS-Umgebungen und -Anwendungen. Dies sind Mitgliede r Ihrer Organisation oder externe Benutzer, mit denen Sie zusammenarbeiten, und die mit Ihren AWS-Ressourcen über einen Webbrowser, eine Client-Anwendung oder interaktive Befehlszeilen-Tools interagieren.

Maschinenidentitäten: Ihre Service-Anwendungen, betrieblichen Tools und Workloads benötigen eine Identität, um Anforderungen an AWS-Services zu stellen, z. B. um Daten zu lesen. Zu diesen Identitäten gehören Maschinen, die in Ihrer AWS-Umgebung ausgeführt werden, z. B. Amazon EC2-Instances oder AWS Lambda-Funktionen. Sie können auch Maschinenidentitäten für externe Parteien verwalten, die Zugriff benötigen. Darüber hinaus verfügen Sie möglicherweise auch über Maschinen außerhalb von AWS, die Zugriff auf Ihre AWS-Umgebung benötigen.

SICH 3: Wie verwalten Sie Berechtigungen für Personen und Maschinen?

Verwalten Sie Berechtigungen zum Steuern des Zugriffs auf Personen- und Maschinenidentitäten, die Zugriff auf AWS und Ihren Workload benötigen. Berechtigungen steuern, wer worauf und unter welchen Bedingungen zugreifen kann.

Anmeldeinformationen dürfen nicht zwischen Benutzern oder Systemen weitergegeben werden. Der Benutzerzugriff sollten nach dem Prinzip der geringsten Rechte erfolgen, passwortgeschützt sein und nur mittels MFA möglich sein. Der programmgesteuerte Zugriff etwa durch API-Aufrufe von AWS-Services sollte mit eingeschränkten Berechtigungen und temporären Anmeldeinformationen erfolgen, die beispielsweise durch den AWS Security Token Service ausgegeben werden.

AWS bietet Ressourcen, die Ihnen das Identity and Access Management erleichtern. Mehr zu den Best Practices erfahren Sie in unseren praktischen Übungen zu den Themen Verwaltung von Anmeldeinformationen und Authentifizierung, Steuerung des Benutzerzugriffs, und Steuerung des programmgesteuerten Zugriffs.

Erkennung

Aufdeckende Kontrollen bieten Ihnen die Möglichkeit, potenzielle Sicherheitsbedrohungen oder -vorfälle zu erkennen. Die Kontrollmechanismen sind ein wesentlicher Bestandteil von Governance-Frameworks. Sie können zur Unterstützung von Qualitätssicherungsverfahren, zur Einhaltung gesetzlicher Vorgaben und Pflichten sowie zur Erkennung und Abwehr von Bedrohungen genutzt werden. Es gibt unterschiedliche Arten aufdeckender Kontrollen. Eine Bestandserfassung der Ressourcen und ihrer detaillierten Attribute trägt beispielsweise zu einer effektiveren Entscheidungsfindung (und Lebenszyklussteuerung) bei, wenn es darum geht, operative Ausgangswerte festzulegen. Sie können auch durch eine interne Prüfung der mit Informationssystemen verbundenen Steuerelemente sicherstellen, dass Ihre Verfahren den Richtlinien und Anforderungen entsprechen. Basierend auf definierten Bedingungen sind passende automatisierte Benachrichtigungen möglich. Diese Steuerelemente sind wichtige reaktive Faktoren, die es Ihrem Unternehmen ermöglichen, den Umfang anomaler Aktivitäten zu ermitteln und zu verstehen.

In AWS können Sie aufdeckende Kontrollen durch Verarbeitungsprotokolle, Ereignisse und Überwachungsfunktionen implementieren, die eine Prüfung, automatisierte Analyse und Benachrichtigung ermöglichen. Mit CloudTrail-Protokollen, AWS API-Aufrufen und CloudWatch können Sie Kennzahlen überwachen und Benachrichtigungen senden. Der Konfigurationsverlauf ist mit AWS Config einsehbar. Amazon GuardDuty ist ein verwalteter Service zur Bedrohungserkennung, der Ihre AWS-Konten und -Workloads zu deren Schutz fortlaufend auf böswillige oder unbefugte Verhaltensweisen überwacht. Mit Serviceprotokollen etwa von Amazon Simple Storage Service (Amazon S3) können Sie Zugriffsanfragen protokollieren.

In der folgenden Frage geht es um Überlegungen zur Sicherheit.

SICH 4: Wie erkennen und untersuchen Sie Sicherheitsereignisse?

Erfassen und analysieren Sie Ereignisse mithilfe von Protokollen und Kennzahlen, um Einblick zu erhalten. Ergreifen Sie Maßnahmen bei Sicherheitsereignissen und potenziellen Bedrohungen, um Ihren Workload zu schützen.

Die Protokollverwaltung ist für eine Well-Architected-Workload wichtig, um so vielfältige Bereiche wie Sicherheit, Forensik sowie die Einhaltung gesetzlicher Vorgaben abzudecken. Zur Ermittlung potenzieller Sicherheitsvorfälle müssen diese Protokolle analysiert und bei Bedarf entsprechende Maßnahmen ergriffen werden. AWS bietet Funktionen, die die Protokollverwaltung erleichtern. Sie

können damit einen Lebenszyklus für die Datenaufbewahrung festlegen oder angeben, wo Daten gespeichert, archiviert oder schließlich gelöscht werden. Dies vereinfacht die vorhersehbare und zuverlässige Datenverarbeitung und senkt die Kosten.

Schutz der Infrastruktur

Zum Schutz der Infrastruktur sind Steuermethoden wie etwa eine tiefgreifende Abwehr erforderlich, um Best Practices sowie organisatorische und gesetzliche Verpflichtungen zu erfüllen. Die Nutzung dieser Methoden ist für erfolgreiche, kontinuierliche Betriebsabläufe sowohl in der Cloud als auch lokal ausschlaggebend.

AWS ermöglicht die Überprüfung zustandsbehafteter und zustandsloser Pakete. Sie können dafür wahlweise AWS-native Technologien oder im AWS Marketplace angebotene Partnerprodukte und -services nutzen. Amazon Virtual Private Cloud (Amazon VPC) wird empfohlen, um eine private, sichere und skalierbare Umgebung zu erstellen, in der Sie Ihre Topologie, einschließlich Gateways, Routing-Tabellen sowie öffentlichen und privaten Subnetzen definieren können.

In den folgenden Fragen geht es um Überlegungen zur Sicherheit.

SEC 5: Wie schützen Sie Ihre Netzwerkressourcen?

Alle Workloads, die über eine Art Netzwerkverbindung verfügen, unabhängig davon, ob es sich um das Internet oder ein privates Netzwerk handelt, erfordern mehrere Abwehrebenen, um Schutz vor externen und internen Netzwerkbedrohungen sicherzustellen.

SICH 6: Wie schützen Sie Ihre Datenverarbeitungsressourcen?

Datenverarbeitungsressourcen in Ihrem Workload erfordern mehrere Ebenen der Abwehr zum Schutz vor externen und internen Bedrohungen. Zu den Datenverarbeitungsressourcen zählen EC2-Instances, Container, AWS Lambda-Funktionen, Datenbankservices, IoT-Geräte und mehr.

Ungeachtet der Umgebung sollten mehrere Abwehrebenen vorhanden sein. Was den Schutz der Infrastruktur anbelangt, gelten viele der Konzepte und Methoden für Cloud- und lokale Modelle gleichermaßen. Das Erzwingen des Grenzschutzes, die Überwachung von Ein- und Ausgangspunkten sowie die umfassende Protokollierung, Überwachung und Benachrichtigung sind für einen effektiven Informationssicherheitsplan wichtig.

AWS-Kunden können die Konfiguration der Amazon Elastic Compute Cloud (Amazon EC2) sowie von Amazon Elastic Container Service-Containern (Amazon ECS) und AWS Elastic Beanstalk-Instances anpassen oder härten und in einem unveränderlichen Amazon Machine Image (AMI) speichern. Dadurch erhalten alle neuen virtuellen Server (Instances), die mit diesem AMI gestartet werden, diese gehärtete Konfiguration. Dabei spielt es keine Rolle, ob sie durch Auto Scaling oder manuell ausgelöst wurden.

Datenschutz

Vor der Entwicklung eines Systems sollten grundlegende Sicherheitspraktiken implementiert werden. Mittels Datenklassifizierung lassen sich beispielsweise organisatorische Daten nach Sensitivität kategorisieren. Die Verschlüsselung macht sie zudem für unbefugte Benutzer unleserlich. Derartige Tools und Techniken sind unabdinglich, um Ihr Unternehmen vor finanziellen Verlusten zu schützen und gesetzliche Vorgaben zu erfüllen.

In AWS können Sie Daten mit folgenden Maßnahmen schützen:

- Als AWS-Kunde behalten Sie die vollständige Kontrolle über Ihre Daten.
- AWS erleichtert Ihnen die Datenverschlüsselung und die Schlüsselverwaltung, einschließlich einer regulären Schlüsselrotation. Sie können diese auf einfache Weise selbst verwalten oder von AWS automatisieren lassen.
- Sie haben Zugriff auf detaillierte Protokolle mit wichtigen Angaben etwa zu Dateizugriffen und -änderungen.
- Die Speichersysteme von AWS zeichnen sich durch eine exzeptionelle Ausfallsicherheit aus. Amazon S3 Standard, S3 Standard-IA, S3 One Zone-IA und Amazon Glacier bieten beispielsweise eine einjährige Objektlanglebigkeit von 99,999999999 %. Dies entspricht einem jährlichen erwarteten Verlust von 0,000000001 % der Objekte.
- Die Versionierung, die in ein umfassenderes Verfahren zur Datenlebenszyklusverwaltung eingebunden sein kann, bietet Schutz vor versehentlichen Überschreibungen, Löschungen und ähnlichen Gefahren.
- AWS veranlasst niemals eine Verschiebung von Daten zwischen Regionen. Die in einer Region platzierten Inhalte bleiben in dieser Region, sofern Sie dies nicht ausdrücklich mithilfe einer Funktion oder eines Services veranlassen.

In den folgenden Fragen geht es um Überlegungen zur Sicherheit.

SICH 7: Wie klassifizieren Sie Ihre Daten?

Die Datenklassifizierung bietet eine Möglichkeit, Daten basierend auf Wichtigkeit und Sensibilität zu kategorisieren, um Ihnen dabei zu helfen, angemessene Schutz- und Aufbewahrungskontrollen zu bestimmen.

SICH 8: Wie schützen Sie Ihre Daten im Ruhezustand?

Schützen Sie Ihre Daten im Ruhezustand, indem Sie mehrere Kontrollen implementieren, um das Risiko eines unbefugten Zugriffs oder eines Missbrauchs zu reduzieren.

SICH 9: Wie schützen Sie Ihre Daten bei der Übertragung?

Schützen Sie Ihre Daten während der Übertragung, indem Sie mehrere Kontrollen implementieren, um das Risiko eines unbefugten Zugriffs oder Verlusts zu reduzieren.

AWS bietet mehrere Möglichkeiten zur Verschlüsselung von Daten im Ruhezustand und während der Übertragung. Unsere Services enthalten Funktionen, die die Verschlüsselung Ihrer Daten erleichtern. Wir haben beispielsweise in Amazon S3 eine serverseitige Verschlüsselung (Server-Side Encryption, SSE) implementiert, die die Speicherung Ihrer Daten in verschlüsselter Form vereinfacht. Sie können auch das komplette Ver- und -Entschlüsselungsverfahren mit HTTPS (generell als SSL-Terminierung bekannt) mit Elastic Load Balancing (ELB) arrangieren.

Vorfallsreaktion

Obwohl die präventiven und aufdeckenden Kontrollen mittlerweile extrem ausgereift sind, sollte Ihr Unternehmen dennoch Verfahren etablieren, um auf Sicherheitsvorfälle reagieren und mögliche Auswirkungen mindern zu können. Wie effektiv Ihre Teams bei einem Vorfall reagieren können, um Systeme zu isolieren oder zu bergen und Betriebsabläufe in einem bekanntermaßen funktionierenden Zustand wiederherzustellen, hängt stark von der Architektur des Workloads ab. Indem Sie sich mit entsprechenden Tools und Zugriffsmöglichkeiten auf Sicherheitsvorfälle vorbereiten und die Vorfallreaktion regelmäßig im Rahmen von Gamedays üben, stellen Sie eine zeitnahe Untersuchung und Wiederherstellung sicher.

In AWS ermöglichen die folgenden Praktiken eine effektive Vorfallreaktion:

- Eine detaillierte Protokollierung wichtiger Informationen etwa zu Dateizugriffen und -änderungen.
- Ereignisse können automatisch verarbeitet werden und Tools auslösen, die Reaktionen über AWS APIs automatisieren.
- Sie können vorab mit AWS CloudFormation entsprechende Tools und einen "Reinraum" bereitstellen. Sie erhalten dadurch eine sichere, isolierte Umgebung für forensische Untersuchungen.

In der folgenden Frage geht es um Überlegungen zur Sicherheit.

SICH 10: Wie können Sie Vorfälle voraussagen, darauf reagieren und diese beheben?

Die Vorbereitung ist entscheidend für eine rechtzeitige und effektive Untersuchung, Reaktion auf und Wiederherstellung nach Sicherheitsvorfällen, um Unterbrechungen der Geschäftsabläufe zu minimieren.

Wichtig ist, dass Sie eine Möglichkeit haben, Ihrem Sicherheitsteam für forensische Zwecke schnell Zugriff gewähren zu können. Automatisieren Sie sowohl die Isolation von Instances als auch die Erfassung von Daten und Zuständen.

Ressourcen

Werfen Sie einen Blick auf die folgenden Ressourcen, um mehr über unsere bewährten Methoden für die Sicherheit zu erfahren.

Dokumentation

- AWS Cloud-Sicherheit
- AWS-Compliance
- AWS-Sicherheitsblog

Whitepaper

- Säule „Sicherheit“
- Übersicht über AWS-Sicherheit
- AWS – Risiko und Compliance

Video

- AWS-Sicherheitsstatus der Union
- Übersicht über die gemeinsame Verantwortlichkeit

Zuverlässigkeit

Die Säule „Zuverlässigkeit“ umfasst die Fähigkeit eines Workloads, die beabsichtigte Funktion erwartungsgemäß korrekt und konsistent auszuführen. Dies umfasst die Möglichkeit, den Workload während des gesamten Lebenszyklus zu betreiben und zu testen. Dieses Dokument bietet umfassende Informationen mit Best Practices für die Implementierung zuverlässiger Workloads in AWS.

Die Säule der Zuverlässigkeit bietet einen Überblick über Designprinzipien, bewährte Methoden und Fragen. Verbindliche Leitlinien zur Implementierung finden Sie im Whitepaper zur Säule der Zuverlässigkeit.

Themen

- Designprinzipien
- Definition
- Bewährte Methoden
- Ressourcen

Designprinzipien

Es gibt fünf Designprinzipien für die Zuverlässigkeit in der Cloud:

- Automatische Wiederherstellung nach einem Fehler: Durch die Überwachung wichtiger Leistungskennzahlen (KPIs, Key Performance Indicators) eines Workloads können Sie die Automatisierung auslösen, sobald ein Schwellenwert überschritten wurde. Diese KPIs sollten als Kennzahlen für den Geschäftswert und nicht als technische Aspekte für den Betrieb des Service betrachtet werden. Dies ermöglicht eine automatische Benachrichtigung bei und Verfolgung von Fehlern sowie die Einleitung einer automatisierten Wiederherstellung, die eine Fehlerumgehung bietet oder den Fehler behebt. Bei einer ausgefeilteren Automatisierung ist es möglich, Fehler vor ihrem eigentlichen Auftreten zu antizipieren und zu beheben.

- Testen von Wiederherstellungsverfahren: In einer lokalen Umgebung werden Tests häufig durchgeführt, um nachzuweisen, dass der Workload in einem bestimmten Szenario funktioniert. Mit den Tests werden in der Regel keine Wiederherstellungsstrategien validiert. In der Cloud können Sie testen, in welchen Situationen die Workload Fehler produziert, und Sie können die Wiederherstellungsverfahren validieren. Mit der Automatisierung können Sie verschiedene Fehler simulieren oder Szenarios reproduzieren, die zuvor zu Fehlern geführt haben. Diese Vorgehensweise legt Fehlerpfade offen, die Sie testen und beheben können, bevor ein echtes Fehlerszenario auftritt. Dadurch werden die Risiken verringert.
- Horizontales Skalieren zur Erhöhung der aggregierten Workload-Verfügbarkeit: Ersetzen Sie eine große Ressource durch mehrere kleine Ressourcen, um die Auswirkung eines einzelnen Fehlers auf den Gesamt-Workload zu reduzieren. Verteilen Sie Anfragen auf mehrere kleinere Ressourcen, damit sie keine gemeinsame Fehlerquelle aufweisen.
- Genaue Analyse der verfügbaren Kapazität: Eine häufige Fehlerursache bei lokalen Workloads ist die Ressourcensättigung. Ein solches Szenario liegt vor, wenn die Anforderungen an den Workload dessen Kapazität überschreiten (dies ist häufig das Ziel von Denial-of-Service-Angriffen). In der Cloud können Sie die Nachfrage und die Workload-Auslastung überwachen und das Hinzufügen oder Entfernen von Ressourcen automatisieren, um den Bedarf ohne Über- oder Unterbereitstellung stets optimal zu erfüllen. Es gibt weiterhin Grenzen, aber einige Kontingente können gesteuert und andere verwaltet werden (siehe "Service Quotas und Einschränkungen verwalten").
- Verwalten von Änderungen an der Automatisierung: Änderungen an Ihrer Infrastruktur sollten über die Automatisierung vorgenommen werden. Zu den Änderungen, die verwaltet werden müssen, gehören Änderungen an der Automatisierung, die anschließend nachverfolgt und überprüft werden können.

Definition

Die bewährten Methoden für Zuverlässigkeit in der Cloud lassen sich in vier Bereiche einteilen:

- Grundlagen
- Workload-Architektur
- Änderungsmanagement
- Fehlerverwaltung

Um Zuverlässigkeit zu erreichen, müssen Sie mit den Grundlagen beginnen – einer Umgebung, in der Servicekontingente und die Netzwerktopologie für die Workload angemessen sind. Die Workload-Architektur des verteilten Systems muss so ausgelegt sein, dass Ausfälle verhindert und abgemildert werden. Die Workload muss Änderungen in Bezug auf den Bedarf oder die Anforderungen verarbeiten und so konzipiert sein, dass sie Fehler erkennt und sie automatisch selbst behebt.

Bewährte Methoden

Themen

Grundlagen

Grundlegende Anforderungen sind diejenigen, deren Umfang über einen einzelnen Workload oder ein einzelnes Projekt hinausgeht. Vor dem Aufbau der Architektur eines System sollten grundlegende Anforderungen, die sich auf die Zuverlässigkeit auswirken, implementiert werden. So müssen Sie beispielsweise Ihre Rechenzentren mit einer ausreichenden Netzwerkbandbreite versorgen.

In AWS sind die meisten dieser grundlegenden Anforderungen bereits berücksichtigt oder können nach Bedarf erfüllt werden. Die Cloud bietet nahezu unbegrenzte Möglichkeiten. Daher liegt es in der Verantwortung von AWS, die Anforderungen in Bezug auf ausreichende Netzwerk- und Rechenkapazität zu erfüllen. Sie können die Ressourcengröße und die Zuweisungen nach Bedarf ändern.

In den folgenden Fragen geht es um Überlegungen zur Zuverlässigkeit. (Eine Liste der Fragen und bewährten Methoden zur Zuverlässigkeit finden Sie im Anhang).

ZUV 1: Was ist bei der Verwaltung von Servicekontingenten und Einschränkungen zu beachten?

Für cloudbasierte Workload-Architekturen gibt es Servicekontingente (die auch als Service Limits bezeichnet werden). Diese Kontingente dienen dazu, nicht versehentlich mehr Ressource n bereitzustellen als nötig und Anfrageraten für API-Vorgänge zu begrenzen, um Services vor Missbrauch zu schützen. Darüber hinaus gibt es Ressourceneinschränkungen, z. B. die Rate, mit

ZUV 1: Was ist bei der Verwaltung von Servicekontingenten und Einschränkungen zu beachten?

der Bits durch ein Glasfaserkabel geschleust werden können, oder die Speichermenge auf einer physischen Festplatte.

ZUV 2: Was ist bei der Planung der Netzwerktopologie zu beachten?

Workloads existieren häufig in mehreren Umgebungen. Dazu gehören mehrere Cloud-Umg ebungen (öffentlich zugängliche und private) und möglicherweise die vorhandene Infrastruktur Ihres Rechenzentrums. Die Pläne müssen Netzwerkaspekte umfassen, wie z. B. die Konnektivität innerhalb und zwischen Systemen, die Verwaltung öffentlicher und privater IP-Adressen und die Auflösung von Domänennamen.

Für cloudbasierte Workload-Architekturen gibt es Servicekontingente (die auch als Service Limits bezeichnet werden). Diese Kontingente sollen verhindern, dass versehentlich mehr Ressourcen bereitgestellt werden als nötig. Zudem begrenzen sie die Anfrageraten für API-Vorgänge, um Services vor Missbrauch zu schützen. Workloads existieren häufig in mehreren Umgebungen. Diese Kontingente müssen Sie für alle Workload-Umgebungen überwachen und verwalten. Dazu gehören mehrere Cloud-Umgebungen (öffentlich zugängliche und private) und möglicherweise die vorhandene Infrastruktur Ihres Rechenzentrums. Die Pläne müssen Netzwerkaspekte umfassen, wie z. B. Konnektivität innerhalb und zwischen Systemen, Verwaltung öffentlicher und privater IP-Adressen und Auflösung von Domänennamen.

Workload-Architektur

Ausgangspunkt für einen zuverlässigen Workload sind vorab getroffene Designentscheidungen für Software und Infrastruktur. Ihre Auswahl in puncto Architektur wirkt sich in allen fünf Well-Architected-Säulen auf das Verhalten der Workload aus. Zur Gewährleistung von Zuverlässigkeit sind bestimmte Muster zu befolgen.

Bei AWS haben Entwickler von Workloads die Wahl zwischen verschiedenen Sprachen und Technologien. AWS SDKs vereinfachen die Codierung durch die Bereitstellung sprachspezifischer APIs für AWS-Services. Diese SDKs und die Auswahl an Sprachen ermöglichen es Entwicklern, die hier aufgeführten bewährten Methoden zur Gewährleistung von Zuverlässigkeit zu implementieren. Entwickler können sich auch darüber informieren, wie Software von Amazon erstellt und betrieben wird. Die Amazon Builders' Library.

In den folgenden Fragen geht es um Überlegungen zur Zuverlässigkeit.

ZUV 3: Wie entwerfen Sie Ihre Workload-Service-Architektur?

Erstellen Sie hoch skalierbare und zuverlässige Workloads mithilfe einer serviceorientierten Architektur (SOA) oder einer Microservices-Architektur. Eine serviceorientierte Architektur (SOA) hat zum Ziel, Softwarekomponenten über Service-Schnittstellen wiederverwendbar zu machen. Die Microservices-Architektur geht noch weiter, um Komponenten kleiner und einfacher zu machen.

ZUV 4: Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle vermieden werden?

Verteilte Systeme nutzen Kommunikationsnetzwerke, um Komponenten wie Server oder Services miteinander zu verbinden. Ihre Workload muss trotz Datenverlust oder höherer Latenz in diesen Netzwerken zuverlässig ausgeführt werden. Komponenten des verteilten Systems müssen so funktionieren, dass sie keine negativen Auswirkungen auf andere Komponenten oder die Workload haben. Diese bewährten Methoden verhindern Ausfälle und verbessern die mittlere Zeit zwischen Ausfällen (MTBF).

ZUV 5: Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle abgemildert oder bewältigt werden?

Verteilte Systeme nutzen Kommunikationsnetzwerke, um Komponenten (wie Server oder Services) miteinander zu verbinden. Ihre Workload muss trotz Datenverlust oder höherer Latenz in diesen Netzwerken zuverlässig ausgeführt werden. Komponenten des verteilten Systems müssen so funktionieren, dass sie keine negativen Auswirkungen auf andere Komponenten oder die Workload haben. Mit den folgenden bewährten Methoden können Workloads Belastungen oder Ausfällen standhalten, schneller wiederhergestellt werden und die Auswirkungen solcher Beeinträchtigungen verringern. Das Ergebnis ist eine verbesserte mittlere Reparaturzeit (MTTR).

Änderungsverwaltung

Änderungen an Ihrem Workload oder der Umgebung müssen vorausgesehen und berücksichtigt werden, um einen zuverlässigen Betrieb der Workload zu erreichen. Zu diesen Änderungen gehören durch äußere Faktoren hervorgerufene Änderungen (z. B. Bedarfsspitzen) sowie interne Änderungen wie Funktionsbereitstellungen und Sicherheitspatches.

Mit AWS können Sie das Verhalten eines Workloads überwachen und die Reaktion auf KPIs automatisieren. Beispielsweise kann die Workload bei einer zunehmenden Zahl von Benutzern zusätzliche Server hinzufügen. Sie können kontrollieren und steuern, welche Benutzer Änderungen an der Workload vornehmen dürfen, und die Historie dieser Änderungen überprüfen.

In den folgenden Fragen geht es um Überlegungen zur Zuverlässigkeit.

ZUV 6: Was ist bei der Überwachung von Workload-Ressourcen zu beachten?

Protokolle und Metriken sind wertvolle Tools, um einen Einblick in den Zustand Ihrer Workloads zu gewinnen. Sie können Ihre Workload so konfigurieren, dass Protokolle und Metriken überwacht und bei Über- oder Unterschreiten von Schwellenwerten oder wichtigen Ereignissen Benachric htigungen gesendet werden. Dank der Überwachung kann die Workload erkennen, wenn Schwellenwerte für eine niedrige Leistung unterschritten werden oder Ausfälle auftreten, sodass als Reaktion drauf eine automatische Wiederherstellung erfolgen kann.

ZUV 7: Wie lässt sich der Workload so gestalten, dass er sich an Bedarfsänderungen anpasst?

Eine skalierbare Workload bietet die Elastizität, Ressourcen automatisch entsprechend dem aktuellen Bedarf hinzuzufügen oder zu entfernen.

ZUV 8: Wie implementieren Sie Änderungen?

Kontrollierte Änderungen sind erforderlich, um neue Funktionen bereitzustellen und um sicherzus tellen, dass die Workloads und die Betriebsumgebung bekannte Software ausführen und auf vorhersagbare Weise durch Patches aktualisiert oder ersetzt werden können. Wenn diese Änderungen nicht kontrolliert stattfinden, ist es schwierig, ihre Auswirkungen vorherzusagen oder daraus entstehende Probleme zu beheben.

Wenn Sie eine Workload so gestalten, dass Ressourcen als Reaktion auf Bedarfsänderungen automatisch hinzugefügt und entfernt werden, erhöht das nicht nur die Zuverlässigkeit. Vielmehr sorgt diese Vorgehensweise auch dafür, dass geschäftlicher Erfolg nicht zu einer Belastung wird. Bei einer vorhandenen Überwachung wird Ihr Team automatisch benachrichtigt, wenn KPIs von erwarteten Normen abweichen. Mit dem automatischen Protokollieren von Änderungen an Ihrer Umgebung können Sie auf Aktionen prüfen, die sich möglicherweise auf die Zuverlässigkeit ausgewirkt haben, und diese schnell identifizieren. Mit der Kontrolle und Steuerung des Änderungsmanagements können Sie die Regeln durchsetzen, die für die benötigte Zuverlässigkeit sorgen.

Fehlerverwaltung

In Systemen mit großer Komplexität ist es wahrscheinlich, dass Fehler auftreten. Zur Gewährleistung von Zuverlässigkeit muss Ihr Workload auftretende Fehler erkennen und Maßnahmen ergreifen, um Auswirkungen auf die Verfügbarkeit zu vermeiden. Workloads müssen Ausfälle verkraften sowie Probleme automatisch beheben können.

Mit AWS können Sie automatisch auf überwachte Daten reagieren. Wenn eine bestimmte Kennzahl beispielsweise einen Schwellenwert überschreitet, können Sie eine automatische Maßnahme zur Behebung dieses Problems auslösen. Statt also zu versuchen, eine fehlerhafte Ressource, die Teil Ihrer Produktionsumgebung ist, zu diagnostizieren und zu reparieren, können Sie sie durch eine neue Ressource ersetzen und die Analyse der fehlerhaften Ressource extern vornehmen. Da Sie in der Cloud temporäre Versionen eines gesamten Systems zu geringen Kosten aufstellen können, können Sie automatisiertes Testen verwenden, um vollständige Wiederherstellungsprozesse zu überprüfen.

In den folgenden Fragen geht es um Überlegungen zur Zuverlässigkeit.

ZUV 9: Was ist bei der Sicherung von Daten zu beachten?

Sichern Sie Daten, Anwendungen und Konfigurationen, um die Anforderungen im Hinblick auf das Recovery Time Objective (RTO, Wiederherstellungsdauer) und das Recovery Point Objective (RPO, Wiederherstellungszeitpunkt) zu erfüllen.

ZUV 10: Wie schützen Sie Ihren Workload mithilfe der Fehlerisolierung?

Fehlerisolierte Grenzen beschränken die Auswirkungen eines Ausfalls innerhalb eines Workloads auf eine begrenzte Anzahl von Komponenten. Komponenten außerhalb der Grenze sind vom

ZUV 10: Wie schützen Sie Ihren Workload mithilfe der Fehlerisolierung?

Ausfall nicht betroffen. Wenn Sie mehrere fehlerisolierte Grenzen verwenden, können Sie die Auswirkungen auf Ihren Workload einschränken.

ZUV 11: Wie lassen sich Workloads so gestalten, dass sie Komponentenausfälle verkraften?

Workloads, für die eine hohe Verfügbarkeit und eine niedrige mittlere Reparaturzeit erforderlich sind, müssen auf Ausfallsicherheit ausgelegt sein.

ZUV 12: Wie lässt sich die Zuverlässigkeit testen?

Nachdem Sie Ihre Workload so konzipiert haben, dass sie den Belastungen der Produktion standhält, sind Tests die einzige Möglichkeit, sie auf die erwartete Funktionalität und Ausfallsi cherheit hin zu testen.

ZUV 13: Was ist bei der Planung der Notfallwiederherstellung zu beachten?

Backups und redundante Workload-Komponenten sind der Ausgangspunkt Ihrer Strategie für die Notfallwiederherstellung. RTO und RPO sind Ihre Ziele für die Wiederherstellung Ihrer Workload. Legen Sie diese Ziele entsprechend den geschäftlichen Anforderungen fest. Implementieren Sie eine Strategie, um diese Ziele zu erreichen. Berücksichtigen Sie dabei Standorte und Funktionen von Workload-Ressourcen und -Daten. Die Wahrscheinlichkeit von Disruptionen und die Kosten von Wiederherstellungen sind ebenfalls wichtige Faktoren bei der Ermittlung des Unternehm enswerts, den Notfallwiederherstellungen von Workloads bieten.

Sichern Sie Ihre Daten regelmäßig und stellen Sie anhand von Tests der Sicherungsdateien sicher, dass Sie Wiederherstellungen nach logischen und physischen Fehlern durchführen können. Ein Schlüssel zur Verwaltung von Fehlern ist das regelmäßige und automatisierte Testen von Workloads, um Ausfälle hervorzurufen, und das anschließende Beobachten des Wiederherstellungsverhaltens. Führen Sie diese Tests regelmäßig durch, auch nach größeren Workload-Änderungen. Verfolgen Sie KPIs aktiv wie auch das Recovery Time Objective (RTO, Wiederherstellungsdauer) und das Recovery Point Objective (RPO, Wiederherstellungszeitpunkt), um die Ausfallsicherheit

einer Workload (insbesondere unter Fehlertestszenarios) zu bewerten. Die Verfolgung von KPIs unterstützt Sie bei der Identifizierung und Milderung einzelner Fehlerquellen. Hierbei geht es darum, Ihre Prozesse zur Wiederherstellung von Workloads gründlich zu testen, damit Sie darauf vertrauen können, dass Sie alle Daten wiederherstellen und Ihre Kunden unterbrechungsfrei bedienen können. Und zwar selbst dann, wenn länger anhaltende Probleme auftreten. Mit Ihren Wiederherstellungsprozessen sollten Sie sich genauso vertraut machen wie mit Ihren normalen Produktionsprozessen.

Ressourcen

Werfen Sie einen Blick auf die folgenden Ressourcen, um mehr über unsere bewährten Methoden für die Zuverlässigkeit zu erfahren.

Dokumentation

- AWS-Dokumentation
- Globale AWS-Infrastruktur
- AWS Auto Scaling: Funktionsweise von Skalierungsplänen
- Was ist AWS Backup?

Whitepaper

- Säule „Zuverlässigkeit“: AWS Well-Architected
- Implementieren von Microservices in AWS

Leistungseffizienz

Die Säule ""Leistungseffizienz"" umfasst die Fähigkeit, Rechenressourcen effizient entsprechend den Systemanforderungen zu nutzen und diese Effizienz aufrechtzuerhalten, während sich die Nachfrage ändert und die Technologie weiterentwickelt.

Die Säule der Leistungseffizienz bietet einen Überblick über Designprinzipien, bewährte Methoden und Fragen. Verbindliche Leitlinien zur Implementierung finden Sie im Whitepaper zur Säule der Leistungseffizienz.

Themen

Designprinzipien

Es gibt fünf Designprinzipien für die Leistungseffizienz in der Cloud:

- Demokratisieren fortschrittlicher Technologien: Vereinfachen Sie die Implementierung fortschrittlicher Technologien für Ihr Team, indem Sie komplexe Aufgaben an Ihren Cloud-Anbieter delegieren. Statt Ihr IT-Team aufzufordern, sich näher über das Hosten und Ausführen einer neuen Technologie zu informieren, sollten Sie die Technologie als Service nutzen. Es gibt Technologien, wie etwa die NoSQL-Datenbanken, das Transcodieren von Medien sowie Machine Learning, die spezielles Fachwissen erfordern. In der Cloud kann Ihr Team diese Technologien als Service nutzen und sich auf die Produktentwicklung konzentrieren, ohne sich um die Bereitstellung und Verwaltung von Ressourcen kümmern zu müssen.
- Globale Verteilung in wenigen Minuten: Durch die Bereitstellung Ihres Workloads in mehreren AWS-Regionen auf der ganzen Welt können Sie Ihren Kunden geringere Latenz und eine bessere Erfahrung bei minimalen Kosten bieten.
- Nutzen von serverlosen Architekturen: Aufgrund der in der Cloud verwendeten serverlosen Architekturen müssen Sie für herkömmliche Rechenaktivitäten keine physischen Server mehr ausführen und verwalten. Serverlose Speicherservices können beispielsweise als statische Websites genutzt werden, wodurch sich Webserver erübrigen. Ihren Code können Sie von Ereignisservices hosten lassen. Auf diese Weise entfällt nicht nur die Verwaltung physischer Server, sondern auch die Transaktionskosten sinken, da verwaltete Services in der Cloud-Umgebung ausgeführt werden.
- Vermehrtes Experimentieren: Mit virtuellen und automatisierbaren Ressourcen können Sie schnell unterschiedliche Konfigurationen, Instance- oder Speichertypen miteinander vergleichen.
- Aufbringen von technischem Verständnis: Befassen Sie sich mit der Verwendungsweise von Cloud-Services und nutzen Sie stets den Technologieansatz, der für Ihre Workload-Ziele am besten geeignet ist. Berücksichtigen Sie bei der Auswahl des passenden Datenbank- oder Speicherkonzepts beispielsweise die Datenzugriffsmuster.

Definition

Es gibt vier bewährte Methoden für die Leistungseffizienz in der Cloud:

- Auswahl
- Prüfung
- Überwachung
- Kompromisse

Um eine leistungsstarke Architektur sicherzustellen, empfiehlt sich für deren Entwicklung ein datenbasierter Ansatz. Sammeln Sie zu allen Aspekten der Architektur Daten, angefangen vom allgemeinen Design bis hin zur Auswahl und Konfiguration der Ressourcentypen.

Durch regelmäßiges Überprüfen Ihrer Auswahl stellen Sie die bestmögliche Nutzung der sich fortlaufend weiterentwickelnden AWS Cloud sicher. Durch Überwachung erkennen Sie Abweichungen von der erwarteten Leistung. Zur Leistungssteigerung der Architektur können Sie auch Kompromisse eingehen, beispielsweise durch Komprimierung oder Caching, oder indem Sie hinsichtlich der Konsistenz mehr Toleranz einräumen.

Bewährte Methoden

Themen

Auswahl

Die optimale Lösung für einen bestimmten Workload variiert und Lösungen bestehen häufig aus einer Kombination mehrerer Ansätze. Gut geplante Workloads nutzen mehrere Lösungen und bieten verschiedene Möglichkeiten zur Leistungsoptimierung.

AWS-Ressourcen sind in vielen Typen und Konfigurationen verfügbar, wodurch es einfacher ist, einen Ansatz zu finden, der Ihren Workload-Anforderungen weitgehend entspricht. Sie können zudem Optionen nutzen, die sich in Ihrer lokalen Infrastruktur nicht ohne Weiteres umsetzen ließen. Nehmen

wir beispielsweise den verwalteten Service Amazon DynamoDB. Dieser bietet eine vollständig verwaltete NoSQL-Datenbank mit einer Latenz im einstelligen Millisekundenbereich ungeachtet des Volumens.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz. (Eine Liste der Fragen und bewährten Methoden zur Leistungseffizienz finden Sie im Anhang).

LEIST 1: Was ist bei der Wahl einer leistungsfähigen Architektur zu beachten?

Oft sind mehrere Ansätze erforderlich, um die optimale Leistung für eine Workload zu erzielen. Gut geplante Systeme nutzen mehrere Lösungen und Funktionen zur Leistungsoptimierung.

Verwenden Sie einen datengestützten Ansatz bei der Auswahl der Muster und der Implementierung Ihrer Architektur, um eine kostengünstige Lösung zu erzielen. AWS Solutions Architects, AWS Reference Architectures und AWS Partner Network (APN) können Sie mit Branchenwissen bei der Auswahl der Architektur unterstützen. Für ihre Optimierung sind jedoch anhand von Benchmarking oder Belastungstests erfasste Daten erforderlich.

Ihre Architektur wird vermutlich auf einer Reihe unterschiedlicher Ansätze basieren (z. B. ereignisgesteuert, ETL oder Pipeline). Implementiert wird sie mit den AWS-Services, die zur Optimierung ihrer Leistung beitragen. In den folgenden Abschnitten erörtern wir die vier Hauptressourcen, die berücksichtigt werden sollten (Datenverarbeitung, Speicher, Datenbank und Netzwerk).

Datenverarbeitung

Durch die Auswahl von Rechenressourcen, die Ihre Bedürfnisse und Leistungsanforderungen erfüllen und dabei eine hohe Kosteneffizienz bieten, können Sie mit derselben Anzahl von Ressourcen mehr erreichen. Beachten Sie bei der Bewertung von Datenverarbeitungsoptionen Ihre Anforderungen im Hinblick auf die Workload-Leistung und die Kosten. Treffen Sie auf dieser Grundlage fundierte Entscheidungen.

In AWS gibt es drei Arten der Datenverarbeitung: Instances, Container und Funktionen.

- Instances sind virtualisierte Server, deren Funktionen mit einer Schaltfläche oder einem API-Aufruf geändert werden können. Da Ressourcenentscheidungen in der Cloud flexibel sind, können Sie mit verschiedenen Servertypen experimentieren. AWS bietet diese virtuellen Server-Instances in unterschiedlichen Varianten und Größen mit einer umfassenden Auswahl an Optionen, einschließlich Solid-State-Laufwerken (SSDs) und Grafikprozessoren (GPUs).

- Container dienen zur Virtualisierung des Betriebssystems. Sie können damit eine Anwendung und deren Abhängigkeiten in von der Ressource isolierten Prozessen ausführen. AWS Fargate bietet serverlose Datenverarbeitung für Container. Amazon EC2 kann verwendet werden, wenn Sie Kontrolle über die Installation, Konfiguration und Verwaltung Ihrer Datenverarbeitungsumgebung benötigen. Zudem haben Sie die Auswahl unter mehreren Plattformen zur Container-Orchestrierung: Amazon Elastic Container Service (ECS) oder Amazon Elastic Kubernetes Service (EKS).
- Funktionen Damit wird die Ausführungsumgebung vom auszuführenden Code abstrahiert. Mit AWS Lambda können Sie beispielsweise Code ohne eine Instance ausführen.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 2: Was ist bei der Wahl der Datenverarbeitungslösung zu beachten?

Die optimale Datenverarbeitungslösung für eine Workload ist vom Anwendungsdesign sowie von Nutzungsmustern und Konfigurationseinstellungen abhängig. Architekturen können unterschi edliche Datenverarbeitungslösungen für verschiedene Komponenten verwenden und unterschi edliche Funktionen zur Leistungsverbesserung unterstützen. Die Wahl der falschen Datenvera rbeitungslösung für eine Architektur kann die Leistungseffizienz schmälern.

Machen Sie sich bei der Datenverarbeitung die verfügbaren Elastizitätsmechanismen zunutze, um eine ausreichende Kapazität sicherzustellen und die Leistung bei sich ändernden Anforderungen aufrechtzuerhalten.

Speicher

Cloud-Speicher ist eine entscheidende Komponente des Cloud Computing, da hier die Informationen vorgehalten werden, die von der Workload genutzt werden. Cloud-Speicher ist in der Regel zuverlässiger, skalierbarer und sicherer als herkömmliche lokale Speichersysteme. Für Ihre Workload stehen Objekt-, Block- und Dateispeicherservices sowie verschiedene Optionen zur Cloud-Datenmigration zur Auswahl.

In AWS ist Speicher in drei Formen verfügbar: Objekt-, Block- und Dateispeicher:

- Objektspeicher bietet eine skalierbare, robuste Plattform, damit Daten überall im Internet zugänglich sind. Das gilt für benutzergenerierte Inhalte, aktive Archive, serverlose Datenverarbeitung, Big Data-Speicher oder die Sicherung und Wiederherstellung. Bei Amazon

Simple Storage Service (Amazon S3) handelt es sich um einen Objektspeicherservice mit branchenführender Skalierbarkeit, Datenverfügbarkeit, Sicherheit und Leistung. Amazon S3 ist auf eine Verfügbarkeit von 99,999999999 % (elf Neunen) ausgelegt und speichert Daten für Millionen von Anwendungen für Unternehmen weltweit.

- Blockspeicher bietet hochverfügbaren, konsistenten Blockspeicher mit geringer Latenz für virtuelle Hosts. Er ist vergleichbar mit Direct Attached Storage (DAS) oder einem Storage Area Network (SAN). Amazon Elastic Block Store (Amazon EBS) ist auf Workloads ausgelegt, die einen persistenten, für EC2-Instances zugänglichen Speicher benötigen. So können Sie Anwendungen in Sachen Speicherkapazität, Leistung und Kosten optimieren.
- Dateispeicher bietet auf mehreren Systemen Zugriff auf ein gemeinsam genutztes Dateisystem. Dateispeicherlösungen wie Amazon Elastic File System (EFS) eignen sich ideal für Anwendungsfälle wie große Inhalts-Repositorys, Entwicklungsumgebungen, Medienspeicher oder Hauptverzeichnisse von Benutzern. Amazon FSx macht das Starten und Ausführen beliebter Dateisysteme einfach und kostengünstig. Somit können Sie die umfangreichen Funktionen und die hohe Leistung weit verbreiteter Open-Source- und kommerzieller Dateisysteme nutzen.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 3: Was ist bei der Wahl der Speicherlösung zu beachten?

Die optimale Speicherlösung für ein System richtet sich nach der Zugriffsmethode (Block, Datei oder Objekt), den Zugriffsmustern (Zufallsprinzip oder sequenziell), dem erforderlichen Durchsatz , der Zugriffshäufigkeit (online, offline, Archiv), der Aktualisierungshäufigkeit (WORM, dynamisch) sowie den Einschränkungen hinsichtlich Verfügbarkeit und Langlebigkeit. Gut geplante Systeme nutzen mehrere Speicherlösungen und bieten unterschiedliche Möglichkeiten zur Leistungs optimierung und effizienten Ressourcennutzung.

Bei der Auswahl einer Speicherlösung ist wichtig, dass diese Ihren Zugriffsmustern entspricht, um die gewünschte Leistung zu erzielen.

Datenbank

Die Cloud bietet speziell entwickelte Datenbankservices, die verschiedene Probleme in Verbindung mit Ihrer Workload lösen. Sie haben die Wahl aus zahlreichen speziell entwickelten Datenbankmodulen, darunter relationale, Schlüssel-Werte-, Dokument-, In-Memory-, Graph-, Zeitreihen- und Ledger-Datenbanken. Durch die Auswahl der besten Datenbank zur Lösung eines

bestimmten Problems (oder mehrerer Probleme) können Sie sich von restriktiven, einheitlichen monolithischen Datenbanken lösen und sich auf die Entwicklung von Anwendungen konzentrieren, die den Leistungsanforderungen Ihrer Kunden gerecht werden.

In AWS haben Sie die Wahl aus zahlreichen speziell entwickelten Datenbankmodulen, darunter relationale, Schlüssel-Werte-, Dokument-, In-Memory-, Graph-, Zeitreihen- und Ledger-Datenbanken. Mit AWS-Datenbanken müssen Sie sich nicht um Aufgaben zur Datenbankverwaltung kümmern, wie etwa die Bereitstellung von Servern, das Einspielen von Patches, die Einrichtung, die Konfiguration, Backups oder die Wiederherstellung. AWS überwacht kontinuierlich Ihre Cluster, damit Ihre Workloads unterbrechungsfrei ausgeführt werden, und bietet selbstreparierenden Speicher und eine automatisierte Skalierung. So können Sie sich ganz auf die Entwicklung höherwertiger Anwendungen konzentrieren.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 4: Was ist bei der Wahl der Datenbanklösung zu beachten?

Welche Datenbanklösung sich am besten für ein System eignet, hängt von der erforderlichen Verfügbarkeit, Konsistenz, Partitionstoleranz, Latenz, Langlebigkeit, Skalierbarkeit und Abfragefä higkeit ab. Viele Systeme nutzen für verschiedene Untersysteme unterschiedliche Datenbank lösungen und unterstützen unterschiedliche Funktionen zur Leistungsoptimierung. Die Wahl der falschen Datenbanklösung und -funktionen kann die Leistungseffizienz eines Systems schmälern.

Das Datenbankkonzept für Ihre Workload hat erhebliche Auswirkungen auf die Leistungseffizienz. Häufig erfolgt die Auswahl in diesem Bereich nach Unternehmensvorgaben statt auf der Grundlage eines datenbasierten Ansatzes. Ebenso wie beim Speicher sollten auch hier unbedingt die Zugriffsmuster der Workload berücksichtigt werden. Auch gilt zu prüfen, ob andere nicht datenbankgestützte Lösungen möglicherweise effizienter wären (z. B. eine Graph-, Zeitreihen- oder -In-Memory-Datenbank).

Netzwerk

Da das Netzwerk alle Workload-Komponenten miteinander verbindet, kann es große positive und negative Auswirkungen auf die Leistung und das Verhalten von Workloads haben. Zudem gibt es Workloads, die stark von der Netzwerkleistung abhängig sind. Ein Beispiel hierfür ist das High Performance Computing (HPC), für das zur Steigerung der Cluster-Leistung umfassende Netzwerkkenntnisse benötigt werden. Sie müssen die Workload-Anforderungen für Bandbreite, Latenz, Jitter und Durchsatz ermitteln.

In AWS wird das Netzwerk virtualisiert und es sind unterschiedliche Typen und Konfigurationen verfügbar. Das erleichtert Ihnen die Anpassung Ihrer Netzwerkmethoden an die eigenen Anforderungen. AWS bietet zur Optimierung des Netzwerkdatenverkehrs Produktfunktionen wie Enhanced Networking, für Amazon EBS optimierte Instances, Amazon S3 Transfer Acceleration sowie den dynamischen Amazon CloudFront-Service. Zur Verbesserung der Latenz und der Stabilität des Netzwerks finden Sie in AWS zudem Netzwerkfunktionen wie die latenzbasierte Weiterleitung mit Amazon Route 53, Amazon VPC-Endpunkte, AWS Direct Connect und AWS Global Accelerator).

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 5: Was ist beim Konfigurieren der Netzwerklösung zu beachten?

Welche Netzwerklösung für eine Workload optimal ist, richtet sich nach der Latenz, dem erforderl ichen Durchsatz, dem Jitter und der Bandbreite. Die Standortoptionen sind von den physischen Einschränkungen abhängig, z. B. von Benutzer- oder lokalen Ressourcen. Diese Einschränkungen können durch Edge-Standorte oder die Ressourcenplatzierung wettgemacht werden.

Bei der Bereitstellung des Netzwerks müssen Sie den Standort berücksichtigen. Zur Reduzierung der Latenz haben Sie die Möglichkeit, Ressourcen in der Nähe ihres Verwendungsorts zu platzieren. Verwenden Sie Netzwerkmetriken, um Änderungen an der Netzwerkkonfiguration vorzunehmen, wenn sich der Workload ändert. Mit den entsprechenden Regionen, Platzierungsgruppen und Edge-Services können Sie die Leistung erheblich steigern. Da cloudbasierte Netzwerke schnell umgebaut oder geändert werden können, müssen Sie Ihre Netzwerkarchitektur im Laufe der Zeit weiterentwickeln, um weiterhin eine effiziente Leistung zu erzielen.

Prüfen Sie die Angaben.

Da sich Cloud-Technologien schnell weiterentwickeln, müssen Sie zur kontinuierlichen Leistungssteigerung dafür sorgen, dass für Workload-Komponenten die neuesten Technologien und Ansätze verwendet werden. Sie müssen kontinuierlich Änderungen an Ihren Workload-Komponenten in Erwägung ziehen, damit Sie die Leistungs- und Kostenziele erreichen. Mit neuen Technologien wie Machine Learning und künstlicher Intelligenz (KI) können Sie Kundenerfahrungen ganz neu gestalten und für alle geschäftlichen Workloads Neuerungen einführen.

Profitieren Sie von den ständigen AWS-Innovationen, deren Grundlage die Anforderungen der Kunden sind. Wir stellen regelmäßig neue Regionen, Edge-Standorte, Services und Funktionen zur Verfügung. Jedes Release kann eine positive Auswirkung auf die Leistungseffizienz Ihrer Architektur haben.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 6: Wie profitiert Ihr Workload von neuen Releases?

Bei der Architektur von Workloads sind die Wahlmöglichkeiten begrenzt. Im Laufe der Zeit werden jedoch immer wieder neue Technologien und Ansätze zur Leistungsoptimierung von Workloads entwickelt.

Wenn Architekturen eine schlechte Leistung aufweisen, liegt dies normalerweise daran, dass ein Prozess zur Überprüfung der Leistung fehlt oder fehlerhaft ist. Sie können ein solches Leistungsprüfverfahren jederzeit implementieren, um durch Anwendung des PDCA-Zyklus (Plan-Do-Check-Act) von Deming iterative Verbesserungen zu fördern.

Überwachung

Nach Implementierung des Workloads müssen Sie die Leistung überwachen, damit Sie vorhandene Probleme beheben können, bevor sich Auswirkungen für Ihre Kunden ergeben. Lassen Sie sich mithilfe von Überwachungsmetriken benachrichtigen, wenn Schwellenwerte überschritten werden.

Amazon CloudWatch ist ein Überwachungsservice, der Ihnen Daten und verwertbare Einblicke bietet. Damit können Sie den Workload überwachen, auf systemweite Leistungsänderungen reagieren und die Ressourcennutzung optimieren. Zudem erhalten Sie einen Gesamtüberblick über den Betriebszustand. CloudWatch erfasst Überwachungs- und Betriebsdaten in Form von Protokollen, Metriken und Ereignissen von Workloads, die in AWS und auf lokalen Servern ausgeführt werden. AWS X-Ray hilft Entwicklern beim Analysieren und Debuggen von verteilten Produktionsanwendungen. Mit AWS X-Ray können Sie Einblicke in die Leistung von Anwendungen gewinnen, Ursachen erkennen und Leistungsengpässe ermitteln. Anhand dieser Informationen können Sie schnell reagieren und die kontinuierliche Verfügbarkeit Ihres Workloads sicherstellen.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 7: Wie lassen sich Ressourcen überwachen, um sicherzustellen, dass sie funktionieren?

Die Systemleistung kann sich mit der Zeit verschlechtern. Überwachen Sie die Systemleistung, um eine Verschlechterung frühzeitig zu erkennen und ihr entgegenzuwirken, etwa indem Sie interne oder externe Faktoren wie das Betriebssystem oder die Anwendungslast korrigieren.

Dass keine Falschmeldungen (False Positives) angezeigt werden, ist für eine effektive Überwachungslösung von entscheidender Bedeutung. Automatisierte Trigger vereiteln Benutzerfehler und können die Fehlerbehebung beschleunigen. Planen Sie Ernstfallübungen ein, bei denen Sie Ihre Benachrichtigungslösung mithilfe von Simulationen in der Produktionsumgebung testen, damit Probleme richtig erkannt werden.

Kompromisse

Bei der Entwicklung von Lösungen können Kompromisse helfen, den optimalen Ansatz zu wählen. Je nach Situation können Sie beispielsweise die Latenz oder die Zeit reduzieren, um die Leistung zu erhöhen, indem Sie bedingte Abstriche bei der Konsistenz, der Langlebigkeit und dem Speicherplatz machen.

AWS ermöglicht Ihnen, globale Veröffentlichungen innerhalb weniger Minuten vorzunehmen. Sie können damit Ressourcen weltweit an verschiedenen Standorten bereitstellen, um die Entfernung zu Endbenutzern und damit die Latenz zu reduzieren. Des Weiteren haben Sie die Möglichkeit, in Informationsspeichern (z. B. Datenbanksystemen) Lesereplikate bereitzustellen, um die Last für die primäre Datenbank zu reduzieren.

In der folgenden Frage geht es um Überlegungen zur Leistungseffizienz.

LEIST 8: Wie lässt sich Leistung durch Kompromisse verbessern?
Durch die Festlegung von Kompromissen beim Gestalten von Lösungen lässt sich der optimale Ansatz einfacher bestimmen. Leistung lässt sich oft durch Zugeständnisse in anderen Bereichen verbessern, etwa bei Konsistenz, Beständigkeit, Zeit und Latenz.

Wenn Sie Änderungen an Ihrer Workload vornehmen, sollten Sie Metriken erfassen und bewerten, um die Auswirkungen dieser Änderungen eindeutig zu bestimmen. Messen Sie die Auswirkungen auf System und Endbenutzer, um nachzuvollziehen, wie Ihre Kompromisse den Workload beeinflussen. Stellen Sie anhand eines systematischen Ansatzes fest (z. B. Lasttests), ob sich die Leistung durch den Kompromiss tatsächlich verbessert.

Ressourcen

Weitere Informationen zu bewährten Methoden für die Leistungseffizienz finden Sie in den folgenden Ressourcen.

Dokumentation

- Amazon S3 Leistungsoptimierung
- Amazon EBS Volume-Leistung

Whitepaper

- Säule „Leistungseffizienz“

Video

- AWS re:Invent 2019: Amazon EC2-Grundlagen (CMP211-R2)
- AWS re:Invent 2019: Leadership Session: Der aktuelle Speicherstatus (STG201-L)
- AWS re:Invent 2019: Leadership Session: Speziell entwickelte Datenbanken von AWS (DAT209-L)
- AWS re:Invent 2019: Konnektivität mit AWS und Hybrid-AWS-Netzwerkarchitekturen (NET317-R1)
- AWS re:Invent 2019: Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems (CMP303-R2)
- AWS re:Invent 2019: Erweitern Sie den Umfang auf Ihre ersten 10 Millionen Benutzer (ARC211-R)

Kostenoptimierung

Die Säule Kostenoptimierung umfasst die Fähigkeit, Systeme so auszuführen, dass sie geschäftlichen Wert bei geringstmöglichen Kosten liefern.

Die Säule der Kostenoptimierung bietet einen Überblick über Designprinzipien, bewährte Methoden und Fragen. Verbindliche Leitlinien zur Implementierung finden Sie im Whitepaper zur Säule der Kostenoptimierung.

Themen

- Designprinzipien
- Definition
- Bewährte Methoden
- Ressourcen

Designprinzipien

Es gibt fünf Designprinzipien für die Kostenoptimierung in der Cloud:

- Implementieren des Cloud-Finanzmanagements: Um finanziellen Erfolg zu haben und die Wertschöpfung in der Cloud zu beschleunigen, müssen Sie in Cloud-Finanzmanagement/ Kostenoptimierung investieren. Ihr Unternehmen muss Zeit und Ressourcen aufwenden, um Know-how in diesem neuen Bereich des Technologie- und Nutzungsmanagements aufzubauen. Wie bei Ihren Funktionen zur Sicherheit oder betriebliche Exzellenz müssen Sie Fähigkeiten durch Wissensaufbau, Programme, Ressourcen und Prozesse aufbauen, damit Sie zu einer kosteneffizienten Organisation werden können.
- Verbrauchsmodell einführen: Zahlen Sie nur für die benötigten Computing-Ressourcen, und erhöhen oder verringern Sie die Nutzung auf Basis Ihrer Geschäftsanforderungen und nicht durch aufwändige Prognosen. Entwicklungs- und Testumgebungen werden in einer normalen Arbeitswoche beispielsweise nur acht Stunden pro Tag benötigt. Sie können diese Ressourcen anhalten, wenn sie nicht verwendet werden und damit potenzielle Einsparungen von 75 % (40 Stunden vs. 168 Stunden) erzielen.
- Gesamteffizienz messen: Messen Sie die geschäftliche Leistung des Workloads und die mit der Bereitstellung verknüpften Kosten. Verwenden Sie diese Kennzahlen, um die Gewinne zu ermitteln, die Sie durch die Erhöhung der Leistung und die Reduzierung der Kosten erzielen.
- Kein Geld mehr für undifferenzierte, aufwendige Arbeiten ausgeben: AWS erledigt die aufwendigsten Arbeiten im Rechenzentrum bezüglich Server-Racks, -Stacks und -Stromversorgung. Außerdem entfällt der betriebliche Aufwand für die Verwaltung von Betriebssystemen und Anwendungen mit verwalteten Services. So können Sie sich auf Ihre Kunden und Geschäftsprojekte anstatt auf die IT-Infrastruktur konzentrieren.
- Ausgaben analysieren und zuordnen: Mit der Cloud ist es einfacher, die Nutzung und die Kosten von Systemen genau zu ermitteln und auf Basis dieser Daten eine transparente Zuordnung der IT-Kosten auf einzelne Workload-Besitzer durchzuführen. Auf diese Weise erhalten Sie Unterstützung bei der Messung der Umsatzrendite (ROI) und Workload-Eigentümer erhalten die Möglichkeit, ihre Ressourcen zu optimieren und die Kosten zu reduzieren.

Definition

Es gibt fünf bewährte Methoden für die Kostenoptimierung in der Cloud:

- Praxis für Cloud-Finanzmanagement

- Ausgabenerkennung und Nutzungsbewusstsein
- Kostengünstige Ressourcen
- Verwaltung von Nachfrage und Bereitstellung von Ressourcen
- Optimierung im Laufe der Zeit

Wie bei den anderen Säulen innerhalb des Well-Architected Framework sind Kompromisse unvermeidbar, so müssen Sie beispielsweise entscheiden, ob Sie die Markteinführungsgeschwindigkeit oder die Kosten optimieren möchten. In manchen Fällen ist es sinnvoll, die Priorität auf Geschwindigkeit zu legen, z. B. verbunden mit einer raschen Markteinführung, der Bereitstellung neuer Funktionen oder einer simplen Fristerfüllung, statt im Vorfeld in Kostenoptimierung zu investieren. Konzeptionelle Entscheidungen werden gelegentlich durch Eile statt auf Basis von Daten getroffen, und man ist immer der Versuchung ausgesetzt, einem potenziellen Szenario zu viel Bedeutung beizumessen, statt Zeit in die Bestimmung der kostengünstigsten Bereitstellung zu investieren. Dies führt häufig übermäßigen und mangelhaft optimierten Bereitstellungen. Es ist jedoch die richtige Wahl, wenn Sie Ressourcen aus Ihrer lokalen Umgebung in die Cloud verlagern und die Optimierung anschließend durchführen möchten. Wenn Sie vorab genügend Arbeit in eine Strategie zur Kostenoptimierung investieren, können Sie die wirtschaftlichen Vorteile der Cloud schneller nutzen, indem Sie eine konsistente Einhaltung bewährter Methoden sicherstellen und Überbereitstellungen vermeiden. In den folgenden Abschnitten finden Sie Techniken und bewährte Methoden sowohl für die erste als auch die fortlaufende Implementierung von Cloud-Finanzmanagement und Kostenoptimierung für Ihre Workloads.

Bewährte Methoden

Themen

Praxis für Cloud-Finanzmanagement

Mit der Einführung der Cloud können Technologieteams dank verkürzter Genehmigungs-, Beschaffungs- und Infrastrukturbereitstellungszyklen schneller innovieren. Ein neuer Ansatz für das

Finanzmanagement in der Cloud ist erforderlich, um geschäftlichen Nutzen und finanziellen Erfolg zu erzielen. Dieser Ansatz ist das Cloud-Finanzmanagement. Es baut Funktionen in Ihrer gesamten Organisation auf, indem organisationsweit Wissensaufbau, Programme, Ressourcen und Prozesse implementiert werden.

Viele Organisationen bestehen aus vielen verschiedenen Einheiten mit unterschiedlichen Prioritäten. Durch die Fähigkeit, Ihre Organisation an mehreren vereinbarten Finanzzielen auszurichten und ihr die Mechanismen zur Erreichung der Ziele bereitzustellen, wird die Effizienz der Organisation gesteigert. Ein leistungsfähiges Unternehmen innoviert und entwickelt schneller, ist agiler und passt sich einfacher an beliebige interne oder externe Faktoren an.

In AWS können Sie Cost Explorer und optional Amazon Athena und Amazon QuickSight mit dem Kosten- und Nutzungsbericht (Cost and Usage Report, CUR) verwenden. So können Sie in Ihrer gesamten Organisation ein Kosten- und Nutzungsbewusstsein schaffen. AWS-Budgets bietet proaktive Benachrichtigungen zu Kosten und Nutzung. Die AWS-Blogs bieten Informationen zu neuen Services und Funktionen, damit Sie immer über neue Serviceversionen auf dem Laufenden sind.

In der folgenden Frage geht es um Überlegungen zur Kostenoptimierung. (Eine Liste der Fragen und bewährten Methoden zur Kostenoptimierung finden Sie im Anhang).

KOSTEN 1: Wie implementieren Sie das Cloud Financial Management?

Die Implementierung von Cloud Financial Management (CFM) ermöglicht es Unternehmen, geschäftlichen Nutzen und finanziellen Erfolg zu erzielen, wenn sie ihre Kosten und Nutzung optimieren und auf AWS skalieren.

Beim Aufbau einer Kostenoptimierungsfunktion sollten Sie Teammitglieder einsetzen und das Team um Experten für CFM und Kostenoptimierung ergänzen. Bestehende Teammitglieder wissen, wie die Organisation derzeit funktioniert und Verbesserungen schnell implementiert werden können. Erwägen Sie auch, Personen mit ergänzenden oder speziellen Kenntnissen, wie im Bereich Analyse oder Projektmanagement, mit einzubinden.

Wenn Sie in Ihrer Organisation ein Kostenbewusstsein implementieren, verbessern Sie vorhandene Programme oder bauen auf diesen auf. Es geht viel schneller, bestehende Prozesse und Programme zu ergänzen, als sie neu zu erstellen. So werden die Ergebnisse viel schneller erreicht.

Ausgabenerkennung und Nutzungsbewusstsein

Die erhöhte Flexibilität und Agilität der Cloud fördert Innovationen und schnelle Entwicklungen und Bereitstellungen. Diese Merkmale eliminieren die manuellen Prozesse und den Zeitaufwand für die Bereitstellung einer lokalen Infrastruktur, einschließlich der Identifizierung von Hardware-Spezifikationen, dem Verhandeln von Preisen, der Verwaltung von Bestellungen, der Planung von Lieferungen und schließlich der Bereitstellung der Ressourcen. Die einfache Nutzung und die nahezu unbegrenzte On-Demand-Verfügbarkeit macht neue Wege erforderlich, über Ausgaben nachzudenken.

Viele Unternehmen bestehen aus einer Vielzahl von Systemen, die von unterschiedlichen Teams betrieben werden. Die Möglichkeit, die Ressourcenkosten der jeweiligen Organisation oder den jeweiligen Produkteigentümer zuzuordnen, fördert ein effizientes Nutzungsverhalten und hilft, Verschwendung von Ressourcen einzudämmen. Mit einer präzisen Kostenzuordnung wissen Sie, welche Produkte wirklich profitabel sind, und können fundiertere Entscheidungen in Bezug auf die Budgetaufteilung treffen.

In AWS erstellen Sie mit AWS Organizations oder AWS Control Tower eine Kontostruktur, die eine Trennung ermöglicht und Sie bei der Zuordnung Ihrer Kosten und Nutzung unterstützt. Sie können auch das Ressourcen-Tagging verwenden, um Geschäfts- und Organisationsinformationen auf Ihre Nutzung und Kosten anzuwenden. Verwenden Sie AWS Cost Explorer, um Einblicke in Ihre Kosten und Nutzung zu erhalten, oder erstellen Sie benutzerdefinierte Dashboards und Analysen mit Amazon Athena und Amazon QuickSight. Die Kontrolle Ihrer Kosten und Nutzung erfolgt durch Benachrichtigungen über AWS-Budgets sowie Kontrollen mithilfe von AWS Identity and Access Management (IAM) und Service Quotas.

In den folgenden Fragen geht es um Überlegungen zur Kostenoptimierung.

KOSTEN 2: Wie können Sie die Nutzung steuern?

Definieren Sie Richtlinien und Verfahren, um sicherzustellen, dass sich die Kosten auf dem Weg zur Erreichung Ihrer Ziele in einem angemessenen Rahmen bewegen. Durch den Einsatz eines Kontrollsystems können Sie Innovationen vorantreiben, ohne das Budget zu überschreiten.

KOSTEN 3: Wie können Sie die Nutzung und Kosten überwachen?

Definieren Sie Richtlinien und Verfahren, um Ihre Kosten überwachen und richtig zuordnen zu können. Dadurch können Sie die Kosteneffizienz des Workloads bewerten und verbessern.

KOSTEN 4: Wie können Sie Ressourcen außer Betrieb nehmen?

Implementieren Sie vom Beginn bis zum Abschluss eines Projekts eine Änderungskontrolle und Ressourcenverwaltung. Auf diese Weise können Sie ungenutzte Ressourcen herunterfahren oder beenden, um Verschwendungen zu minimieren.

Sie können Tags für die Kostenzuordnung verwenden, um Ihre Nutzung und Kosten in AWS zu kategorisieren und zu verfolgen. Wenn Sie Tags auf Ihre AWS-Ressourcen anwenden (z. B. EC2-Instances oder S3-Buckets), generiert AWS einen Kosten- und Nutzungsbericht mit Ihrer Nutzung und Ihren Tags. Sie können Tags anwenden, die für Unternehmenskategorien stehen (z. B. Kostenstellen, Workload-Namen oder Besitzer), um Ihre Kosten verschiedenen Services zuzuordnen.

Achten Sie darauf, dass Sie den richtigen Detail- und Granularitätsgrad für die Kosten- und Nutzungsberichterstattung und -überwachung verwenden. Um allgemeine Erkenntnisse zu gewinnen und Trends zu erkennen, verwenden Sie die tägliche Granularität mit AWS Cost Explorer. Für tiefgehendere Analysen und Prüfungen verwenden Sie die stündliche Granularität in AWS Cost Explorer oder Amazon Athena und Amazon QuickSight sowie den Kosten- und Nutzungsbericht (CUR) mit stündlicher Granularität.

Durch die Kombination von mit Tags gekennzeichneten Ressourcen und Entitätslebenszyklus-Tracking (Mitarbeiter, Projekte) können Sie verwaiste Ressourcen oder Projekte identifizieren, die für das Unternehmen keinen Wert mehr generieren und außer Betrieb genommen werden sollten. Sie können Abrechnungsbenachrichtigungen einrichten, um Sie über prognostizierte Budgetüberschreitungen zu informieren.

Kostengünstige Ressourcen

Die Verwendung geeigneter Instances und Ressourcen für Ihren Workload ist für Kosteneinsparungen von entscheidender Bedeutung. Die Ausführung eines Berichtsprozesses kann auf kleineren Servern beispielsweise bis zu fünf Stunden dauern, auf einem doppelt so teuren großen

Server jedoch lediglich eine Stunde. Auf beiden Servern erhalten Sie dasselbe Ergebnis, der kleinere Server generiert über den Ausführungszeitraum jedoch höhere Kosten.

Architektonisch gute Workloads verwenden die kostengünstigsten Ressourcen; dieses Verhalten kann eine signifikante und positive wirtschaftliche Auswirkung haben. Sie haben außerdem die Möglichkeit, verwaltete Services für die Kostenreduzierung zu verwenden. So können Sie für die E-Mail-Zustellung beispielsweise einen Service nutzen, bei dem die Kosten nach der Anzahl der versendeten Nachrichten berechnet werden, statt Server für diese Aufgabe bereithalten zu müssen.

AWS bietet eine Vielzahl flexibler und kosteneffektiver Preisoptionen für den Erwerb von Instances von Amazon EC2 und anderen Services auf eine Weise, die Ihre Anforderungen ideal erfüllt. On demand Instances zahlen Sie auf Stundenbasis für die genutzte Rechenkapazität und gehen keine Mindestverpflichtungen ein. Savings Plans und Reserved Instances bieten Einsparungen von bis zu 75 % gegenüber On-Demand-Preisen. Mit Spot-Instances können Sie ungenutzte Amazon EC2-Kapazität nutzen und von Einsparungen von bis zu 90 % im Vergleich zum On-Demand-Preis profitieren. Spot Instances eignen sich, wenn das System eine Flotte von Servern toleriert, bei der einzelne Server dynamisch aktiviert und deaktiviert werden können, wie z. B. bei zustandslosen Webservern, bei der Stapelverarbeitung oder bei der Nutzung von HPC und Big Data.

Auch mit der Auswahl geeigneter Services ist es möglich, Nutzung und Kosten zu reduzieren. So können Sie beispielsweise CloudFront nutzen, um das Datenübertragungsvolumen zu reduzieren, oder Kosten vollständig eliminieren, z. B. mit Amazon Aurora on RDS, mit dem Sie kostspielige Datenbanklizensierungskosten vermeiden können.

In den folgenden Fragen geht es um Überlegungen zur Kostenoptimierung.

KOSTEN 5: Wie können Sie die Kosten bei der Auswahl von Services einschätzen?

Bei Amazon EC2, Amazon EBS und Amazon S3 handelt es sich um AWS-Services, die als einzelne Bausteine angeboten werden. Verwaltete Services, etwa Amazon RDS und Amazon DynamoDB, sind AWS-Services auf einer höheren Ebene oder Anwendungsebene. Wenn Sie sich für die richtigen Bausteine und verwalteten Services entscheiden, können Sie die Kosten dieses Workloads optimieren. Durch die Nutzung von verwalteten Services können Sie einen Großteil Ihres administrativen und betrieblichen Overheads reduzieren oder beseitigen und damit Kapazität en für anwendungs- und geschäftsbezogene Aktivitäten gewinnen.

KOSTEN 6: Wie können Sie bei der Auswahl des Ressourcentyps, -umfangs und der Anzahl der Ressourcen Kostenziele erfüllen?

Stellen Sie sicher, dass Sie den geeigneten Ressourcenumfang und die Anzahl der Ressourcen für die jeweilige Aufgabe auswählen. Durch die Auswahl des kostengünstigsten Typs, Umfangs und der kostengünstigsten Anzahl minimieren Sie die Verschwendung von Ressourcen.

KOSTEN 7: Wie können Sie Kosten mithilfe von Preismodellen senken?

Verwenden Sie das Preismodell, das sich für Ihre Ressourcen am besten eignet. So halten Sie die Ausgaben möglichst niedrig.

KOSTEN 8: Wie können Sie die Kosten für Datenübertragungen planen?

Damit Sie architekturbezogene Entscheidungen zur Kostenminimierung treffen können, müssen Sie unbedingt die Datenübertragungskosten einplanen und überwachen. Eine geringfügige, aber effektive Änderung an der Architektur kann Ihre Betriebskosten über einen längeren Zeitraum hinweg erheblich senken.

Durch das Einkalkulieren der Kosten während der Serviceauswahl und die Verwendung von Tools wie Cost Explorer und AWS Trusted Advisor zur regelmäßigen Überprüfung Ihrer AWS-Nutzung können Sie Ihre Nutzung aktiv überwachen und Ihre Bereitstellungen entsprechend anpassen.

Verwaltung von Nachfrage und Bereitstellung von Ressourcen

Wenn Sie in die Cloud wechseln, zahlen Sie nur für die genutzten Ressourcen. Sie können Ressourcen so bereitstellen, dass sie dem Workload-Bedarf zum jeweiligen Zeitpunkt entsprechen. Dadurch werden kostspielige Überbereitstellungen überflüssig. Sie können den Bedarf auch anpassen, indem Sie eine Drosselung, einen Puffer oder eine Warteschlange verwenden, um den Bedarf zu glätten und ihn mit weniger Ressourcen zu erfüllen, was zu niedrigeren Kosten führt. Außerdem können Sie ihn mit einem Batch-Service zu einem späteren Zeitpunkt verarbeiten.

In AWS können Sie Ressourcen automatisch so bereitstellen, dass sie den Workload-Bedarf erfüllen. Durch Auto Scaling mit bedarfs- oder zeitbasiertem Ansatz können Sie Ressourcen nach Bedarf hinzufügen und entfernen. Wenn Sie in der Lage sind, Bedarfsänderungen zu antizipieren, können

Sie mehr Kosten einsparen und zugleich sicherstellen, dass Ihre Ressourcen Ihren Workload-Anforderungen entsprechen. Sie können Amazon API Gateway verwenden, um eine Drosselung zu implementieren, oder Amazon SQS einsetzen, um eine Warteschlange für Ihren Workload zu implementieren. Mit beiden können Sie den Bedarf für Ihre Workload-Komponenten anpassen.

In der folgenden Frage geht es um Überlegungen zur Kostenoptimierung.

KOSTEN 9: Wie verwalten Sie die Nachfrage und stellen Ressourcen bereit?

Stellen Sie bei einem Workload mit ausgewogenen Ausgaben und Leistungen sicher, dass alles, wofür Sie bezahlen, genutzt wird, und vermeiden Sie eine erhebliche Unterauslastung der Instances. Eine verschobene Auslastungsmetrik in einer der Richtungen wirkt sich nachteilig auf Ihr Unternehmen aus, entweder im Hinblick auf die Betriebskosten (verschlechterte Leistung aufgrund von Überbelegung) oder auf die verschwendeten AWS-Ausgaben (aufgrund von Überversorgung).

Wenn Sie planen, dass Ressourcen für Bedarf und Bereitstellung geändert werden können, denken Sie auch an die Nutzungsmuster, die Zeit für die Bereitstellung neuer Ressourcen und die Vorhersehbarkeit des Bedarfsmusters. Stellen Sie beim Verwalten des Bedarfs sicher, dass Ihre Warteschlange oder Ihr Puffer korrekt dimensioniert ist und Sie in der erforderlichen Zeit auf den Workload-Bedarf reagieren.

Optimierung im Laufe der Zeit

Im Zuge der Veröffentlichung neuer Services und Funktionen durch AWS empfiehlt es sich, dass Sie Ihre bestehenden Entscheidungen zur Architektur überdenken, um sicherzustellen, dass diese weiterhin so kostengünstig wie möglich sind. Wenn sich Ihre Anforderungen ändern, zögern Sie nicht, und nehmen Sie Ressourcen, ganze Services und Systeme, die Sie nicht mehr benötigen, außer Betrieb.

Durch die Implementierung neuer Funktionen oder Ressourcentypen können Sie Ihren Workload inkrementell optimieren und gleichzeitig den Aufwand für die Implementierung der Änderung minimieren. Dadurch wird die Effizienz im Laufe der Zeit kontinuierlich verbessert und sichergestellt, dass Sie stets die aktuellste Technologie nutzen, um die Betriebskosten zu senken. Sie können mit neuen Services auch Komponenten des Workloads ersetzen oder ihm neue Komponenten hinzufügen. Dies kann zu erheblichen Effizienzsteigerungen führen. Daher ist es wichtig, Ihren Workload regelmäßig zu überprüfen und neue Services und Funktionen zu implementieren.

In der folgenden Frage geht es um Überlegungen zur Kostenoptimierung.

KOSTEN 10: Wie können Sie neue Services bewerten?

Im Zuge der Veröffentlichung neuer Services und Funktionen durch AWS empfiehlt es sich, dass Sie Ihre bestehenden Entscheidungen zur Architektur überdenken, um sicherzustellen, dass diese weiterhin so kostengünstig wie möglich sind.

Wenn Sie Ihre Bereitstellungen regelmäßig überprüfen, sollten Sie auch bewerten, wie Sie mit neueren Services möglicherweise Geld sparen können. Mit Amazon Aurora on RDS können Sie beispielsweise die Kosten für relationale Datenbanken reduzieren. Wenn Sie serverlose Technologie wie Lambda verwenden, müssen Sie Instances nicht mehr betreiben und verwalten, um Code auszuführen.

Ressourcen

Weitere Informationen zu bewährten Methoden für die Kostenoptimierung finden Sie in den folgenden Ressourcen.

Dokumentation

- AWS-Dokumentation

Whitepaper

- Säule „Kostenoptimierung“

Nachhaltigkeit

Bei der Säule „Nachhaltigkeit“ geht es um Auswirkungen auf die Umwelt, insbesondere um Energieverbrauch und -effizienz, da diese wichtige Faktoren für Architekten sind, die ihre direkten Aktionen zur Reduzierung des Ressourcenverbrauchs beeinflussen. Verbindliche Leitlinien zur Implementierung finden Sie im Whitepaper zur Säule der Nachhaltigkeit.

Themen

- Designprinzipien
- Definition

- Bewährte Methoden

Designprinzipien

Es gibt sechs Designprinzipien für die Nachhaltigkeit in der Cloud:

- Verstehen Sie Ihre Auswirkungen: Messen Sie die Auswirkungen Ihrer Cloud-Workloads und modellieren Sie diese Auswirkungen für die Zukunft. berücksichtigen Sie dabei alle relevanten Faktoren, darunter Auswirkungen durch die Verwendung Ihrer Produkte durch Kunden sowie solche durch deren Außerbetriebnahme und Entsorgung. Vergleichen Sie den produktiven Output mit den Gesamtauswirkungen Ihrer Cloud-Workloads, indem Sie die für jede Arbeitseinheit erforderlichen Ressourcen und die damit verbundenen Emissionen ermitteln. Anhand dieser Daten können Sie Leistungskennzahlen (KPIs) einrichten, Möglichkeiten zur Verbesserung der Produktivität bei gleichzeitiger Reduzierung der Auswirkungen finden und berechnen, wie sich vorgeschlagene Änderungen im Zeitverlauf auswirken werden.
- Legen Sie Nachhaltigkeitsziele fest: Formulieren Sie für alle Cloud-Workloads langfristige Nachhaltigkeitsziele wie etwa die Reduzierung der pro Transaktion erforderlichen Computing- und Speicherressourcen. Modellieren Sie den ROI von Verbesserungen in Bezug auf die Nachhaltigkeit vorhandener Workloads. Stellen Sie den Besitzern die nötigen Ressourcen zur Verfügung, um in Nachhaltigkeitsziele investieren zu können. Planen Sie wachstumsorientiert und gestalten Sie Ihre Workloads so, dass das Wachstum mit geringeren Auswirkungen einhergeht – gemessen in einer sinnvollen Einheit, etwa pro Benutzer oder pro Transaktion. Ziele helfen Ihnen, die allgemeinen Nachhaltigkeitsziele Ihres Unternehmens oder Ihrer Organisation zu erreichen, Rückschritte zu identifizieren und Bereiche mit Verbesserungsmöglichkeiten zu priorisieren.
- Maximieren Sie Ihre Auslastung: Sorgen Sie für Workloads angemessenen Umfangs und nutzen Sie effiziente Designprinzipien, um hohe Auslastung zu gewährleisten und die Energieeffizienz der zugrunde liegenden Hardware so zu maximieren. Zwei Hosts mit 30 % Auslastung sind aufgrund des grundlegenden Energieverbrauchs pro Host weniger effizient als ein Host mit 60 % Auslastung. Gleichzeitig sollten Sie nicht genutzte Ressourcen, Verarbeitungsvorgänge und Speicher beseitigen oder minimieren, um den Gesamtenergieverbrauch für Ihren Workload zu senken.
- Antizipieren und nutzen Sie neue und effizientere Hardware- und Software-Angebote: Unterstützen Sie die Verbesserungen, die Ihre Partner und Lieferanten in früheren Prozessphasen vornehmen, um die Auswirkungen Ihrer Cloud-Workloads zu reduzieren. Achten Sie stets auf neue und effizientere Hardware- und Software-Angebote. Planen Sie für Flexibilität, damit neue effiziente Technologien schnell eingeführt werden können.

- Verwenden Sie verwaltete Services: Die gemeinsame Nutzung von Services über eine breite Kundenbasis hinweg hilft dabei, die Ressourcennutzung zu maximieren und dadurch den Umfang der Infrastruktur zu verringern, der für die Unterstützung Ihrer Cloud-Workloads erforderlich ist. So können Kunden die Auswirkungen allgemeiner Rechenzentrumskomponenten wie Energieversorgung und Netzwerk teilen, indem sie Workloads zur AWS Cloud migrieren und verwaltete Services einführen, z. B. AWS Fargate für Serverless-Container. Dabei kann AWS skalierbar ausgeführt werden und ist für einen effizienten Betrieb verantwortlich. Verwenden Sie verwaltete Services, die dabei helfen können, Ihre Auswirkungen zu verringern, wie etwa die automatische Verschiebung selten genutzter Daten in „kalte“ Speicher mit Amazon S3 Lifecycle-Konfigurationen oder Amazon EC2 Auto Scaling, um Ihre Kapazitäten an die jeweiligen Anforderungen anzupassen.
- Reduzieren Sie die nachgelagerten Auswirkungen Ihrer Cloud-Workloads: Senken Sie den Energie- oder Ressourcenverbrauch für die Nutzung Ihrer Services. Reduzieren oder beseitigen Sie die Erfordernis einer Geräteaktualisierung auf Kundenseite, wenn sie Ihre Services nutzen möchten. Verwenden Sie in Ihren Tests Gerätefarmen, um die zu erwartenden Auswirkungen zu verstehen, und führen Sie Tests mit Kunden durch, um die tatsächlichen Auswirkungen der Nutzung Ihrer Services zu erkennen.

Definition

Es gibt sechs bewährte Methoden für die Nachhaltigkeit in der Cloud:

- Auswahl von Regionen
- Verhaltensmuster von Benutzern
- Software- und Architekturmuster
- Datenmuster
- Hardwaremuster
- Entwicklungs- und Bereitstellungsprozess

Nachhaltigkeit in der Cloud ist ein kontinuierliches Bestreben, das sich in erster Linie auf die Reduzierung des Energieverbrauchs und die Effizienz aller Komponenten eines Workloads konzentriert. Dazu muss der maximale Nutzen aus den bereitgestellten Ressourcen gezogen und die insgesamt erforderlichen Ressourcen müssen minimiert werden. Diese Bemühung kann von der anfänglichen Auswahl einer effizienten Programmiersprache, der Einführung moderner Algorithmen, der Nutzung effizienter Datenspeichertechniken, der Bereitstellung einer korrekt dimensionierten

und effizienten Recheninfrastruktur bis hin zur Minimierung der Anforderungen an leistungsstarke Endbenutzerhardware reichen.

Bewährte Methoden

Themen

Auswahl von Regionen

Wählen Sie die Regionen, in denen Sie Ihre Workloads implementieren, anhand Ihrer geschäftlichen Anforderungen und Ihrer Nachhaltigkeitsziele aus.

In der folgenden Frage geht es um Überlegungen zur Nachhaltigkeit. (Eine Liste der Fragen und bewährten Methoden zur Nachhaltigkeit finden Sie im Anhang.)

SUS 1: Wie wählen Sie Regionen aus, um Ihre Nachhaltigkeitsziele zu unterstützen?
Wählen Sie Regionen in der Nähe von Amazon-Projekten für erneuerbare Energien aus. Es sollte sich um Regionen handeln, in denen das Stromnetz nachweislich geringere Kohlendioxidemissi onen generiert als andere Standorte (oder Regionen).

Verhaltensmuster von Benutzern

Die Art und Weise, wie Benutzer Ihre Workloads und andere Ressourcen nutzen, kann Sie bei der Identifizierung von Verbesserungen unterstützen, um Nachhaltigkeitsziele zu erreichen. Skalieren Sie Ihre Infrastruktur, um die Benutzerlast kontinuierlich anzupassen. Sorgen Sie dafür, dass zur Unterstützung der Benutzer stets nur die mindestens erforderlichen Ressourcen bereitgestellt

werden. Richten Sie Service-Levels an den Kundenanforderungen aus. Positionieren Sie Ressourcen so, dass die für ihre Nutzung erforderlichen Netzwerkkapazitäten begrenzt werden. Entfernen Sie vorhandene, nicht verwendete Komponenten. Identifizieren Sie erstellte, aber nicht verwendete Komponenten und beenden Sie ihre Generierung. Stellen Sie Teammitgliedern Geräte zur Verfügung, die ihre Anforderungen bei geringstmöglichen Auswirkungen auf die Nachhaltigkeit erfüllen.

In der folgenden Frage geht es um diese Überlegungen zur Nachhaltigkeit:

SUS 2: Wie können Sie Verhaltensmuster von Benutzern zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Die Art und Weise, wie Benutzer Ihre Workloads und andere Ressourcen nutzen, kann Sie bei der Identifizierung von Verbesserungen unterstützen, um Nachhaltigkeitsziele zu erreichen. Skalieren Sie Ihre Infrastruktur, um die Benutzerlast kontinuierlich anzupassen. Sorgen Sie dafür, dass zur Unterstützung der Benutzer stets nur die mindestens erforderlichen Ressourcen bereitgestellt werden. Richten Sie Service-Levels an den Kundenanforderungen aus. Positionieren Sie Ressourcen so, dass die für ihre Nutzung erforderlichen Netzwerkkapazitäten begrenzt werden. Entfernen Sie vorhandene, nicht verwendete Komponenten. Identifizieren Sie erstellte, aber nicht verwendete Komponenten und beenden Sie ihre Generierung. Stellen Sie Teammitgliedern Geräte zur Verfügung, die ihre Anforderungen bei geringstmöglichen Auswirkungen auf die Nachhaltigkeit erfüllen.

Skalieren der Infrastruktur anhand der Benutzerlast: Identifizieren Sie Zeiträume mit geringer oder gar keiner Nutzung und skalieren Sie Ressourcen, um überschüssige Kapazitäten zu entfernen und die Effizienz zu verbessern.

Ausrichten von SLAs an Nachhaltigkeitszielen: Definieren und aktualisieren Sie Service Level Agreements (SLAs), darunter die Zeiträume für Verfügbarkeit und Datenaufbewahrung, um den Ressourcenaufwand für Ihre Workloads zu minimieren und gleichzeitig geschäftliche Anforderungen weiter erfüllen zu können.

Beenden der Erstellung und Wartung nicht verwendeter Komponenten: Analysieren Sie Anwendungskomponenten (wie vorab kompilierte Berichte, Datensätze und statische Bilder) sowie Zugriffsmuster für Komponenten, um Redundanzen, eine zu geringe Auslastung und mögliche Kandidaten für die Außerbetriebnahme zu identifizieren. Konsolidieren Sie generierte Komponenten mit redundanten Inhalten (z. B. monatliche Berichte mit sich überschneidenden oder gemeinsam

genutzten Datensätzen und Ausgaben), um für duplizierte Ausgaben genutzte Ressourcen zu eliminieren. Deaktivieren Sie nicht verwendete Komponenten (z. B. Bilder von Produkten, die nicht mehr verkauft werden), um genutzte Ressourcen freizugeben und die Zahl der Ressourcen zu reduzieren, die zur Unterstützung von Workloads verwendet werden.

Optimieren der geografischen Platzierung von Workloads für Benutzerstandorte: Analysieren Sie Netzwerkzugriffsmuster, um zu erkennen, aus welchen geographischen Regionen Ihre Kunden Verbindungen herstellen. Wählen Sie Regionen und Services im Hinblick auf die Reduzierung der Distanz für den Netzwerkdatenverkehr aus, um die Zahl der Netzwerkressourcen zu verringern, die zur Unterstützung von Workloads benötigt werden.

Optimieren von Ressourcen für Teammitglieder im Hinblick auf die ausgeführten Aktivitäten: Optimieren Sie die Ressourcen, die Teammitgliedern zur Verfügung gestellt werden, um negative Auswirkungen auf die Nachhaltigkeit zu minimieren und gleichzeitig ihre Anforderungen zu erfüllen. Beispielsweise können Sie komplexe Vorgänge wie Rendering und Kompilierung auf intensiv genutzten, geteilten Cloud-Desktops statt auf weniger ausgelasteten Einzelbenutzersystemen mit hohem Energieverbrauch ausführen.

Software- und Architekturmuster

Implementieren Sie Muster für den Lastausgleich und die Wahrung einer konsistent hohen Nutzung der bereitgestellten Ressourcen, um die Zahl der genutzten Ressourcen zu minimieren. Komponenten werden möglicherweise aufgrund von Änderungen des Benutzerverhaltens über die Zeit nicht mehr genutzt. Prüfen Sie Muster und Architekturen, um nicht ausreichend genutzte Komponenten zu konsolidieren und so die Nutzung insgesamt zu erhöhen. Nehmen Sie Komponenten außer Betrieb, die nicht mehr benötigt werden. Identifizieren Sie die Leistung Ihrer Workload-Komponenten und optimieren Sie die Komponenten, die die meisten Ressourcen verbrauchen. Achten Sie auf die Geräte, mit denen Ihre Kunden auf Ihre Services zugreifen, und implementieren Sie Muster, um den Bedarf für Geräte-Upgrades zu minimieren.

In den folgenden Fragen geht es um Überlegungen zur Nachhaltigkeit:

SUS 3: Wie können Sie Software- und Architekturmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Implementieren Sie Muster für den Lastausgleich und die Wahrung einer konsistent hohen Nutzung der bereitgestellten Ressourcen, um die Zahl der genutzten Ressourcen zu minimiere n. Komponenten werden möglicherweise aufgrund von Änderungen des Benutzerverhaltens

SUS 3: Wie können Sie Software- und Architekturmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

über die Zeit nicht mehr genutzt. Prüfen Sie Muster und Architekturen, um nicht ausreichend genutzte Komponenten zu konsolidieren und so die Nutzung insgesamt zu erhöhen. Nehmen Sie Komponenten außer Betrieb, die nicht mehr benötigt werden. Identifizieren Sie die Leistung Ihrer Workload-Komponenten und optimieren Sie die Komponenten, die die meisten Ressourcen verbrauchen. Achten Sie auf die Geräte, mit denen Ihre Kunden auf Ihre Services zugreifen, und implementieren Sie Muster, um den Bedarf für Geräte-Upgrades zu minimieren.

Optimieren von Software und Architektur für asynchrone und geplante Aufträge: Verwenden Sie effiziente Softwaredesigns und Architekturen, um die Zahl der für einzelne Arbeitseinheiten im Durchschnitt benötigten Ressourcen zu minimieren. Implementieren Sie Mechanismen für die gleichmäßige Nutzung von Komponenten, um die Zahl der Ressourcen zu reduzieren, die zwischen Aufgaben nicht genutzt werden, und die Auswirkungen von Lastspitzen zu minimieren.

Entfernen von Workload-Komponenten mit geringer oder keiner Nutzung oder Faktorwechsel: Überwachen Sie die Workload-Aktivität, um Änderungen bei der Nutzung einzelner Komponenten über die Zeit zu erkennen. Entfernen Sie ungenutzte Komponenten, die nicht mehr benötigt werden. Setzen Sie wenig genutzte Ressourcen neu ein, um die Verschwendung von Ressourcen zu begrenzen.

Optimieren von Codebereichen, die die meiste Zeit oder die meisten Ressourcen verbrauchen: Überwachen Sie die Workload-Aktivität, um die Anwendungskomponenten zu identifizieren, die die meisten Ressourcen verbrauchen. Optimieren Sie den Code, der innerhalb dieser Komponenten ausgeführt wird, um die Ressourcennutzung zu minimieren und die Leistung zu maximieren.

Optimieren der Auswirkungen auf Geräte und Ausrüstung von Kunden: Identifizieren Sie die Geräte und Einrichtungen, mit denen Ihre Kunden Ihre Services nutzen, ihren voraussichtlichen Lebenszyklus und die finanziellen und nachhaltigkeitsbezogenen Auswirkungen der Ersetzung dieser Komponenten. Implementieren Sie Softwaremuster und Architekturen, die es für Kunden unnötig machen, Geräte zu ersetzen oder ihre Ausrüstung zu aktualisieren. Implementieren Sie beispielsweise neue Funktionen, die Code verwenden, der mit älterer Hardware und älteren Betriebssystemversionen abwärtskompatibel ist, oder gestalten Sie die Größe von Nutzlasten so, dass sie die Speicherkapazitäten der Zielgeräte nicht überschreiten.

Verwenden von Softwaremustern und Architekturen, die Datenzugriffs- und Speichermuster optimal unterstützen: Identifizieren Sie, wie Daten in Ihrem Workload verwendet, von Benutzern genutzt,

übertragen und gespeichert werden. Wählen Sie Technologien aus, die die Anforderungen an Datenverarbeitung und -speicherung minimieren.

Datenmuster

Implementieren Sie Muster für den Lastausgleich und die Wahrung einer konsistent hohen Nutzung der bereitgestellten Ressourcen, um die Zahl der genutzten Ressourcen zu minimieren. Komponenten werden möglicherweise aufgrund von Änderungen des Benutzerverhaltens über die Zeit nicht mehr genutzt. Prüfen Sie Muster und Architekturen, um nicht ausreichend genutzte Komponenten zu konsolidieren und so die Nutzung insgesamt zu erhöhen. Nehmen Sie Komponenten außer Betrieb, die nicht mehr benötigt werden. Identifizieren Sie die Leistung Ihrer Workload-Komponenten und optimieren Sie die Komponenten, die die meisten Ressourcen verbrauchen. Achten Sie auf die Geräte, mit denen Ihre Kunden auf Ihre Services zugreifen, und implementieren Sie Muster, um den Bedarf für Geräte-Upgrades zu minimieren.

In der folgenden Frage geht es um Überlegungen zur Nachhaltigkeit:

SUS 4: Wie können Sie Datenzugriffs- und -nutzungsmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Implementieren Sie Verfahren für die Datenverwaltung, die den zur Unterstützung Ihres Workloads bereitgestellten Speicher und die für dessen Nutzung erforderlichen Ressourcen reduzieren. Identifizieren Sie Ihre Daten und verwenden Sie Speichertechnologien und Konfigurationen, die den Unternehmenswert und die Nutzung der Daten optimal unterstützen. Verschieben Sie die Daten während des Lebenszyklus zu effizienteren Speichern mit geringerer Leistung, wenn die Anforderungen abnehmen. Löschen Sie Daten, die nicht mehr benötigt werden.

Implementieren einer Richtlinie für die Klassifizierung von Daten: Klassifizieren Sie Daten, um ihre Bedeutung für geschäftliche Ergebnisse zu verstehen. Nutzen Sie diese Informationen, um festzulegen, wann Daten in einen energieeffizienteren Speicher übertragen oder auf sichere Weise gelöscht werden können.

Verwenden von Technologien, die Datenzugriff und Speichermuster unterstützen: Nutzen Sie einen Speicher, der den Zugriff auf Ihre Daten und ihre Speicherung jeweils optimal unterstützt, um die Zahl der bereitgestellten Ressourcen zu minimieren und gleichzeitig den Workload zu unterstützen. Beispielsweise verbrauchen SSD-Laufwerke mehr Energie als magnetische Laufwerke und sollten nur für aktive Datenanwendungsfälle eingesetzt werden. Verwenden Sie für Daten, auf die nicht häufig zugegriffen wird, einen energieeffizienten Archivierungsspeicher.

Verwenden von Lebenszyklusrichtlinien zum Löschen nicht notwendiger Daten: Verwalten Sie den Lebenszyklus aller Daten und setzen Sie automatisch Löschfristen durch, um die Speicheranforderungen Ihres Workloads insgesamt zu minimieren.

Minimieren übermäßiger Bereitstellungen im Blockspeicher: Erstellen Sie zur Minimierung des insgesamt bereitgestellten Speichers Blockspeicher mit Größenzuweisungen entsprechend dem jeweiligen Workload. Verwenden Sie elastische Volumes, um den Speicher bei wachsenden Datenmengen erweitern zu können, ohne die Größe des an Computing-Ressourcen angefügten Speichers ändern zu müssen. Überprüfen Sie elastische Volumes regelmäßig und verkleinern Sie zu große Volumes, um sie an den aktuellen Datenumfang anzupassen.

Entfernen nicht benötigter oder redundanter Daten: Duplizieren Sie Daten nur wie notwendig, um den insgesamt genutzten Speicher zu minimieren. Verwenden Sie Backup-Technologien, die Daten auf Datei- und Blockebene deduplizieren. Verwenden Sie Konfigurationen mit Redundant Array of Independent Drives (RAID) nur, wenn dies zur Erfüllung von SLAs notwendig ist.

Verwenden geteilter Dateisysteme oder Objektspeicher für den Zugriff auf allgemeine Daten: Verwenden Sie geteilten Speicher und zentrale Datenquellen, um Datenduplizierungen zu vermeiden und den Gesamtspeicherbedarf des Workloads zu reduzieren. Rufen Sie Daten nur wie notwendig aus dem geteilten Speicher ab. Trennen Sie nicht genutzte Volumes, um Ressourcen freizugeben. Minimieren Sie Datenübertragungen über Netzwerke hinweg. Verwenden Sie stattdessen einen geteilten Speicher und greifen Sie über regionale Datenspeicher auf die Daten zu, um die Zahl der Netzwerkressourcen zu minimieren, die für Datenübertragungen für Ihren Workload benötigt werden.

Sichern von Daten nur in dem Fall, dass ihre erneute Erstellung schwierig ist: Sichern Sie zur Minimierung der Speichernutzung nur Daten, die einen Unternehmenswert besitzen oder zur Erfüllung von Compliance-Anforderungen benötigt werden. Prüfen Sie Backup-Richtlinien und vermeiden Sie einen flüchtigen Speicher, der in einem Wiederherstellungsszenario keinen Wert bietet.

Hardwaremuster

Suchen Sie nach Möglichkeiten, die Auswirkungen auf die Nachhaltigkeit Ihrer Workloads durch Änderungen der Methoden für die Hardwareverwaltung zu reduzieren. Minimieren Sie den Umfang der für die Bereitstellung erforderlichen Hardware und wählen Sie die jeweils effizienteste Hardware für den jeweiligen Workload aus.

In der folgenden Frage geht es um Überlegungen zur Nachhaltigkeit:

SUS 5: Wie können Hardwareverwaltung und Nutzungsverfahren Ihre Nachhaltigkeitsziele unterstützen?

Suchen Sie nach Möglichkeiten, die Auswirkungen auf die Nachhaltigkeit Ihrer Workloads durch Änderungen der Methoden für die Hardwareverwaltung zu reduzieren. Minimieren Sie den Umfang der für die Bereitstellung erforderlichen Hardware und wählen Sie die jeweils effizienteste Hardware für den jeweiligen Workload aus.

Verwenden der geringstmöglichen Menge an Hardware zur Erfüllung Ihrer Anforderungen: Mit den Möglichkeiten der Cloud können Sie häufige Änderungen für Ihre Workload-Implementierungen ausführen. Aktualisieren Sie bereitgestellte Komponenten, wenn sich Ihre Anforderungen ändern.

Überwachen Sie kontinuierlich die Einführung neuer Instance-Typen und nutzen Sie Verbesserungen bei der Energieeffizienz, einschließlich Instance-Typen, die zur Unterstützung spezifischer Workloads bestimmt sind, wie z. B. Machine-Learning-Trainings und -Inferenzen und Videotranskodierung.

Verwenden von Instance-Typen mit den geringsten Auswirkungen: Mit verwalteten Services geht die Verantwortung für die Wahrung einer hohen durchschnittlichen Nutzung und die Optimierung der Nachhaltigkeit der bereitgestellten Hardware auf AWS über. Mit verwalteten Services können Sie die nachhaltigkeitsbezogenen Auswirkungen des Service über alle Mandanten des Service verteilen und so Ihren Beitrag verringern.

Optimieren der GPU-Nutzung: Grafikverarbeitungseinheiten (Graphics Processing Units, GPUs) können sehr viel Energie verbrauchen. Zahlreiche GPU-Workloads sind hoch variabel, z. B. Rendern, Transkodieren sowie Machine-Learning-Trainings und -Modellierungen. Führen Sie GPU-Instances nur für die benötigte Zeit aus und automatisieren Sie ihre Außerbetriebnahme, wenn sie nicht benötigt werden, um den Ressourcenverbrauch zu minimieren.

Entwicklungs- und Bereitstellungsmuster

Reduzieren Sie nachhaltigkeitsbezogene Auswirkungen, indem Sie Ihre Entwicklungs-, Test- und Bereitstellungsmethoden ändern.

In der folgenden Frage geht es um Überlegungen zur Nachhaltigkeit:

SUS 6: Wie können Ihre Entwicklungs- und Bereitstellungsprozesse Ihre Nachhaltigkeitsziele unterstützen?

Reduzieren Sie nachhaltigkeitsbezogene Auswirkungen, indem Sie Ihre Entwicklungs-, Test- und Bereitstellungsmethoden ändern.

Einführen von Methoden, die schnelle Verbesserungen für die Nachhaltigkeit ermöglichen: Testen und validieren Sie potenzielle Verbesserungen, bevor Sie sie für die Produktion bereitstellen. Berücksichtigen Sie die Testkosten bei der Berechnung des potenziellen zukünftigen Nutzens einer Verbesserung. Entwickeln Sie kostengünstige Testmethoden, um kleine Verbesserungen einzuführen.

Konstantes Aktualisieren Ihres Workloads: Aktuelle Betriebssysteme, Bibliotheken und Anwendungen können die Workload-Effizienz verbessern und die Nutzung effizienterer Technologien unterstützen. Eine aktuelle Software kann darüber hinaus Funktionen für eine genauere Messung der Auswirkungen Ihres Workloads bereitstellen, da die Anbieter mit ihrer Software ebenfalls Nachhaltigkeitsziele erfüllen müssen.

Höhere Auslastung von Entwicklungsumgebungen: Verwenden Sie Automatisierung und Infrastructure-as-Code, um Vorproduktionsumgebungen bei Bedarf in Betrieb und bei Nichtverwendung wieder außer Betrieb zu nehmen. Eine typische Vorgehensweise besteht in der Planung von Verfügbarkeitszeiten, die mit den Arbeitszeiten der Entwicklungsteams übereinstimmen. Der Ruhezustand ist ein nützliches Tool, um den aktuellen Status beizubehalten und Instances nur zu aktivieren, wenn sie benötigt werden. Verwenden Sie Instance-Typen mit Burst-Kapazität, Spot-Instances, Elastic Database-Services, Containern und anderen Technologien, um Entwicklungs- und Testkapazität an die Nutzung anzupassen.

Verwenden verwalteter Gerätefarmen für Tests verwenden: Verwaltete Gerätefarmen verteilen die nachhaltigkeitsbezogenen Auswirkungen der Hardwarefertigung und der Ressourcennutzung über zahlreiche Beteiligte. Verwaltete Gerätefarmen stellen verschiedene Gerätetypen bereit, unterstützen auch ältere und weniger verbreitete Hardware und vermeiden nachhaltigkeitsbezogene Auswirkungen durch unnötige Geräte-Upgrades seitens Kunden.

Ressourcen

Werfen Sie einen Blick auf die folgenden Ressourcen, um mehr über unsere bewährten Methoden für die Nachhaltigkeit zu erfahren.

Whitepaper

- Säule „Nachhaltigkeit“

Video

- The Climate Pledge

Die Überprüfung

Architekturen müssen nach einheitlichen Gesichtspunkten überprüft werden. Wenn dabei niemand an den Pranger gestellt wird, ist eine Voraussetzung für tief schürfende Analysen gegeben. Der Prozess sollte nicht schwerfällig sein (Stunden, nicht Tage) und als Konversation angelegt sein, nicht als Prüfung. Architekturen werden überprüft, um festzustellen, ob kritische Mängel vorliegen, gegen die etwas unternommen werden muss – oder um festzustellen, ob bestimmte Bereiche nachgebessert werden können. Am Ende der Überprüfung stehen Maßnahmen, die dem Kunden, der mit dem Workload arbeitet, ein angenehmeres Erlebnis ermöglichen.

Wie bereits im Abschnitt "Architektur-Überlegungen" angesprochen, ist es in Ihrem Interesse, dass jedes Teammitglied Verantwortung für die Qualität der Architektur übernimmt. Wir empfehlen, dass die Teammitglieder, die die Architektur entwerfen, mit Hilfe des Well-Architected Framework ihre Architektur fortlaufend überprüfen, anstatt eine formelle Überprüfungsbesprechung anzusetzen. Findet die Überprüfung fortlaufend statt, können Ihre Teammitglieder parallel mit der Entwicklung der Architektur Antworten aktualisieren und mit jeder neuen Funktion die Architektur verbessern.

Das AWS Well-Architected Framework ist ähnlich aufgebaut wie der interne AWS-Prozess zur Überprüfung von Systemen und Services. Der architektonische Ansatz wird beeinflusst von konzeptionellen Grundsätzen und Fragen, die sicherstellen, dass Bereiche nicht vernachlässigt werden, die häufig in der Ursachenanalyse auftauchen. Tritt an einem internen System, AWS-Service oder bei einem Kunden ein schwerwiegendes Problem auf, untersuchen wir die Ursachenanalyse auf Verbesserungsmöglichkeiten für unsere Überprüfungsprozesse.

Die Überprüfungen müssen an wichtigen Meilensteinen des Produktzyklus erfolgen – früh in der Entwurfsphase, um einseitige Türen zu vermeiden, an denen schwer nachzubessern ist. Und zuletzt schließlich kurz vor dem Go-Live. Viele Entscheidungen können rückgängig gemacht werden; es gibt zwei Möglichkeiten. Für diese Entscheidungen reicht ein schlanker Prozess. Gibt es nur eine Möglichkeit, kann diese nur schwer oder gar nicht rückgängig gemacht werden und muss genauer inspiziert werden, bevor sie gewählt wird. Nachdem Sie in Produktion gehen, verändert sich Ihr Workload weiter, da neue Funktionen hinzukommen und Sie Technologieimplementierungen anpassen. Die Architektur eines Workloads verändert sich mit der Zeit. Treffen Sie durchdachte Hygienemaßnahmen, um zu verhindern, dass die Qualität seiner architektonischen Merkmale im Zuge der Weiterentwicklung nachlässt. Wenn Sie an der Architektur signifikante Änderungen vornehmen, müssen Sie bestimmte Hygieneprozesse befolgen, z. B. eine Überprüfung nach dem Well-Architected-Prinzip.

Wenn die Überprüfung als einmalige Momentaufnahme oder unabhängige Messung vorgesehen ist, müssen alle wichtigen Beteiligten in die Konversation eingebunden sein. Häufig ist die Überprüfung der Punkt, an dem einem Team das erste Mal richtig klar wird, was es implementiert hat. Wird der Workload eines anderen Teams überprüft, ist es sinnvoll, mehrere informelle Konversationen über seine Architektur einzuplanen. In diesen Gesprächen erhalten Sie Antworten auf die meisten Fragen. Im Anschluss daran können Sie in ein oder zwei Besprechungen Punkte abklären und ausführlich auf Unklarheiten oder eventuelle Risiken eingehen.

Damit Ihre Besprechungen erfolgreich verlaufen, empfehlen wir folgende Ausstattung:

- Besprechungszimmer mit Whiteboards
- Diagramme und Entwurfsnotizen ausgedruckt auf Papier
- Liste der Fragen, die sich nicht mit herkömmlichen Mitteln beantworten lassen (z. B. „Werden die Daten verschlüsselt?“)

Nach der Überprüfung sollten Sie eine Liste mit Problemen vorliegen haben. Welche Sie priorisieren, hängt vom geschäftlichen Kontext ab. Berücksichtigen Sie auch, wie sich diese Probleme auf die tägliche Arbeit Ihres Teams auswirken. Wenn Sie die Probleme frühzeitig angehen, gewinnen Sie vielleicht Zeit. Zeit, in der Sie geschäftlichen Mehrwert schaffen können, anstatt sich um wiederkehrende Probleme zu kümmern. Während Sie die Probleme aus der Welt schaffen, können Sie Ihre Überprüfung aktualisieren und so verfolgen, wie sich die Architektur verbessert.

Wie hilfreich eine Überprüfung war, zeigt sich erst danach. Neue Teams widersetzen sich möglicherweise zuerst. Sie können Einwänden der Teams entgegnen, indem Sie sie über die Vorteile einer Überprüfung aufklären:

- „Wir sind zu beschäftigt!“ (Häufig im Vorfeld großer Produktstarts zu hören)
  - Wenn ihr euch auf einen großen Launch vorbereitet, sollte der möglichst glatt über die Bühne gehen. Die Überprüfung deckt Schwachstellen auf, die ihr vielleicht übersehen habt.
  - Wir empfehlen, dass ihr früh im Produktzyklus Überprüfungen einbaut, um Risiken aufzudecken und einen Auffangplan auszuarbeiten, der auf die Roadmap für die Feature-Bereitstellung abgestimmt ist.
- „Wir haben nicht die Zeit, um mit den Ergebnissen etwas anzufangen!“ (Oft zu hören, wenn ein unverrückbares Ereignis näher rückt, z. B. eine große Sportveranstaltung, auf das alles ausgerichtet ist)

  - Diese Ereignisse lassen sich nicht verschieben. Wollt ihr da wirklich reingehen, ohne die Risiken eurer Architektur zu kennen? Selbst wenn ihr nicht alle Probleme wegbekommt, könnt ihr euch immer noch mit Playbooks helfen, wenn sie tatsächlich eintreten.
- „Wir möchten nicht, dass andere die Geheimnisse unserer Lösungsimplementierung kennenlernen!“
  - Wenn Sie die Aufmerksamkeit des Teams auf die Fragen im Well-Architected Framework richten, erkennen sie, dass keine der Fragen kommerziell oder technisch sensible Informationen herauszieht.

Wenn Sie mit Teams aus Ihrer Organisation mehrere Überprüfungen durchführen, identifizieren Sie möglicherweise thematische Fragen. So könnte sich beispielsweise herausstellen, dass mehrere Teams in einer bestimmten Säule oder einem bestimmten Themengebiet mehrere zusammenhängende Probleme haben. Werfen Sie einen ganzheitlichen Blick auf all Ihre Überprüfungen und identifizieren Sie Mechanismen, Trainings oder Principal-Engineer-Vorträge, mit deren Hilfe sich diese thematischen Fragen angehen lassen.

Fazit

Das AWS Well-Architected Framework liefert über alle sechs Säulen hinweg bewährte architektonische Methoden für die Entwicklung und den Betrieb zuverlässiger, sicherer, effizienter, kosteneffizienter und nachhaltiger Systeme in der Cloud. Die Fragen aus dem Framework erlauben Ihnen, bestehende und geplante Architekturen zu überprüfen. Außerdem sind darin bewährte AWS-Methoden für die fünf Säulen enthalten. Als fester Bestandteil Ihres Architekturdesigns fördert das Framework stabile und effiziente Systeme. Anschließend können Sie sich auf Ihre funktionalen Anforderungen konzentrieren.

Mitwirkende

Dieses Dokument ist unter der Mitarbeit folgender Personen und Organisationen entstanden:

- Brian Carlson, Operations Lead Well-Architected, Amazon Web Services
- Ben Potter, Security Lead Well-Architected, Amazon Web Services
- Seth Eliot, Reliability Lead Well-Architected, Amazon Web Services
- Eric Pullen, Sr. Solutions Architect, Amazon Web Services
- Rodney Lester, Principal Solutions Architect, Amazon Web Services
- Jon Steele, Sr. Technical Account Manager, Amazon Web Services
- Max Ramsay, Principal Security Solutions Architect, Amazon Web Services
- Callum Hughes, Solutions Architect, Amazon Web Services
- Aden Leirer, Content Program Manager Well-Architected, Amazon Web Services

Weitere Informationen

AWS-Architekturzentrum

AWS Cloud-Compliance

AWS Well-Architected-Partnerprogramm

AWS Well-Architected Tool

AWS Well-Architected-Homepage

Whitepaper zur Säule für die betriebliche Exzellenz

Whitepaper der Säule für Sicherheit

Whitepaper zur Säule der Zuverlässigkeit

Whitepaper zur Säule der Leistungseffizienz

Whitepaper zur Säule der Kostenoptimierung

Whitepaper zur Säule der Nachhaltigkeit

Die Amazon Builders' Library

Dokumentversionen

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

Änderung	Beschreibung	Datum
Kleineres Update	Eine Definition für Grad des Aufwands wurde hinzugefügt und bewährte Methoden im Anhang wurden aktualisiert.	October 20, 2022
Whitepaper aktualisiert	Die Säule „Nachhaltigkeit“ wurde hinzugefügt und Links wurden aktualisiert.	December 2, 2021
Größere Aktualisierung	Die Säule „Nachhaltigkeit“ wurde zum Framework hinzugefügt.	November 20, 2021
Kleineres Update	Nicht inklusive Sprache wurde entfernt.	April 22, 2021
Kleineres Update	Zahlreiche Links wurden repariert.	March 10, 2021
Kleineres Update	Kleinere redaktionelle Änderungen im gesamten Dokument.	July 15, 2020
Updates für das neue Framework	Prüfung und Umformulierung der meisten Fragen und Antworten.	July 8, 2020
Whitepaper aktualisiert	Ergänzung des AWS Well-Architected Tool sowie von Links zu AWS Well-Architected Labs und AWS	July 1, 2019

	Well-Architected Partnern, kleinere Fehlerbehebungen zur Aktivierung mehrerer Sprachversionen des Frameworks.	
Whitepaper aktualisiert	Die meisten Fragen und Antworten wurden noch einmal durchgelesen und umgeschri eben, damit die Fragen jeweils nur ein Thema behandeln. Dabei wurden einige Fragen in mehrere Einzelfragen aufgeteil t. Häufig verwendete Begriffe (Workload, Komponente usw.) wurden definiert. Darstellu ng der Fragen im Textkorpus wurde bearbeitet, um Platz zu schaffen für Erläuterungen.	November 1, 2018
Whitepaper aktualisiert	Fragentext ist nach mehreren Updates einfacher formulier t, Antworten sind standardi siert, und die Lesbarkeit wurde verbessert.	June 1, 2018
Whitepaper aktualisiert	"Betriebliche Exzellenz" wurde vor die anderen Säulen gesetzt und umgeschrieben. Umfasst jetzt die anderen Säulen. Die anderen Säulen wurden aktualisiert, um der Weiterentwicklung von AWS Rechnung zu tragen.	November 1, 2017

Whitepaper aktualisiert	Aktualisierung des Framework . Dieses enthält jetzt die Säule "Betriebliche Exzellenz". Die anderen Säulen wurden überarbeitet und aktualisiert. Dabei wurden Doppelnen nungen ausgeräumt und Erkenntnisse aus Überprüfu ngen bei mehreren Tausend Kunden aufgenommen.	November 1, 2016
Kleinere Updates	Anhang wurde mit aktuellen Amazon CloudWatch Logs Informationen aktualisiert.	November 1, 2015
Erstveröffentlichung	AWS Well-Architected Framework wurde veröffent licht.	October 1, 2015

Anhang: Fragen und bewährte Methoden

Themen

Operational Excellence

Themen

Organisation

Fragen

OPS 1 Wie können Sie Ihre Prioritäten bestimmen?

Alle Beteiligten müssen verstehen, welchen Anteil sie am geschäftlichen Erfolg haben. Setzen Sie sich gemeinsame Ziele, damit Sie die Prioritäten für Ressourcen festlegen können. Dadurch erzielen Ihre Bemühungen den größtmöglichen Nutzen.

Bewährte Methoden

- OPS01-BP01 Bedürfnisse externer Kunden bewerten
- OPS01-BP02 Bedürfnisse interner Kunden bewerten
- OPS01-BP03 Bewerten der Governance-Anforderungen
- OPS01-BP04 Bewerten der Compliance-Anforderungen
- OPS01-BP05 Bewerten der Bedrohungsszenarien
- OPS01-BP06 Bewerten von Kompromissen
- OPS01-BP07 Abwägen von Vorteilen und Risiken

OPS01-BP01 Bedürfnisse externer Kunden bewerten

Binden Sie alle wichtigen Beteiligten ein, einschließlich Geschäfts-, Entwicklungs- und Betriebsteams, um zu bestimmen, welche Bereiche verstärkt auf die Bedürfnisse der externen Kunden ausgerichtet werden müssen. Dadurch wird sichergestellt, dass Sie mit der betrieblichen Unterstützung vertraut sind, die erforderlich ist, um die gewünschten geschäftlichen Ergebnisse zu erzielen.

Gängige Antimuster:

- Sie haben sich entschieden, außerhalb der Kerngeschäftszeiten keinen Kundenservice zu bieten, aber Sie haben dazu keine historischen Supportanfragedaten analysiert. Daher wissen Sie nicht, ob diese Entscheidung Auswirkungen auf Ihre Kunden hat.
- Sie entwickeln eine neue Funktion, haben aber Ihre Kunden nicht miteinbezogen, um herauszufinden, ob die Funktion erwünscht ist und wie sie genau aussehen sollte. Außerdem haben Sie keine Tests durchgeführt, um die Nachfrage und die Methode der Bereitstellung zu validieren.

Vorteile der Einführung dieser bewährten Methode: Kunden, deren Anforderungen erfüllt sind, bleiben mit höherer Wahrscheinlichkeit als Kunden erhalten. Die Bewertung und das Verständnis externer Kundenbedürfnisse liefert die Grundlage dafür, wie Sie Ihre Anstrengungen zur Bereitstellung eines geschäftlichen Mehrwerts priorisieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Kenntnis der geschäftlichen Anforderungen: Der geschäftliche Erfolg basiert auf gemeinsamen Zielen und der Kommunikation zwischen allen Beteiligten, zu denen auch die Teams aus den Bereichen Betriebswirtschaft, Entwicklung und Operationen gehören.

- Überprüfen der geschäftlichen Ziele, Anforderungen und Prioritäten externer Kunden: Führen Sie wichtige Beteiligte zusammen, einschließlich Geschäfts-, Entwicklungs- und Betriebsteams, um die Ziele, Anforderungen und Prioritäten externer Kunden zu besprechen. Dadurch wird sichergestellt, dass Sie mit der betrieblichen Unterstützung vertraut sind, die erforderlich ist, um die gewünschten Geschäfts- und Kundenergebnisse zu erzielen.
- Schaffen eines gemeinsamen Verständnisses: Sorgen Sie dafür, dass alle Beteiligten die Geschäftsfunktionen des Workloads und die Rollen der einzelnen Teams bei den Workload-spezifischen betrieblichen Abläufen kennen. Außerdem sollte bekannt sein, wie diese Faktoren Ihre gemeinsamen Geschäftsziele mit internen und externen Kunden beeinflussen.

Ressourcen

Zugehörige Dokumente:

- AWS Well-Architected Framework-Konzepte – Feedbackschleife

OPS01-BP02 Bedürfnisse interner Kunden bewerten

Binden Sie alle wichtigen Beteiligten ein, einschließlich Geschäfts-, Entwicklungs- und Betriebsteams, um zu bestimmen, welche Bereiche verstärkt auf die Bedürfnisse der internen Kunden ausgerichtet werden müssen. Dadurch wird sichergestellt, dass Sie mit der betrieblichen Unterstützung vertraut sind, die erforderlich ist, um geschäftliche Ergebnisse zu erzielen.

Anhand Ihrer etablierten Prioritäten können Sie dann erkennen, an welchen Stellen die Verbesserungsbemühungen konzentriert werden sollten (z. B. Teamfähigkeiten entwickeln, die Workload-Leistung verbessern, Kosten senken, Runbooks automatisieren oder die Überwachung ausbauen). Wenn sich Anforderungen ändern, aktualisieren Sie Ihre Prioritäten entsprechend.

Gängige Antimuster:

- Sie haben sich entschieden, die Zuweisung von IP-Adressen für Ihre Produktteams zu ändern, um die Netzwerkverwaltung zu vereinfachen. Dabei haben Sie jedoch nicht mit den Mitarbeitern gesprochen. Sie wissen also nicht, welche Auswirkungen diese Änderung auf Ihre Produktteams haben wird.
- Sie implementieren ein neues Entwicklungstool, haben aber Ihre internen Kunden nicht einbezogen, um herauszufinden, ob das Tool benötigt wird oder mit den Abläufen der Kunden kompatibel ist.

- Sie implementieren ein neues Überwachungssystem, haben aber Ihre internen Kunden nicht kontaktiert, um herauszufinden, ob spezifische Überwachungs- oder Berichtsanforderungen vorliegen, die berücksichtigt werden sollten.

Vorteile der Einführung dieser bewährten Methode: Die Bewertung und das Verständnis interner Kundenbedürfnisse liefert die Grundlage dafür, wie Sie Ihre Anstrengungen zur Bereitstellung eines geschäftlichen Mehrwerts priorisieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Kenntnis der geschäftlichen Anforderungen: Der geschäftliche Erfolg basiert auf gemeinsamen Zielen und der Kommunikation zwischen allen Beteiligten, zu denen auch die Teams aus den Bereichen Geschäft, Entwicklung und Betrieb gehören.
  - Überprüfen der geschäftlichen Ziele, Anforderungen und Prioritäten interner Kunden: Führen Sie wichtige Beteiligte zusammen, einschließlich Geschäfts-, Entwicklungs- und Betriebsteams, um die Ziele, Anforderungen und Prioritäten interner Kunden zu besprechen. Dadurch wird sichergestellt, dass Sie mit der betrieblichen Unterstützung vertraut sind, die erforderlich ist, um die gewünschten Geschäfts- und Kundenergebnisse zu erzielen.
  - Übereinstimmendes Verständnis: Sorgen Sie dafür, dass alle Beteiligten die Geschäftsfunktionen des Workloads und die Rollen der einzelnen Teams bei den Workload-spezifischen Betriebsabläufen kennen. Außerdem sollte bekannt sein, wie diese Faktoren Ihre gemeinsamen Geschäftsziele mit internen und externen Kunden beeinflussen.

Ressourcen

Zugehörige Dokumente:

- AWS Well-Architected Framework-Konzepte – Feedbackschleife

OPS01-BP03 Bewerten der Governance-Anforderungen

Stellen Sie sicher, dass Sie sich mit den Richtlinien oder Verpflichtungen Ihres Unternehmens vertraut machen, die bestimmte Schwerpunkte vorgeben oder hervorheben können. Bewerten Sie interne Faktoren wie Unternehmensrichtlinien, Standards und Anforderungen. Stellen Sie sicher, dass Mechanismen vorliegen, um Änderungen an der Governance zu identifizieren. Wenn keine

Governance-Anforderungen festgestellt werden, stellen Sie sicher, dass diese Prüfung mit der erforderlichen Sorgfalt durchgeführt wurde.

Gängige Antimuster:

- Sie werden einer Prüfung unterzogen und aufgefordert, einen Nachweis der Compliance mit der internen Governance zu erbringen. Sie haben keine Ahnung, ob die Compliance-Anforderungen erfüllt werden, da Sie Ihre Compliance-Anforderungen noch gar nicht etabliert haben.
- Sie haben eine Kompromittierung erlitten, die zu finanziellen Verlusten geführt hat. Sie stellen fest, dass die Versicherung, die den finanziellen Verlust gedeckt hätte, von der Implementierung bestimmter Sicherheitskontrollen in Ihrem Unternehmen abhängt, die nicht vorhanden sind und von Ihrer Governance verlangt werden.
- Ihr Administratorkonto wurde kompromittiert. Dies hat zur Folge, dass Ihre Unternehmenswebsite manipuliert und das Vertrauen Ihrer Kunden beschädigt wurde. Ihre interne Governance erfordert die Verwendung von Multi-Faktor-Authentifizierung (MFA), um Administratorkonten zu sichern. Sie haben Ihr Administratorkonto nicht mit MFA gesichert und müssen mit entsprechenden Disziplinarmaßnahmen rechnen.

Vorteile der Einführung dieser bewährten Methode: Die Bewertung und das Verständnis der Governance-Anforderungen, die Ihr Unternehmen auf den Workload anwendet, liefert die Grundlage dafür, wie Sie Ihre Anstrengungen zur Bereitstellung eines geschäftlichen Mehrwerts priorisieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Kenntnis der Governance-Anforderungen: Bewerten Sie interne Governance-Faktoren wie Programm- oder Unternehmensrichtlinien, problem- oder systemspezifische Richtlinien, Standards, Verfahren, Baselines und Richtlinien. Stellen Sie sicher, dass Mechanismen vorliegen, um Änderungen an der Governance zu identifizieren. Wenn keine Governance-Anforderungen festgestellt werden, stellen Sie sicher, dass diese Prüfung mit der erforderlichen Sorgfalt durchgeführt wurde.

Ressourcen

Zugehörige Dokumente:

- AWS Cloud-Compliance

OPS01-BP04 Bewerten der Compliance-Anforderungen

Bewerten Sie externe Faktoren, wie z. B. gesetzliche Compliance-Anforderungen und Branchenstandards, um sicherzustellen, dass Sie sich der Richtlinien oder Verpflichtungen bewusst sind, die einen bestimmten Fokus erfordern oder verstärken können. Wenn keine Compliance-Anforderungen festgestellt werden, stellen Sie sicher, dass diese Prüfung mit der erforderlichen Sorgfalt durchgeführt wurde.

Gängige Antimuster:

- Sie werden einer Prüfung unterzogen und aufgefordert, einen Nachweis der Compliance mit den Branchenvorschriften zu erbringen. Sie haben keine Ahnung, ob die Compliance-Anforderungen erfüllt werden, da Sie Ihre Compliance-Anforderungen noch gar nicht etabliert haben.
- Ihr Administratorkonto wurde kompromittiert. Dies hat zur Folge, dass Kundendaten heruntergeladen und das Vertrauen Ihrer Kunden beschädigt wurde. Die branchenweit anerkannten bewährten Methoden erfordern die Verwendung von MFA, um Administratorkonten zu sichern. Sie haben Ihr Administratorkonto nicht mit MFA gesichert und werden durch Ihre Kunden rechtlich belangt.

Vorteile der Einführung dieser bewährten Methode: Die Bewertung und das Verständnis der Compliance-Anforderungen für Ihren Workload liefern die Grundlage dafür, wie Sie Ihre Anstrengungen zur Bereitstellung eines geschäftlichen Mehrwerts priorisieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Kenntnis der Compliance-Anforderungen: Bewerten Sie externe Faktoren, wie z. B. gesetzliche Compliance-Anforderungen und Branchenstandards, um sicherzustellen, dass Sie sich der Richtlinien oder Verpflichtungen bewusst sind, die einen bestimmten Fokus erfordern oder verstärken können. Wenn keine Compliance-Anforderungen festgestellt werden, stellen Sie sicher, dass die Prüfung mit der erforderlichen Sorgfalt durchgeführt wurde.
  - Kenntnis der gesetzlichen Compliance-Anforderungen: Ermitteln Sie gesetzliche Compliance-Anforderungen, die Sie von Gesetzes wegen erfüllen müssen. Konzentrieren Sie Ihre Anstrengungen auf diese Anforderungen. Beispiele hierfür sind Verpflichtungen aus Datenschutzgesetzen.
    - AWS-Compliance
    - AWS-Compliance-Programme

    - Aktuelle Neuigkeiten zur AWS-Compliance
- Kenntnis der Branchenstandards und bewährten Methoden: Bestimmen Sie Branchenstandards und bewährte Methoden, die für Ihren Workload gelten, z. B. den Payment Card Industry Data Security Standard (PCI DSS). Konzentrieren Sie Ihre Anstrengungen auf diese Anforderungen.
    - AWS-Compliance-Programme
- Kenntnis der internen Compliance-Anforderungen: Ermitteln Sie Compliance-Anforderungen und Best Practices, die von Ihrer Organisation festgelegt werden. Konzentrieren Sie Ihre Anstrengungen auf diese Anforderungen. Beispiele hierfür sind Richtlinien zur Informationssicherheit und Datenklassifizierungsstandards.

Ressourcen

Zugehörige Dokumente:

- AWS Cloud-Compliance
- AWS-Compliance
- Aktuelle Neuigkeiten zur AWS-Compliance
- AWS-Compliance-Programme

OPS01-BP05 Bewerten der Bedrohungsszenarien

Bewerten Sie Bedrohungen für das Unternehmen (z. B. Wettbewerb, Geschäftsrisiken und -verpflichtungen, operative Risiken und Bedrohungen der Informationssicherheit) und pflegen Sie aktuelle Informationen in einem Risikoregister. Berücksichtigen Sie die Auswirkungen von Risiken, wenn Sie bestimmen, auf welche Bereiche die Anstrengungen fokussiert werden sollen.

Das Well-Architected Framework legt den Schwerpunkt auf Lernen, Messen und Verbessern. Es bietet einen konsistenten Ansatz, mit dem Sie Architekturen bewerten und Designs implementieren können, die sich im Laufe der Zeit skalieren lassen. AWS bietet das AWS Well-Architected Tool , mit dem Sie Ihren Ansatz vor der Entwicklung, den Status Ihrer Workloads vor der Produktion und den Status Ihrer Workloads in der Produktion überprüfen können. Sie können sie mit den neuesten bewährten Methoden für die AWS-Architektur vergleichen, den Gesamtstatus Ihrer Workloads überwachen und Einblicke in potenzielle Risiken erhalten.

AWS-Kunden haben auch die Möglichkeit, die Architektur ihrer geschäftskritischen Workloads auf die Einhaltung bewährter AWS-Methoden hin überprüfen zu lassen (Well-Architected Review). Für

Enterprise Support-Kunden kommt auch eine Betriebsüberprüfungin Frage, die ihnen helfen soll, Lücken in ihrem Ansatz für den Betrieb in der Cloud zu identifizieren.

Aufgrund der teamübergreifenden Natur dieser Überprüfungen erhalten Sie ein allgemeines Verständnis Ihrer Workloads und können erkennen, wie Team-Rollen zum Erfolg beitragen. Die bei den Überprüfungen gefundenen Punkte können Ihnen beim Festlegen Ihrer Prioritäten helfen.

AWS Trusted Advisor bietet als Tool Zugriff auf verschiedene wichtige Prüfungen, die Optimierungsempfehlungen ausgeben. Diese Informationen können Ihnen beim Festlegen Ihrer Prioritäten helfen. Kunden mit Business und Enterprise Support erhalten Zugriff auf weitere Prüfungen in den Bereichen Sicherheit, Zuverlässigkeit, Leistung und Kostenoptimierung, die beim Festlegen von Prioritäten noch hilfreicher sind.

Gängige Antimuster:

- Sie verwenden in Ihrem Produkt eine alte Version einer Softwarebibliothek. Ihnen ist nicht bewusst, dass für die Bibliothek Sicherheitsaktualisierungen vorliegen, mit denen Probleme behoben werden, die unbeabsichtigte Auswirkungen auf Ihren Workload haben können.
- Ein Mitbewerber hat soeben eine Version seines Produkts veröffentlicht, in der viele Probleme behoben werden, die Kunden an Ihrem Produkt bemängeln. Die Behebung dieser bekannten Probleme hatte für Sie bisher keine Priorität.
- Regulierungsbehörden nehmen Unternehmen wie Ihres, die nicht den gesetzlichen Compliance-Anforderungen entsprechen, verstärkt ins Visier. Sie haben Ihre ausstehenden Compliance-Anforderungen nicht priorisiert.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die Bedrohungen für Ihr Unternehmen und Ihren Workload identifizieren und verstehen, können Sie bestimmen, welche Bedrohungen angegangen werden müssen, wo die Prioritäten liegen und welche Ressourcen dafür erforderlich sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Bedrohungslandschaft bewerten: Bewerten Sie Bedrohungen für das Unternehmen (z. B. Konkurrenz, Geschäftsrisiken und -verpflichtungen, operative Risiken und Bedrohungen der Informationssicherheit), damit Sie die jeweiligen Auswirkungen berücksichtigen können, wenn Sie bestimmen, auf welche Bereiche die operativen Anstrengungen konzentriert werden sollten.

- Aktuelle AWS-Sicherheitsmitteilungen
- AWS Trusted Advisor
- Verwalten eines Bedrohungsmodells: Erstellen und verwalten Sie ein Bedrohungsmodell, in dem potenzielle Bedrohungen, geplante und vorhandene Maßnahmen und deren Priorität festgehalten werden. Untersuchen Sie, wie wahrscheinlich es ist, dass sich Bedrohungen als Vorfälle äußern, wie hoch die Kosten für die Wiederherstellung nach diesen Vorfällen sind, welche Schäden zu erwarten sind und wie viel es kostet, diese Vorfälle zu verhindern. Überarbeiten Sie die Prioritäten, wenn sich der Inhalt des Bedrohungsmodells ändert.

Ressourcen

Zugehörige Dokumente:

- AWS Cloud-Compliance
- Aktuelle AWS-Sicherheitsmitteilungen
- AWS Trusted Advisor

OPS01-BP06 Bewerten von Kompromissen

Bewerten Sie die Auswirkungen von Kompromissen zwischen konkurrierenden Interessen oder alternativen Ansätzen, um fundiert zu entscheiden, auf welche Bereiche die operativen Anstrengungen konzentriert werden sollten, oder eine geeignete Handlungsweise zu wählen. Beispielsweise kann die Beschleunigung der Markteinführung neuer Funktionen einer Kostenoptimierung vorgezogen werden oder Sie können eine relationale Datenbank für nicht relationale Daten wählen, um die Migration eines Systems zu vereinfachen, anstatt zu einer für Ihren Datentyp optimierten Datenbank zu migrieren und Ihre Anwendung zu aktualisieren.

AWS kann Ihnen helfen, Ihre Teams über AWS und die verfügbaren Services zu schulen, sodass alle Mitarbeiter wissen, welche Auswirkungen ihre Entscheidungen auf Ihren Workload haben können. Bei der Schulung Ihrer Teams sollten Sie die vom AWS Support (AWS Knowledge Center, AWS-Diskussionsforenund AWS Support Center) bereitgestellten Ressourcen und AWS-Dokumentation nutzen, um Ihre Teams zu schulen. Wenn Sie eine Frage zu AWS haben, können Sie sich über das AWS Support Center an den AWS Support wenden.

AWS stellt in der Amazon Builders' Library auch bewährte Methoden und Muster vor, die wir durch den Betrieb von AWS gelernt haben. Eine Vielzahl weiterer nützlicher Informationen finden Sie im AWS-Blog und im offiziellen AWS-Podcast.

Gängige Antimuster:

- Sie verwenden eine relationale Datenbank, um Zeitreihendaten und nicht relationale Daten zu verwalten. Es gibt Datenbankoptionen, die für Ihre verwendeten Datentypen optimiert sind. Sie sind sich der Vorteile aber nicht bewusst, da Sie die Unterschiede zwischen den Lösungsangeboten nicht evaluiert haben.
- Ihre Investoren fordern, dass Sie die Compliance mit Payment Card Industry Data Security Standards (PCI DSS) nachweisen. Sie denken nicht über die möglichen Kompromisse zwischen der Erfüllung dieser Anfrage und der Fortsetzung Ihrer derzeitigen Entwicklungsaktivitäten nach. Stattdessen fahren Sie mit der Entwicklung fort, ohne einen Compliance-Nachweis zu liefern. Ihre Investoren beenden die Unterstützung Ihres Unternehmens, da sie Bedenken bezüglich der Sicherheit Ihrer Plattform und ihrer Investitionen haben.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die Auswirkungen und Konsequenzen Ihrer Entscheidungen verstehen, können Sie die vorhandenen Optionen priorisieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Kompromisse bewerten: Bewerten Sie die Auswirkungen von Kompromissen bei konkurrierenden Interessen, um fundiert zu entscheiden, auf welche Bereiche die operativen Anstrengungen konzentriert werden sollten. So kann beispielsweise die Beschleunigung der Markteinführung neuer Funktionen einen höheren Stellenwert haben als die Kostenoptimierung.
- AWS kann Ihnen helfen, Ihre Teams über AWS und die verfügbaren Services zu schulen, sodass alle Mitarbeiter wissen, welche Auswirkungen ihre Entscheidungen auf Ihren Workload haben können. Bei der Schulung Ihrer Teams sollten Sie die vom AWS Support (AWS Knowledge Center, AWS Discussion Forums und AWS Support Center) bereitgestellten Ressourcen und AWS-Dokumente nutzen. Wenn Sie eine Frage zu AWS haben, können Sie sich über das AWS Support Center an den AWS Support wenden.
- AWS stellt in der Amazon Builders' Library auch bewährte Methoden und Muster vor, die wir durch den Betrieb von AWS gelernt haben. Eine Vielzahl weiterer nützlicher Informationen finden Sie im AWS-Blog und im offiziellen AWS-Podcast.

Ressourcen

Zugehörige Dokumente:

- AWS-Blog
- AWS Cloud-Compliance
- AWS-Diskussionsforen
- AWS-Dokumentation nutzen,
- AWS Knowledge Center
- AWS Support
- AWS Support Center
- Die Amazon Builders' Library
- im offiziellen AWS-Podcast

OPS01-BP07 Abwägen von Vorteilen und Risiken

Wägen Sie die Vorteile und Risiken ab, um fundiert zu entscheiden, auf welche Bereiche die operativen Anstrengungen konzentriert werden sollten. So kann es beispielsweise sinnvoll sein, einen Workload mit noch offenen Problemen bereitzustellen, um den Kunden wichtige neue Funktionen zur Verfügung zu stellen. Es gibt ggf. die Möglichkeit, die damit verbundenen Risiken zu minimieren, oder es ist zu einem bestimmten Zeitpunkt nicht mehr akzeptabel, dass ein Risiko weiterhin bestehen bleibt. In diesem Fall ergreifen Sie Maßnahmen, um das Risikoproblem zu beheben.

Manchmal kann es vorkommen, dass man zu viel Augenmerk auf eine kleine Auswahl von operativen Prioritäten richtet. Gehen Sie langfristig gut ausgewogen vor, um sicherzustellen, dass erforderliche Fähigkeiten entwickelt und Risiken verwaltet werden. Wenn sich Anforderungen ändern, aktualisieren Sie Ihre Prioritäten entsprechend.

Gängige Antimuster:

- Sie haben sich entschieden, eine Bibliothek einzubinden, die „alle nötigen Funktionen" bietet und von einem Ihrer Entwickler „im Internet gefunden" wurde. Sie haben keine Bewertung der Risiken durchgeführt, die die Einführung dieser Bibliothek aus einer unbekannten Quelle bergen kann, und wissen nicht, ob sie Schwachstellen oder schädlichen Code enthält.
- Sie haben sich entschieden, eine neue Funktion zu entwickeln und bereitzustellen, statt ein vorhandenes Problem zu beheben. Sie haben keine Bewertung der Risiken durchgeführt, die das vorhandene Problem in der bereitgestellten Funktion bergen könnte, und wissen nicht, welche Folgen daraus für Ihre Kunden entstehen.
- Sie haben sich entschieden, eine häufig von Kunden angeforderte Funktion nicht bereitzustellen, weil Ihr Compliance-Team unbestimmte Bedenken geäußert hat.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die verfügbaren Vorteile Ihrer Optionen ermitteln und sich der Risiken für Ihr Unternehmen bewusst sind, können Sie fundierte Entscheidungen treffen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Abwägen von Vorteilen und Risiken: Wägen Sie den Nutzen von Entscheidungen gegen die damit einhergehenden Risiken ab.
  - Ermitteln von Vorteilen: Ermitteln Sie die Vorteile auf Basis der geschäftlichen Ziele, Anforderungen und Prioritäten. Zu diesen Prioritäten können beispielsweise eine kurze Markteinführungszeit, Sicherheit, Zuverlässigkeit, Leistung und Kosten zählen.
  - Ermitteln von Risiken: Ermitteln Sie die Risiken auf Basis der geschäftlichen Ziele, Anforderungen und Prioritäten. Zu diesen Prioritäten können beispielsweise eine kurze Markteinführungszeit, Sicherheit, Zuverlässigkeit, Leistung und Kosten zählen.
  - Abwägen von Vorteilen und Risiken und Treffen fundierter Entscheidungen: Ermitteln Sie die Auswirkungen von Vorteilen und Risiken basierend auf den Zielen, Bedürfnissen und Prioritäten Ihrer wichtigsten Beteiligten, zu denen auch die Bereiche Betriebswirtschaft, Entwicklung und Operationen zählen. Bewerten Sie den Wert eines Vorteils anhand der Wahrscheinlichkeit, dass sich das Risiko tatsächlich bewahrheitet, und anhand der Kosten der jeweiligen Auswirkungen. Eine schnellere Markteinführung zu Lasten der Zuverlässigkeit könnte beispielsweise einen Wettbewerbsvorteil bedeuten. Wenn jedoch Probleme mit der Zuverlässigkeit auftreten, kann dies zu einer verringerten Betriebszeit führen.

OPS 2 Wie strukturieren Sie Ihr Unternehmen, um die gewünschten Geschäftsergebnisse zu erzielen?

Ihre Teams müssen ihre Rolle beim Erreichen von Geschäftsergebnissen verstehen. Teams müssen ihre Rollen beim Erfolg anderer Teams verstehen, die Rolle anderer Teams bei ihrem eigenen Erfolg und sie müssen gemeinsame Ziele haben. Wenn sie Verantwortlichkeit, Zuständigkeit und Entscheidungsfindung nachvollziehen können und wissen, wer dazu berechtigt ist, Entscheidungen zu treffen, können ihre Anstrengungen fokussiert und der Nutzen Ihrer Teams maximiert werden.

Bewährte Methoden

- OPS02-BP01 Ressourcen haben feste Besitzer
- OPS02-BP02 Prozesse und Verfahren haben feste Besitzer

- OPS02-BP03 Betriebsaktivitäten haben feste Besitzer, die für ihre Leistung verantwortlich sind
- OPS02-BP04 Teammitglieder wissen, wofür sie verantwortlich sind
- OPS02-BP05 Mechanismen zur Identifizierung von Verantwortlichkeit und Eigentümerschaft sind vorhanden
- OPS02-BP06 Mechanismen zum Anfordern von Ergänzungen, Änderungen und Ausnahmen sind vorhanden:
- OPS02-BP07 Zuständigkeiten zwischen Teams werden vordefiniert oder ausgehandelt:

OPS02-BP01 Ressourcen haben feste Besitzer

Verschaffen Sie sich einen Überblick darüber, wer der Besitzer der einzelnen Anwendungen, Workloads, Plattformen und Infrastrukturkomponenten ist, welchen geschäftlichen Nutzen diese Komponenten bieten und warum diese Zuständigkeit besteht. Auf dem Verständnis des geschäftlichen Werts dieser einzelnen Komponenten und ihrer Unterstützung der Geschäftsergebnisse basieren die jeweils angewendeten Prozesse und Verfahren.

Vorteile der Einführung dieser bewährten Methode: Anhand der Zuständigkeit kann identifiziert werden, wer Verbesserungen genehmigen, diese Verbesserungen implementieren oder beides durchführen kann.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Ressourcen haben feste Besitzer: Definieren Sie, was Zuständigkeit für die Ressourcen-Anwendungsfälle in Ihrer Umgebung bedeutet. Legen Sie Besitzer für Ressourcen fest und dokumentieren Sie diese. Die Angaben sollten mindestens den Namen, die Kontaktinformationen, die Organisation und das Team beinhalten. Speichern Sie Informationen zur Ressourcenzuständigkeit mithilfe von Metadaten wie Tags oder Ressourcengruppen. Verwenden Sie AWS Organizations, um Konten zu strukturieren und Richtlinien zu implementieren, damit Zuständigkeits- und Kontaktinformationen zuverlässig erfasst werden.
  - Definieren von Zuständigkeitsformen und ihrer Zuweisung: Für das Konzept der Zuständigkeit können in Ihrem Unternehmen je nach Anwendungsfall unterschiedliche Definitionen vorliegen. Möglicherweise möchten Sie einen „Workload-Besitzer“ als Person definieren, die für das Risiko und die Haftung bezüglich des Betriebs eines Workloads zuständig und letztendlich dazu berechtigt ist, Entscheidungen über den Workload zu treffen. Sie können die Zuständigkeit auch im Sinne von finanzieller oder administrativer Verantwortung definieren, wenn die Zuständigkeit

mit einer übergeordneten Organisation zusammengeführt wird. Ein Entwickler kann z. B. Besitzer seiner Entwicklungsumgebung und für Vorfälle verantwortlich sein, die im Betrieb dieser Umgebung auftreten. Der jeweilige Produktleiter kann die Verantwortung für die finanziellen Kosten im Zusammenhang mit dem Betrieb der Entwicklungsumgebungen tragen.

- Definieren der Zuständigkeit für eine Organisation, ein Konto, eine Sammlung von Ressourcen oder einzelne Komponenten: Definieren und dokumentieren Sie die Zuständigkeit an einem zugänglichen Ort, der zur Unterstützung der Ermittlung organisiert ist. Aktualisieren Sie Definitionen und Zuständigkeitsdetails, wenn sie sich ändern.
- Erfassen der Zuständigkeit in den Metadaten der Ressourcen: Erfassen Sie die Ressourcenzuständigkeit mithilfe von Metadaten wie Tags oder Ressourcengruppen und geben Sie Zuständigkeits- und Kontaktinformationen an. Verwenden Sie AWS Organizations, um Konten zu strukturieren und sicherzustellen, dass Zuständigkeits- und Kontaktinformationen erfasst werden.

OPS02-BP02 Prozesse und Verfahren haben feste Besitzer

Verschaffen Sie sich einen Überblick darüber, wer für die Definition einzelner Prozesse und Verfahren zuständig ist, warum diese spezifischen Prozesse und Verfahren verwendet werden und warum diese Zuständigkeit besteht. Wenn Sie wissen, warum bestimmte Prozesse und Verfahren verwendet werden, können Sie Verbesserungsmöglichkeiten identifizieren.

Vorteile der Einführung dieser bewährten Methode: Anhand der Zuständigkeit kann identifiziert werden, wer Verbesserungen genehmigen, diese Verbesserungen implementieren oder beides durchführen kann.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Prozesse und Verfahren haben feste Besitzer, die für ihre Definition verantwortlich sind: Dokumentieren Sie die Prozesse und Verfahren, die in Ihrer Umgebung angewendet werden, sowie die Person oder Personen, die für die Definition verantwortlich sind.
  - Identifizieren von Prozessen und Verfahren: Identifizieren Sie die Betriebsaktivitäten, die zur Unterstützung Ihrer Workloads durchgeführt werden. Dokumentieren Sie diese Aktivitäten an einem auffindbaren Ort.
  - Definieren der Zuständigkeit für die Definition eines Prozesses oder Verfahrens: Legen Sie die Person oder Personen fest, die für die Spezifikation einer Aktivität verantwortlich sind. Sie sind

dafür verantwortlich, sicherzustellen, dass die Aktivität von einem ausreichend qualifizierten Teammitglied durchgeführt wird, das die entsprechenden Berechtigungen, Zugriffsrechte und Tools hat. Wenn bei der Durchführung dieser Aktivität Probleme auftreten, sind die zuständigen Teammitglieder dafür verantwortlich, detailliertes Feedback bereitzustellen, das für die Verbesserung der Aktivität erforderlich ist.

- Erfassen der Zuständigkeit in den Metadaten des Aktivitätsartefakts: Verfahren, die in Services wie AWS Systems Manager (durch Dokumente) und AWS Lambda (als Funktionen) automatisiert werden, unterstützen die Erfassung von Metadateninformationen als Tags. Erfassen Sie die Ressourcenzuständigkeit mithilfe von Tags oder Ressourcengruppen und geben Sie Zuständigkeits- und Kontaktinformationen an. Verwenden Sie AWS Organizations, um Markierungsrichtlinien zu erstellen und zu gewährleisten, dass Zuständigkeits- und Kontaktinformationen erfasst werden.

OPS02-BP03 Betriebsaktivitäten haben feste Besitzer, die für ihre Leistung verantwortlich sind

Verschaffen Sie sich einen Überblick darüber, wer für spezifische Aktivitäten in festgelegten Workloads verantwortlich ist und warum diese Zuständigkeit besteht. Wenn Sie wissen, wer für die Durchführung von Aktivitäten verantwortlich ist, können Sie nachvollziehen, wer die Aktivität durchführen, das Ergebnis validieren und dem Besitzer der Aktivität Feedback geben wird.

Vorteile der Einführung dieser bewährten Methode:i Wenn die verantwortliche Person für die Durchführung einer Aktivität bekannt ist, wissen Sie, wer benachrichtigt werden muss, wenn eine Aktion erforderlich ist, und wer die Aktion ausführen, das Ergebnis validieren und dem Besitzer der Aktivität Feedback geben wird.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Betriebsaktivitäten haben feste Besitzer, die für ihre Leistung verantwortlich sind: Erfassen Sie die Verantwortung für die Durchführung von Prozessen und Verfahren in Ihrer Umgebung.
  - Identifizieren von Prozessen und Verfahren: Identifizieren Sie die Betriebsaktivitäten, die zur Unterstützung Ihrer Workloads durchgeführt werden. Dokumentieren Sie diese Aktivitäten an einem auffindbaren Ort.
  - Definieren der Verantwortlichkeit für die Durchführung von Aktivitäten: Legen Sie das Team fest, das für eine Aktivität verantwortlich ist. Stellen Sie sicher, dass die Teammitglieder die Details der Aktivität und die erforderlichen Qualifikationen haben und über die entsprechenden Berechtigungen, Zugriffsrechte und Tools für die Durchführung der Aktivität verfügen. Sie

müssen die Bedingung kennen, unter denen die Aktivität ausgeführt werden soll (z. B. nach einem Ereignis oder gemäß einem Zeitplan). Diese Informationen sollten leicht auffindbar sein, damit Mitglieder Ihrer Organisation herausfinden können, an wen sie sich für bestimmte Anforderungen wenden müssen (Team oder Person).

OPS02-BP04 Teammitglieder wissen, wofür sie verantwortlich sind

Wenn Ihnen die Verantwortlichkeiten Ihrer Rolle bekannt sind und Sie wissen, wie Sie zu Geschäftsergebnissen beitragen, können Sie Ihre Aufgaben entsprechend priorisieren und die Bedeutung Ihrer Rolle nachvollziehen. Auf diese Weise können Teammitglieder Anforderungen erkennen und entsprechend reagieren.

Vorteile der Einführung dieser bewährten Methode: Das Verständnis Ihrer Verantwortlichkeiten wirkt sich auf Ihre Entscheidungen, Ihre Aktionen und die Übergabe von Aktivitäten an die ordnungsgemäßen Besitzer aus.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Sicherstellen, dass Teammitglieder ihre Rollen und Verantwortlichkeiten verstehen: Legen Sie die Rollen und Verantwortlichkeiten von Teammitgliedern fest und stellen Sie sicher, dass sie die Erwartungen ihrer Rolle verstehen. Diese Informationen sollten leicht auffindbar sein, damit Mitglieder Ihrer Organisation herausfinden können, an wen sie sich für bestimmte Anforderungen wenden müssen (Team oder Person).

OPS02-BP05 Mechanismen zur Identifizierung von Verantwortlichkeit und Eigentümerschaft sind vorhanden

Wenn keine Person oder Personen festgelegt sind, gibt es definierte Eskalationsabläufe, um eine Person zu kontaktieren, die berechtigt ist, die fehlende Zuständigkeit zuzuweisen oder die Erfüllung einer Anforderung zu planen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie wissen, wer verantwortlich oder zuständig ist, können Sie sich an das entsprechende Team oder Teammitglied wenden, um eine Anfrage zu stellen oder eine Aufgabe zu übergeben. Das Vorhandensein einer festgelegten Person, die berechtigt ist, Verantwortlichkeiten oder Zuständigkeiten zuzuweisen oder die Erfüllung von Anforderungen zu planen, reduziert das Risiko, dass Aufgaben liegen bleiben oder Anforderungen nicht erfüllt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Mechanismen zur Identifizierung von Verantwortlichkeit und Eigentümerschaft sind vorhanden: Stellen Sie Mitgliedern Ihrer Organisation zugängliche Mechanismen bereit, um Zuständigkeiten und Verantwortlichkeiten zu ermitteln und zuzuordnen. Auf diese Weise können sie bestimmen, an wen sie sich für bestimmte Anforderungen wenden müssen (Team oder Person).

OPS02-BP06 Mechanismen zum Anfordern von Ergänzungen, Änderungen und Ausnahmen sind vorhanden:

Sie können Anfragen an Besitzer von Prozessen, Verfahren und Ressourcen stellen. Treffen Sie fundierte Entscheidungen, um angemessene Anfragen nach einer Bewertung der Vorteile und Risiken zu genehmigen.

Vorteile der Einführung dieser bewährten Methode: Es ist wichtig, dass Mechanismen vorhanden sind, um Ergänzungen, Änderungen und Ausnahmen zur Unterstützung der Aktivitäten von Teams anzufordern. Ohne diese Option kann der aktuelle Zustand die Innovationsfähigkeit einschränken.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mechanismen zum Anfordern von Ergänzungen, Änderungen und Ausnahmen sind vorhanden: Strenge Standards schränken die Innovationsfähigkeit ein. Stellen Sie Mitgliedern Ihrer Organisation Mechanismen bereit, um Anfragen zur Unterstützung ihrer Geschäftsanforderungen an Besitzer von Prozessen, Verfahren und Ressourcen zu stellen.

OPS02-BP07 Zuständigkeiten zwischen Teams werden vordefiniert oder ausgehandelt:

Es gibt definierte oder ausgehandelte Vereinbarungen zwischen Teams, in denen die Zusammenarbeit und gegenseitige Unterstützung beschrieben wird (z. B. Reaktionszeiten, Service Level Objectives oder Service Level Agreements). Wenn bekannt ist, welche Auswirkungen die Arbeit der Teams auf die Geschäftsergebnisse und die Ergebnisse anderer Teams und Organisationen hat, können Teams ihre Aufgaben priorisieren und entsprechend handeln.

Wenn Verantwortlichkeit und Eigentümerschaft undefiniert oder unbekannt sind, besteht das Risiko, dass sowohl die erforderlichen Aktivitäten nicht rechtzeitig behandelt werden, als auch redundante

und potenziell widersprüchliche Anstrengungen unternommen werden, um diese Anforderungen zu erfüllen.

Vorteile der Einführung dieser bewährten Methode: Durch die Festlegung der Verantwortlichkeiten zwischen Teams, der Ziele und der Methoden für die Kommunikation von Anforderungen werden weniger Anfragen gestellt und die Bereitstellung der erforderlichen Informationen wird gewährleistet. Dadurch können Aufgaben schneller zwischen Teams übergeben und die Geschäftsergebnisse leichter erreicht werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Zuständigkeiten zwischen Teams werden vordefiniert oder ausgehandelt: Durch die Angabe der Methoden, mit denen Teams interagieren, und der Informationen, die für die gegenseitige Unterstützung erforderlich sind, kann die Verzögerung minimiert werden, die durch die iterative Überprüfung und Klärung von Anfragen entsteht. Mit spezifischen Vereinbarungen, in denen Erwartungen definiert sind (z. B. Reaktionszeit oder Ausführungszeit), können Teams effektive Pläne erstellen und Ressourcen entsprechend organisieren.

OPS 3 Wie unterstützt Ihre Unternehmenskultur Ihre Geschäftsergebnisse?

Stellen Sie Ihren Teammitgliedern Unterstützung bereit, damit sie effektiver handeln und Ihr Geschäftsergebnis unterstützen können.

Bewährte Methoden

- OPS03-BP01 Förderung durch die Geschäftsführung
- OPS03-BP02 Teammitglieder sind befugt, Maßnahmen zu ergreifen, wenn Ergebnisse gefährdet sind:
- OPS03-BP03 Eskalation wird empfohlen
- OPS03-BP04 Kommunikation ist zeitnah, klar und umsetzbar
- OPS03-BP05 Experimentieren wird empfohlen
- OPS03-BP06 Teammitglieder werden in die Lage versetzt und ermutigt, ihre Fähigkeiten zu pflegen und zu erweitern:
- OPS03-BP07 Teams mit entsprechenden Ressourcen ausstatten
- OPS03-BP08 Unterschiedliche Meinungen werden innerhalb des Teams und teamübergreifend gefördert und sind erwünscht

OPS03-BP01 Förderung durch die Geschäftsführung

Die Geschäftsführung legt klare Erwartungen für das Unternehmen fest und bewertet den Erfolg. Die Geschäftsführung ist Sponsor, Fürsprecher und treibende Kraft für die Übernahme bewährter Methoden und die Weiterentwicklung des Unternehmens

Vorteile der Einführung dieser bewährten Methode: Eine engagierte Geschäftsführung, klar kommunizierte Erwartungen und gemeinsame Ziele stellen sicher, dass die Teammitglieder wissen, was von ihnen erwartet wird. Mit der Erfolgsevaluierung können die Hindernisse auf dem Weg zum Erfolg identifiziert und durch die Intervention der Geschäftsführung oder ihrer Delegierten behoben werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Förderung durch Geschäftsführung: Die Geschäftsführung legt klare Erwartungen für das Unternehmen fest und bewertet den Erfolg. Die Geschäftsführung ist Sponsor, Fürsprecher und treibende Kraft für die Übernahme bewährter Methoden und die Weiterentwicklung des Unternehmens
  - Festlegen von Erwartungen: Definieren und veröffentlichen Sie Ziele für Ihre Teams einschließlich der Art, wie diese Ziele gemessen werden.
  - Verfolgen der Zielerreichung: Überprüfen Sie regelmäßig die stufenweise Erreichung von Zielen und teilen Sie den entsprechenden Teams die Ergebnisse mit, damit geeignete Maßnahmen ergriffen werden können, wenn angepeilte Ergebnisse gefährdet sind.
  - Bereitstellen der erforderlichen Ressourcen zum Erreichen Ihrer Ziele: Überprüfen Sie regelmäßig, ob die vorhandenen Ressourcen noch ausreichen oder ob aufgrund neuer Informationen, Änderungen an Zielen, Verantwortlichkeiten oder Ihrer Geschäftsumgebung zusätzliche Ressourcen benötigt werden.
  - Unterstützen Ihrer Teams: Bleiben Sie mit Ihren Teams in Verbindung, damit Sie wissen, wie es ihnen ergeht und ob es äußere beeinträchtigende Faktoren gibt. Wenn sich äußere Faktoren negativ auf Ihre Teams auswirken, bewerten Sie die Ziele neu und passen Sie sie entsprechend an. Identifizieren Sie Hindernisse für den Fortschritt Ihrer Teams. Treten Sie für Ihre Teams ein und beseitigen Sie Hindernisse und unnötige Bürden.
  - Treibende Kraft für Übernahme bewährter Methoden: Würdigen Sie bewährte Methoden, die messbare Vorteile bieten, und geben Sie ihren Entwicklern und Anwendern Anerkennung. Ermutigen Sie Ihre Teams zur Annahme dieser Methoden, um die Vorteile noch zu verstärken.

- Treibende Kraft für die Entwicklung Ihrer Teams: Schaffen Sie eine Kultur der kontinuierlichen Verbesserung. Fördern Sie das Wachstum und die Entwicklung sowohl im Persönlichen als auch im Betrieblichen. Setzen Sie langfristige Ziele, die stufenweise Erfolge über einen längeren Zeitraum hinweg erfordern. Passen Sie diese Vision an Ihre Anforderungen, Geschäftsziele und Ihre Geschäftsumgebung an, wenn sie sich ändern.

OPS03-BP02 Teammitglieder sind befugt, Maßnahmen zu ergreifen, wenn Ergebnisse gefährdet sind:

Der/die Verantwortliche des Workload hat klare Anweisungen und Zuständigkeitsbereiche festgelegt, damit alle Teammitglieder direkt reagieren können, wenn die Ziele gefährdet sind. Es werden Eskalationsmechanismen verwendet, damit klare Anweisungen gelten, wenn Ereignisse außerhalb des festgelegten Zuständigkeitsbereichs liegen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie Änderungen frühzeitig testen und validieren, können Sie Probleme mit minimalen Kosten beheben und die Auswirkungen auf Ihre Kunden einschränken. Durch Tests vor der Bereitstellung minimieren Sie die Fehler.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Befugnis der Teammitglieder zu Maßnahmen bei Gefährdung der angepeilten Ergebnisse: Geben Sie Ihren Teammitgliedern die erforderlichen Berechtigungen, Hilfsmittel und Möglichkeiten, damit sie die benötigten Fertigkeiten für eine effektive Reaktion einüben können.
  - Befähigen der Teammitglieder zum Einüben der erforderlichen Fertigkeiten für die Reaktion: Stellen Sie alternative sichere Umgebungen bereit, in denen Prozesse und Verfahren sicher getestet und eingeübt werden können. Führen Sie Ernstfallübungen durch, damit Ihre Teammitglieder Erfahrung beim Reagieren auf reale Vorfälle in simulierten und sicheren Umgebungen sammeln können.
  - Definieren und Bestätigen der Befugnis von Teammitgliedern zum Ergreifen von Maßnahmen: Verschaffen Sie den Teammitgliedern die erforderliche Autorität, um Maßnahmen zu ergreifen, indem Sie ihnen Berechtigungen und Zugriff auf ihre Workloads und Komponenten geben. Sagen Sie ihnen deutlich, dass sie befugt sind, Maßnahmen zu ergreifen, wenn die Ziele gefährdet sind.

OPS03-BP03 Eskalation wird empfohlen

Teammitglieder verfügen über entsprechende Mechanismen und werden ermutigt, Bedenken an Entscheidungsträger und Beteiligte zu eskalieren, wenn ihnen Ziele als gefährdet erscheinen. Die Eskalation sollte früh und oft durchgeführt werden, damit Risiken identifiziert und Vorfälle verhindert werden können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Ermutigen zu einem frühen und häufigen Eskalieren: Bestätigen Sie im Unternehmen, dass die frühe und oftmalige Eskalation die bewährte Methode ist. Bestätigen und akzeptieren Sie im Unternehmen, dass sich Eskalationen zwar als unbegründet herausstellen können, es sich aber trotzdem insgesamt lohnt, wenn ein echter Vorfall dadurch verhindert wird.
  - Bereitstellung eines Mechanismus für die Eskalation: Sorgen Sie für dokumentierte Verfahren, die definieren, wann und wie eine Eskalation erfolgen soll. Dokumentieren Sie eine Abfolge von Personen mit zunehmender Autorität zum Ergreifen oder Bestätigen von Maßnahmen und ihre Kontaktinformationen. Die Eskalation sollte so weit gehen, bis das Teammitglied der Meinung ist, dass das Problem an eine Person übergeben wurde, die damit umgehen kann, oder bis die Person kontaktiert wurde, die für das Risiko und den Betrieb des Workload verantwortlich ist. Letztendlich ist diese Person für alle Entscheidungen zu ihrem Workload verantwortlich. Eskalationen müssen die Art des Risikos, die Bedeutung des Workload, die betroffenen Personen, die Auswirkungen und die Dringlichkeit bzw. den voraussichtlichen Zeitpunkt der Auswirkungen enthalten.
  - Schutz von eskalierenden Mitarbeitern: Stellen Sie eine Richtlinie bereit, die Teammitglieder vor Konsequenzen schützt, wenn sie zu einem nicht reagierenden Entscheidungsträger oder Verantwortlichen eskalieren. Schaffen Sie Mechanismen, durch die überprüft wird, ob dies geschieht, und leiten Sie entsprechende Maßnahmen ein.

OPS03-BP04 Kommunikation ist zeitnah, klar und umsetzbar

Es gibt Mechanismen und sie werden angewandt, um Teammitglieder rechtzeitig über bekannte Risiken und geplante Ereignisse zu informieren. Erforderlicher Kontext, Details und Zeit (wenn möglich) werden bereitgestellt, um festzustellen, ob und welche Maßnahmen erforderlich sind, und um rechtzeitig Maßnahmen ergreifen zu können. Zum Beispiel die Benachrichtigung über Software-Schwachstellen, damit Patches beschleunigt werden können, oder die Benachrichtigung über

geplante Verkaufsaktionen, damit ein Einfrieren von Änderungen implementiert werden kann, um das Risiko einer Service-Unterbrechung zu vermeiden.

Geplante Ereignisse können in einem Änderungskalender oder Wartungsplan aufgezeichnet werden, sodass Teammitglieder feststellen können, welche Aktivitäten ausstehen.

In AWS kann der AWS Systems Manager Change Calendar verwendet werden, um diese Details aufzuzeichnen. Er unterstützt programmgesteuerte Prüfungen des Kalenderstatus, um zu bestimmen, ob der Kalender zu einem bestimmten Zeitpunkt geöffnet oder geschlossen ist. Betriebsaktivitäten können um bestimmte genehmigte Zeitfenster geplant werden, die für potenziell störende Aktivitäten reserviert sind. AWS Systems Manager Maintenance Windows ermöglicht es Ihnen, Aktivitäten für Instances und andere unterstützte Ressourcen zu planen, um die Aktivitäten zu automatisieren und diese Aktivitäten auffindbar zu machen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Rechtzeitige, klare und unterstützende Kommunikation: Es gibt Mechanismen zur Benachrichtigung über Risiken oder geplante Ereignisse auf eine klare und unterstützende Weise mit ausreichend Zeit für geeignete Maßnahmen.
  - Dokumentieren geplanter Aktivitäten in einem Änderungskalender und Bereitstellen von Benachrichtigungen: Stellen Sie eine zugängliche Informationsquelle bereit, der geplante Ereignisse zu entnehmen sind. Stellen Sie Benachrichtigungen zu geplanten Ereignissen vom gleichen System bereit.
  - Verfolgen von Ereignissen und Aktivitäten mit möglichen Auswirkungen auf Ihren Workload: Überwachen Sie Benachrichtigungen zu Schwachstellen und Patch-Informationen, um bestehende Schwachstellen und potenzielle Risiken im Zusammenhang mit den Komponenten Ihrer Workloads zu verstehen. Stellen Sie Benachrichtigungen für die Teammitglieder bereit, damit sie Maßnahmen ergreifen können.

Ressourcen

Zugehörige Dokumente:

- AWS Systems Manager Change Calendar
- AWS Systems Manager Maintenance Windows

OPS03-BP05 Experimentieren wird empfohlen

Durch das Experimentieren werden Lernprozesse beschleunigt und bleiben Teammitglieder interessiert und engagiert. Ein unerwünschtes Ergebnis ist ein erfolgreiches Experiment, das einen Weg identifiziert hat, der nicht zum Erfolg führt. Teammitglieder werden nicht für erfolgreiche Experimente mit unerwünschten Ergebnissen bestraft. Durch Experimente werden Innovationen möglich und Ideen zu Ergebnissen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Ermutigen zum Experimentieren: Ermutigen Sie zum Experimentieren, um Lernprozesse und Innovationen zu fördern.
  - Experimentieren mit vielfältigen Technologien: Ermutigen Sie zum Experimentieren mit Technologien, die jetzt oder in der Zukunft zum Erreichen Ihrer Geschäftsergebnisse beitragen könnten. Das gewonnene Wissen kann die Grundlage für zukünftige Innovationen sein.
  - Experimentieren mit einem klaren Ziel: Ermutigen Sie Ihre Teammitglieder zum Experimentieren mit bestimmten Zielen oder Technologien, die in naher Zukunft nützlich sein könnten. Das gewonnene Wissen kann die Grundlage für Ihre Innovationen sein.
  - Bereitstellen strukturierter Zeit zum Experimentieren: Legen Sie bestimmte Zeiten fest, in denen Teammitglieder von ihren normalen Aufgaben befreit sind, damit sie sich auf das Experimentieren konzentrieren können.
  - Bereitstellen der erforderlichen Ressourcen für Experimente: Finanzieren Sie die erforderlichen Ressourcen zur Durchführung von Experimenten (z. B. Software oder Cloud-Ressourcen).
  - Würdigen des Erfolgs: Würdigen Sie den Wert der Experimente. Berücksichtigen Sie, dass Experimente mit unerwünschten Ergebnissen auch ein Erfolg sind, weil sie zeigen, dass ein bestimmter Weg nicht zum Erfolg führt. Teammitglieder werden nicht für unerwünschte Ergebnisse von Experimenten bestraft.

OPS03-BP06 Teammitglieder werden in die Lage versetzt und ermutigt, ihre Fähigkeiten zu pflegen und zu erweitern:

Teams müssen ihre Fertigkeiten ausbauen, um neue Technologien nutzen und mit veränderten Anforderungen und Aufgaben Ihrer Workloads umgehen zu können. Neue Fertigkeiten im Umgang mit neuen Technologien erhöhen oftmals die Zufriedenheit der Teammitglieder und ermöglichen neue Innovationen. Unterstützen Sie Ihre Teammitglieder beim Erlangen und Bewahren von

Branchenzertifizierungen, mit denen ihre zunehmenden Fertigkeiten bestätigt und anerkannt werden. Führen Sie funktionsübergreifende Schulungen durch, um den Wissenstransfer zu fördern und das Risiko signifikanter Auswirkungen zu reduzieren, wenn Sie qualifizierte und erfahrene Teammitglieder mit kritischem Wissen verlieren. Schaffen Sie spezielle strukturierte Lernzeiten.

AWS stellt Ressourcen bereit, darunter das Erste Schritte – AWS Resource Center, AWS-Blogs, AWS Online Tech Talks, AWS-Veranstaltungen und -Webinaresowie die AWS Well-Architected Labs, die Anleitungen, Beispiele und detaillierte Walkthroughs zur Schulung Ihrer Teams bieten.

AWS stellt in der Amazon Builders' Library auch bewährte Methoden und Muster vor, die wir durch den Betrieb von AWS gelernt haben Die Amazon Builders' Library auch bewährte Methoden und Muster vor, die wir durch den Betrieb von AWS gelernt haben, sowie eine Vielzahl weiterer nützlicher Lernmaterialien im AWS-Blog und im offiziellen AWS-Podcast.

Sie sollten die von AWS bereitgestellten Schulungsressourcen nutzen, z. B. die Well-Architected Labs, den AWS Support (AWS Knowledge Center, AWS Diskussionsforenund AWS Support Center) bereitgestellten Ressourcen und AWS-Dokumentation nutzen, um Ihre Teams zu schulen. Wenn Sie eine Frage zu AWS haben, können Sie sich über das AWS Support Center an den AWS Support wenden.

AWS Training und Zertifizierung bietet einige kostenlose Schulungen durch digitale Kurse im Selbststudium zu den Grundlagen von AWS. Sie können sich auch für eine Schulung registrieren, die von Dozenten geleitet wird, um die AWS-Fähigkeiten und -Fertigkeiten Ihres Teams auszubauen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Teammitglieder werden in die Lage versetzt und ermutigt, ihre Fähigkeiten zu pflegen und zu erweitern: Zur Einführung neuer Technologien, um Innovationen und Änderungen bei Bedarf und Zuständigkeiten bei der Unterstützung Ihrer Workloads zu unterstützen, ist fortlaufende Bildung notwendig.
  - Bereitstellen von Ressourcen für die Weiterbildung: Stellen Sie eine spezielle strukturierte Lernzeit, Schulungsmaterialien und Laborressourcen bereit. Unterstützen Sie die Teilnahme an Konferenzen und bei professionellen Organisationen, die Möglichkeiten zum Lernen von Lehrenden und anderen Fachleuten bieten. Sorgen Sie dafür, dass erfahrene Teammitglieder neueren Teammitgliedern als Mentoren dienen können, oder dass sie sich Arbeitsweisen, Methoden und Fertigkeiten von ihnen abschauen können. Ermutigen Sie dazu, auch etwas über Inhalte zu lernen, die nicht direkt mit der Arbeit zusammenhängen, um den Horizont zu erweitern.

- Teamschulung und teamübergreifende Zusammenarbeit: Planen Sie die kontinuierlichen Weiterbildungsanforderungen Ihrer Teammitglieder mit ein. Schaffen Sie Gelegenheiten für die Teammitglieder, (vorübergehend oder dauerhaft) in anderen Teams zu arbeiten, damit sie ihre Fertigkeiten und bewährten Methoden austauschen können, wovon letztendlich das gesamte Unternehmen profitiert.
- Unterstützen beim Erlangen und Bewahren von Branchenzertifizierungen: Unterstützen Sie Ihre Teammitglieder beim Erlangen und Bewahren von Branchenzertifizierungen, durch die das Gelernte bestätigt wird und die Erfolge anerkannt werden.

Ressourcen

Zugehörige Dokumente:

- Erste Schritte – AWS Resource Center
- AWS-Blogs
- AWS Cloud-Compliance
- AWS Diskussionsforen
- AWS-Dokumentation nutzen,
- AWS Online Tech Talks
- AWS-Veranstaltungen und -Webinare
- AWS Knowledge Center
- AWS Support
- AWS Training und Zertifizierung
- AWS Well-Architected Labs,
- Die Amazon Builders' Library
- im offiziellen AWS-Podcast.

OPS03-BP07 Teams mit entsprechenden Ressourcen ausstatten

Legen Sie eine angemessene Teamgröße fest und stellen Sie die erforderlichen Hilfsmittel und Ressourcen für die Workloads bereit. Die Überlastung von Teammitgliedern erhöht das Risiko von Vorfällen durch menschliches Versagen. Investitionen in Tools und Ressourcen (z. B. Automatisierung für häufige Aufgaben) können die Effektivität Ihres Teams deutlich steigern, wodurch es sich ggf. um zusätzliche Aufgaben kümmern kann.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Angemessene Teamplanung: Stellen Sie sicher, dass Sie die Bedeutung und die maßgeblichen Faktoren des Erfolgs oder Misserfolgs Ihrer Teams kennen. Unterstützen Sie Teams mit erforderlichen Ressourcen.
  - Verstehen der Teamleistung: Messen Sie die Erreichung von Betriebsergebnissen und die Entwicklung von Assets durch Ihre Teams. Verfolgen Sie Änderungen bei dem Output und der Fehlerrate im Zeitverlauf. Sprechen Sie mit Teams, um sich über ihre arbeitsbezogenen Herausforderungen zu informieren (z. B. zunehmende Aufgaben, technologische Veränderungen, Verlust von Mitarbeitern oder steigende Kundenzahl).
  - Verstehen der Auswirkungen auf die Teamleistung: Bleiben Sie mit Ihren Teams in Verbindung, damit Sie wissen, wie es ihnen ergeht und ob es äußere beeinträchtigende Faktoren gibt. Wenn sich äußere Faktoren negativ auf Ihre Teams auswirken, bewerten Sie die Ziele neu und passen Sie sie entsprechend an. Identifizieren Sie Hindernisse für den Fortschritt Ihrer Teams. Treten Sie für Ihre Teams ein und beseitigen Sie Hindernisse und unnötige Bürden.
  - Bereitstellen der erforderlichen Ressourcen für den Erfolg von Teams: Überprüfen Sie regelmäßig, ob die vorhandenen Ressourcen noch ausreichen oder zusätzliche Ressourcen benötigt werden, und unterstützen Sie die Teams durch entsprechende Korrekturen.

OPS03-BP08 Unterschiedliche Meinungen werden innerhalb des Teams und teamübergreifend gefördert und sind erwünscht

Nutzen Sie die funktionsübergreifende Diversität, um verschiedene einzigartige Perspektiven zu erhalten. Nutzen Sie diese Perspektive, um Innovation zu fördern, Ihre Annahmen in Frage zu stellen und das Risiko einer Verzerrung durch automatische Bestätigung zu reduzieren. Erweitern Sie Inklusion, Diversität und Offenheit innerhalb Ihrer Teams, um nützliche Perspektiven zu gewinnen.

Die Unternehmenskultur wirkt sich direkt auf die Zufriedenheit und Bindung der Teammitglieder aus. Ermöglichen Sie die Interaktion und aktivieren Sie die Fähigkeiten Ihrer Teammitglieder für den Erfolg Ihres Unternehmens.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Berücksichtigen unterschiedlicher Meinungen und Perspektiven: Ermutigen Sie alle anderen, einen Beitrag zu leisten. Geben Sie unterrepräsentierten Gruppen eine Stimme. Rotieren Sie die Rollen und Zuständigkeiten in Meetings.
  - Erweitern von Rollen und Zuständigkeiten: Bieten Sie Teammitgliedern die Möglichkeit, Rollen zu übernehmen, die ihnen fremd sind. Sie sammeln Erfahrung und erhalten neue Perspektiven durch die Rolle und den resultierenden Austausch mit neuen Teammitgliedern, zu denen sie möglicherweise andernfalls keinen Kontakt hätten. Sie werden die neue Rolle und die Teammitglieder mit ihren Erfahrungen und Perspektiven bereichern. Aus der erweiterten Perspektive können sich neue Geschäftschancen oder neue Verbesserungsmöglichkeiten ergeben. Lassen Sie Mitglieder innerhalb eines Teams abwechselnd allgemeine Aufgaben übernehmen, die normalerweise andere ausführen, um ihre Anforderungen und Auswirkungen zu verstehen.
  - Bereitstellen einer sicheren und freundlichen Umgebung: Stellen Sie Richtlinien und Kontrollen zum Schutz der geistigen und physischen Sicherheit der Teammitglieder in Ihrem Unternehmen bereit. Die Teammitglieder müssen ohne Angst vor Vergeltung zusammenarbeiten können. Wenn sich Teammitglieder sicher und willkommen fühlen, ist die Wahrscheinlichkeit höher, dass sie engagiert und produktiv bleiben. Je vielfältiger Ihr Unternehmen ist, desto besser können Sie andere verstehen, einschließlich Ihrer Kunden. Wenn Ihre Teammitglieder zufrieden sind, ihre Meinung sagen können und sich ernst genommen fühlen, steigt die Wahrscheinlichkeit, dass sie wertvolle Erkenntnisse mitteilen (z. B. Marketingmöglichkeiten, erforderliche Zugänglichkeit, unerschlossene Marktsegmente, unbehandelte Risiken in Ihrer Umgebung).
  - Ermöglichen der vollständigen Teilnahme von Teammitgliedern: Stellen Sie die Ressourcen bereit, die Ihre Mitarbeiter zur vollständigen Teilnahme an allen arbeitsbezogenen Tätigkeiten benötigen. Teammitglieder haben Fertigkeiten entwickelt, mit denen sie ihre täglichen Herausforderungen meistern. Diese einzigartigen Fertigkeiten können Ihrem Unternehmen einen erheblichen Vorteil bieten. Wenn Sie die Teammitglieder mit den notwendigen Ressourcen ausstatten, werden die Vorteile ihres Beitrags verstärkt.

Vorbereitung

Fragen

- OPS 4 Wie können Sie Ihren Workload so konzipieren, dass sein jeweiliger Zustand klar ersichtlich ist?

- OPS 5 Wie können Sie Fehler reduzieren, die Fehlerbehebung erleichtern und den Ablauf bis zur Produktion verbessern?
- OPS 6 Wie können Sie Bereitstellungsrisiken eindämmen?
- OPS 7 Wie bringen Sie in Erfahrung, ob Sie für die Unterstützung eines Workloads bereit sind?

OPS 4 Wie können Sie Ihren Workload so konzipieren, dass sein jeweiliger Zustand klar ersichtlich ist?

Gestalten Sie Ihren Workload so, dass er die Informationen liefert, die Sie benötigen, um seinen internen Zustand über alle Komponenten (z. B. Metriken, Protokolle und Tracing) hinweg zu verstehen. Auf diese Weise können Sie im Bedarfsfall effektiv reagieren.

Bewährte Methoden

- OPS04-BP01 Implementieren einer Anwendungstelemetrie
- OPS04-BP02 Implementieren und Konfigurieren der Workload-Telemetrie
- OPS04-BP03 Telemetrie von Benutzeraktivitäten implementieren
- OPS04-BP04 Implementieren einer Abhängigkeitstelemetrie
- OPS04-BP05 Implementieren einer Nachvollziehbarkeit von Transaktionen

OPS04-BP01 Implementieren einer Anwendungstelemetrie

Anwendungs-Telemetrie ist die Grundlage für Beobachtbarkeit Ihres Workloads. Ihre Anwendung sollte Telemetriedaten ausgeben, die Aufschluss über den Zustand der Anwendung und das Erreichen von Geschäftsergebnissen geben. Von der Fehlerbehebung bis hin zur Messung der Auswirkungen einer neuen Funktion liefert die Anwendungstelemetrie Informationen darüber, wie Sie Ihren Workload aufbauen, betreiben und weiterentwickeln.

Anwendungstelemetrie besteht aus Metriken und Protokollen. Bei Metriken handelt es sich um Diagnosedaten, wie Ihr Puls oder Ihre Körpertemperatur. Metriken werden gemeinsam verwendet, um den Zustand Ihrer Anwendung zu beschreiben. Das Sammeln von Metriken im Zeitverlauf kann dazu verwendet werden, Grundlinien zu entwickeln und Anomalien zu erkennen. Protokolle sind Meldungen, die die Anwendung ihren internen Zustand oder auftretende Ereignisse betreffend sendet. Fehlercodes, Transaktionskennungen und Benutzeraktionen sind Beispiele für protokollierte Ereignisse.

Gewünschtes Ergebnis:

- Ihre Anwendung gibt Metriken und Protokolle an, die Aufschluss über ihren Zustand und das Erreichen von Geschäftsergebnissen geben.
- Metriken und Protokolle werden zentral für alle Anwendungen im Workload gespeichert.

Gängige Antimuster:

- Ihre Anwendung sendet keine Telemetriedaten. Sie müssen sich darauf verlassen, dass Ihre Kunden Ihnen mitteilen, wenn etwas nicht stimmt.
- Ein Kunde hat gemeldet, dass Ihre Anwendung nicht reagiert. Sie verfügen über keine Telemetrie und können nicht bestätigen, dass das Problem existiert, und es auch nicht einschätzen, ohne die Anwendung selbst zu verwenden, um die aktuelle Benutzererfahrung zu verstehen.

Vorteile der Einführung dieser bewährten Methode:

- Sie können den Zustand Ihrer Anwendung, die Benutzererfahrung und das Erreichen von Geschäftsergebnissen nachvollziehen.
- Auf Änderungen am Zustand Ihrer Anwendung können Sie schnell reagieren.
- Sie können Zustandstrends für Anwendungen entwickeln.
- Sie können fundierte Entscheidungen hinsichtlich der Verbesserung Ihrer Anwendung treffen.
- Anwendungsprobleme lassen sich schneller erkennen und beheben.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Die Implementierung von Anwendungstelemetrie besteht aus drei Schritten: Identifizierung eines Speicherorts für Telemetrie, Identifizierung von Telemetrie, die den Zustand der Anwendung beschreibt, und Instrumentierung der Anwendung, um Telemetrie auszugeben.

Beispiel: Ein E-Commerce-Unternehmen hat eine auf Microservices basierende Architektur. Im Rahmen des Architekturentwurfs wurde eine Anwendungstelemetrie identifiziert, mit deren Hilfe es den Zustand der einzelnen Microservices nachvollziehen kann. Der Warenkorb-Service hat beispielsweise Telemetriedaten zu Ereignissen wie Hinzufügen zum Warenkorb, Verlassen des Warenkorbs und Dauer des Hinzufügens eines Artikels zum Warenkorb ausgegeben. Alle Microservices würden Fehler, Warnungen und Transaktionsinformationen protokollieren. Telemetrie würde zu Speicher- und Analysezwecken an Amazon CloudWatch gesendet.

Implementierungsschritte

Der erste Schritt besteht darin, einen zentralen Speicherort für die Telemetriedaten der Anwendungen in Ihrem Workload zu ermitteln. Wenn Sie keine bestehende Plattform haben, bietet Amazon CloudWatch die Erfassung von Telemetriedaten, Dashboards, Analysen und Fähigkeiten zur Ereigniserzeugung.

Stellen Sie sich folgende Fragen, um herauszufinden, welche Telemetrie Sie benötigen:

- Ist meine Anwendung in einem guten Zustand?
- Erreicht meine Anwendung die gewünschten Geschäftsergebnisse?

  Ihre Anwendung sollte Protokolle und Metriken ausgeben, die gemeinsam eine Antwort auf diese Fragen bieten. Wenn Sie diese Fragen mit der vorhandenen Anwendungstelemetrie nicht beantworten können, arbeiten Sie mit den Ansprechpersonen aus den Bereichen Business und Technik zusammen, um eine Liste von Telemetriedaten zu erstellen, die dies ermöglichen. Sie können Ihr AWS-Konto-Team um fachkundige technische Beratung bitten, wenn Sie neue Anwendungstelemetrie identifizieren und entwickeln.

  Sobald die zusätzliche Anwendungstelemetrie identifiziert wurde, arbeiten Sie mit Ihren Ansprechpartnern aus dem technischen Bereich zusammen, um Ihre Anwendung zu instrumentieren. AWS Distro for OpenTelemetry bietet APIs, Bibliotheken und Agenten, die Anwendungstelemetrie bieten. Dieses Beispiel zeigt, wie man eine JavaScript-Anwendung mit benutzerdefinierten Metriken instrumentiert.

  Kunden, die die Beobachtbarkeits-Services verstehen möchten, die AWS anbietet, können den Workshop zur Beobachtbarkeit eigenständig durchgehen oder Unterstützung von ihrem AWS-Konto-Team anfordern. Dieser Workshop führt Sie durch die Beobachtbarkeitslösungen von AWS und bietet praktische Beispiele für deren Einsatz.

  Für umfassendere Einblicke in die Anwendungstelemetrie lesen Sie den Artikel „Instrumentieren verteilter Systeme für Einblicke in die Betriebsabläufe“ in der Amazon Builder’s Library. Darin wird erklärt, wie Amazon Anwendungen instrumentiert. Er kann als Leitfaden für die Entwicklung eigener Instrumentierungsrichtlinien dienen.

Grad des Aufwands für den Implementierungsplan: Mittel

Ressourcen

Relevante bewährte Methoden:

the section called “OPS04-BP02 Implementieren und Konfigurieren der Workload-Telemetrie” – Anwendungstelemetrie ist ein Bestandteil der Workload-Telemetrie. Sie müssen den Zustand der einzelnen Anwendungen, aus denen der Workload besteht, kennen, um den Zustand des gesamten Workloads zu verstehen.

the section called “OPS04-BP03 Telemetrie von Benutzeraktivitäten implementieren” – Die Telemetrie der Benutzeraktivität ist häufig eine Teilmenge der Anwendungstelemetrie. Benutzeraktivitäten, wie z. B. das Hinzufügen zum Warenkorb, Clickstreams oder abgeschlossene Transaktionen, geben Aufschluss über das Benutzererlebnis.

the section called “OPS04-BP04 Implementieren einer Abhängigkeitstelemetrie” – Abhängigkeitsprüfungen beziehen sich auf die Anwendungstelemetrie und können in Ihre Anwendung instrumentiert werden. Wenn Ihre Anwendung von externen Abhängigkeiten wie DNS oder einer Datenbank abhängig ist, kann Ihre Anwendung Metriken und Protokolle über Erreichbarkeit, Timeouts und andere Ereignisse ausgeben.

the section called “OPS04-BP05 Implementieren einer Nachvollziehbarkeit von Transaktionen” – Für die Verfolgung von Transaktionen über einen Workload hinweg muss jede Anwendung Informationen darüber ausgeben, wie sie gemeinsame Ereignisse verarbeitet. Die Art und Weise, wie die einzelnen Anwendungen mit diesen Ereignissen umgehen, wird über ihre Anwendungstelemetrie übermittelt.

the section called “OPS08-BP02 Definieren von Workload-Metriken” – Workload-Metriken sind die wesentlichen Zustandsindikatoren für Ihren Workload. Wesentliche Anwendungsmetriken sind Teil der Workload-Metriken.

Zugehörige Dokumente:

- AWS Builders' Library – Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- AWS Distro for OpenTelemetry
- AWS Well-Architected Whitepaper zur betrieblichen Exzellenz – Entwerfen von Telemetrie
- Erstellen von Metriken aus Protokollereignissen mit Filtern
- Implementieren von Protokollierung und Überwachung mit Amazon CloudWatch
- Überwachen des Zustands und der Leistung der Anwendung mit AWS Distro for OpenTelemetry

- Neu: Wie Sie eine bessere Überwachung Ihrer benutzerdefinierten Anwendungsmetriken mit dem Amazon CloudWatch-Agent erreichen
- Beobachtbarkeit bei AWS
- Szenario: Metriken in CloudWatch veröffentlichen
- Mit dem Entwickeln beginnen – Effektives Überwachen Ihrer Anwendungen
- Verwenden von CloudWatch mit einem AWS-SDK

Relevante Videos:

- AWS re:Invent 2021 – Observability the open-source way (AWS re:Invent 2021 – Beobachtbarkeit nach dem Open-Source-Prinzip)
- Collect Metrics and Logs from Amazon EC2 instances with the CloudWatch Agent (Erfassen von Metriken und Protokollen aus EC-Instances mit dem CW-Agent)
- How to Easily Setup Application Monitoring for Your AWS Workloads (So richten Sie die Anwendungsüberwachung mühelos für Ihre AWS-Workloads ein) – AWS Online Tech Talks
- Mastering Observability of Your Serverless Applications (Beherrschung der Beobachtbarkeit Ihrer serverlosen Anwendungen) – AWS Online Tech Talks
- Open Source Observability with AWS (Open-Source-Beobachtbarkeit mit AWS) – AWS Virtual Workshop

Zugehörige Beispiele:

- AWS – Protokollierung und Überwachung – Beispielressourcen
- AWS-Lösung: Amazon CloudWatch-Überwachungs-Framework
- AWS-Lösung: Centralized Logging
- Workshop zur Beobachtbarkeit

OPS04-BP02 Implementieren und Konfigurieren der Workload-Telemetrie

Entwickeln und konfigurieren Sie Ihren Workload so, dass Sie Informationen über den jeweiligen internen Zustand und den aktuellen Status erhalten (zum Beispiel über die Menge an API-Aufrufen, HTTP-Statuscodes und Skalierungsereignisse). Ermitteln Sie mithilfe dieser Informationen, wann ein Eingreifen erforderlich ist.

Verwenden Sie einen Service wie Amazon CloudWatch, um Protokolle und Metriken aus Workload-Komponenten zu aggregieren (z. B. API-Protokolle aus AWS CloudTrail, AWS Lambda-Metriken, Amazon VPC-Flow-Protokolleund andere Services).

Gängige Antimuster:

- Ihre Kunden beschweren sich über eine schlechte Leistung. Ihre Anwendung wurde in der letzten Zeit nicht verändert, daher vermuten Sie ein Problem mit einer Workload-Komponente. Sie verfügen über keine Telemetrie, um zu bestimmen, welche Komponenten zur schlechten Leistung beitragen.
- Ihre Anwendung ist nicht erreichbar. Ihnen fehlt die Telemetrie, um festzustellen, ob es sich um ein Netzwerkproblem handelt.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie verstehen, was in Ihrem Workload geschieht, können Sie bei Bedarf reagieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren einer Protokoll- und Metriktelemetrie: Nutzen Sie Ihren Workload, um Informationen über den jeweiligen internen Zustand, den Status und die Erreichung von Geschäftsergebnissen zu erhalten. Ermitteln Sie mithilfe dieser Informationen, wann ein Eingreifen erforderlich ist.
  - Bessere Überwachung Ihrer VMs mit Amazon CloudWatch – AWS Online Tech Talks
  - FunktionsweiseAmazon CloudWatch von
  - Was ist Amazon CloudWatch?
  - Verwenden von Amazon CloudWatch-Metriken
  - Was ist Amazon CloudWatch Logs?
    - Implementieren und Konfigurieren der Workload-Telemetrie: Entwickeln und konfigurieren Sie Ihren Workload so, dass Sie Informationen über den jeweiligen internen Zustand und den aktuellen Status erhalten (zum Beispiel über die Menge an API-Aufrufen, HTTP-Statuscodes und Skalierungsereignisse).
      - Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
      - AWS CloudTrail
      - Was ist AWS CloudTrail?

- VPC Flow Logs

Ressourcen

Zugehörige Dokumente:

- AWS CloudTrail
- Amazon CloudWatch-Dokumentation
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- FunktionsweiseAmazon CloudWatch von
- Verwenden von Amazon CloudWatch-Metriken
- VPC Flow Logs
- Was ist AWS CloudTrail?
- Was ist Amazon CloudWatch Logs?
- Was ist Amazon CloudWatch?

Relevante Videos:

- Verwaltung der Anwendungsleistung in AWS
- Bessere Überwachung Ihrer VMs mit Amazon CloudWatch
- Bessere Überwachung Ihrer VMs mit Amazon CloudWatch – AWS Online Tech Talks

OPS04-BP03 Telemetrie von Benutzeraktivitäten implementieren

Nutzen Sie Ihren Anwendungscode so, dass Sie Informationen zur Benutzeraktivität erhalten, zum Beispiel über Click-Streams oder gestartete, abgebrochene und abgeschlossene Transaktionen. Verwenden Sie diese Informationen, um zu verstehen, wie die Anwendung verwendet wird oder welche Nutzungsmuster sie aufweist, und um festzustellen, wann ein Eingreifen erforderlich ist.

Gängige Antimuster:

- Ihre Entwickler haben eine neue Funktion ohne Benutzertelemetrie bereitgestellt und die Auslastung ist gestiegen. Sie können nicht feststellen, ob die erhöhte Auslastung durch die neue Funktion oder durch ein Problem mit dem neuen Code bedingt ist.

- Ihre Entwickler haben eine neue Funktion ohne Benutzertelemetrie bereitgestellt. Sie können nicht beurteilen, ob Ihre Kunden sie verwenden, ohne sie direkt danach zu fragen.

Vorteile der Einführung dieser bewährten Methode: Erfahren Sie, wie Ihre Kunden Ihre Anwendung verwenden, um Nutzungsmuster und unerwartete Verhaltensweisen zu identifizieren und die Möglichkeit zu erhalten, bei Bedarf zu reagieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Telemetrie von Benutzeraktivitäten implementieren: Entwickeln Sie Ihren Anwendungscode so, dass Sie Informationen zur Benutzeraktivität erhalten (zum Beispiel über Click-Streams oder gestartete, abgebrochene und abgeschlossene Transaktionen). Verwenden Sie diese Informationen, um zu verstehen, wie die Anwendung verwendet wird oder welche Nutzungsmuster sie aufweist, und um festzustellen, wann ein Eingreifen erforderlich ist.

OPS04-BP04 Implementieren einer Abhängigkeitstelemetrie

Entwerfen und konfigurieren Sie Ihren Workload so, dass er Informationen über den Status (z. B. Erreichbarkeit oder Reaktionszeit) der Ressourcen ausgibt, von denen er abhängt. Beispiele für externe Abhängigkeiten können externe Datenbanken, DNS und Netzwerkkonnektivität sein. Ermitteln Sie mithilfe dieser Informationen, wann ein Eingreifen erforderlich ist.

Gängige Antimuster:

- Sie können nicht feststellen, ob der Grund für die Unerreichbarkeit Ihrer Anwendung ein DNS-Problem ist, ohne manuell zu überprüfen, ob der Service Ihres DNS-Anbieters funktioniert.
- Ihre Warenkorb-Anwendung kann keine Transaktionen abschließen. Sie können nicht feststellen, ob es an einem Problem bei Ihrem Kreditkarten-Verarbeitungsanbieter liegt, ohne bei ihm nachzufragen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie den Zustand Ihrer Abhängigkeiten verstehen, können Sie bei Bedarf reagieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Implementieren einer Abhängigkeitstelemetrie: Entwickeln und konfigurieren Sie Ihren Workload so, dass Sie Informationen zum Zustand und Status der Systeme erhalten, auf die er angewiesen ist. Einige Beispiele sind: externe Datenbanken, DNS, Netzwerkkonnektivität und externe Kreditkarten-Verarbeitungsservices.
  - Amazon CloudWatch Agent mit AWS Systems Manager-Integration – einheitliche Metrik- und Protokollerfassung für Linux und Windows
  - Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch Agent mit AWS Systems Manager-Integration – einheitliche Metrik- und Protokollerfassung für Linux und Windows
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent

Zugehörige Beispiele:

- Well-Architected Labs – Überwachung von Abhängigkeiten

OPS04-BP05 Implementieren einer Nachvollziehbarkeit von Transaktionen

Implementieren Sie Ihren Anwendungscode und konfigurieren Sie Ihre Workload-Komponenten so, dass Sie Informationen über den Transaktionsfluss im gesamten Workload erhalten. Verwenden Sie diese Informationen, um zu bestimmen, wann eine Reaktion erforderlich ist, und um Sie bei der Identifizierung der Faktoren zu unterstützen, die zu einem Problem beitragen.

In AWS können Sie verteilte Ablaufverfolgungsservices wie AWS X-Rayverwenden, um Ablaufverfolgungen zu sammeln und aufzuzeichnen, während Transaktionen durch Ihren Workload geleitet werden, Karten generieren, um zu sehen, wie Transaktionen über Ihren Workload und Ihre Services fließen, Einblicke in die Beziehungen zwischen Komponenten gewinnen und Probleme in Echtzeit identifizieren und analysieren.

Gängige Antimuster:

- Sie haben eine serverlose Microservices-Architektur implementiert, die mehrere Konten umfasst. Ihre Kunden melden vorübergehende Leistungsprobleme. Sie können nicht feststellen, welche Funktion oder Komponente ursächlich ist, da Sie nicht nachvollziehen und bestimmen können, in welchem Anwendungsbereich das Leistungsproblem entsteht und wodurch es verursacht wird.
- Sie versuchen, festzustellen, wo sich die Flaschenhälse bei der Leistung in Ihrem Workload befinden, damit sie bei der Entwicklung behoben werden können. Sie können die Beziehung zwischen Ihren Anwendungskomponenten und den Services, mit denen sie interagieren, nicht sehen, um festzustellen, wo sich die Engpässe befinden, da Sie nicht über die Nachvollziehbarkeit verfügen, die es Ihnen ermöglichen würde, die spezifischen Services und Pfade aufzuschlüsseln, die die Anwendungsleistung beeinträchtigen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie den Transaktionsfluss in Ihrem gesamten Workload verstehen, können Sie das erwartete Verhalten Ihrer Workload-Transaktionen und Abweichungen vom erwarteten Verhalten in Ihrem gesamten Workload verstehen, sodass Sie bei Bedarf reagieren können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Implementieren einer Nachvollziehbarkeit von Transaktionen: Entwickeln Sie Ihre Anwendung und Ihren Workload so, dass Informationen über den Transaktionsfluss aller Systemkomponenten übermittelt werden, z. B. Transaktionsstufe, aktive Komponente und Zeit bis zum Abschluss der Aktivität. Mithilfe dieser Informationen können Sie feststellen, was gerade bearbeitet wird, was bereits abgeschlossen wurde und welche Ergebnisse die abgeschlossenen Aktivitäten haben. Dadurch können Sie feststellen, wann ein Eingreifen erforderlich ist. Beispielsweise können ungewöhnlich lange Transaktionsreaktionszeiten innerhalb einer Komponente auf Probleme mit dieser Komponente hinweisen.
  - AWS X-Ray
  - Was ist AWS X-Ray?

Ressourcen

Zugehörige Dokumente:

- AWS X-Ray
- Was ist AWS X-Ray?

OPS 5 Wie können Sie Fehler reduzieren, die Fehlerbehebung erleichtern und den Ablauf bis zur Produktion verbessern?

Verwenden Sie Strategien, die die Übertragung von Änderungen auf die Produktionsumgebung verbessern und Refactoring, schnelles Feedback zur Qualität sowie eine schnelle Fehlerbehebung ermöglichen. Dadurch fließen nützliche Änderungen schneller in die Produktion ein und es treten bei der Bereitstellung weniger Probleme auf. Zudem können Probleme, die durch Bereitstellungsaktivitäten verursacht werden, schnell aufgespürt und gelöst werden.

Bewährte Methoden

- OPS05-BP01 Verwendung einer Versionskontrolle
- OPS05-BP02 Testen und Validieren von Änderungen
- OPS05-BP03 Einsatz von Systemen zur Konfigurationsverwaltung
- OPS05-BP04 Einsatz von Systemen zur Build- und Bereitstellungsverwaltung.
- OPS05-BP05 Durchführen der Patch-Verwaltung
- OPS05-BP06 Gemeinsame Design-Standards
- OPS05-BP07 Implementieren von Verfahren zur Verbesserung der Codequalität
- OPS05-BP08 Verwenden mehrerer Umgebungen
- Häufige, kleine, umkehrbare Änderungen vornehmen:
- OPS05-BP10 Vollständige Automatisierung von Integration und Bereitstellung

OPS05-BP01 Verwendung einer Versionskontrolle

Ermöglichen Sie die Verfolgung von Änderungen und Releases mithilfe einer Versionskontrolle.

Viele AWS-Services bieten Versionskontrollfunktionen. Verwenden Sie ein Revisions- oder Quellcodeverwaltungssystem wie AWS CodeCommit, um Code und andere Artefakte zu verwalten, z. B. versionsgesteuerte AWS CloudFormation -Vorlagen Ihrer Infrastruktur.

Gängige Antimuster:

- Sie haben Ihren Code auf Ihrer Workstation entwickelt und gespeichert. Es ist ein Speicherfehler bei der Workstation aufgetreten, der nicht rückgängig gemacht werden kann, und Sie haben den Code verloren.

- Nachdem Sie den vorhandenen Code mit Ihren Änderungen überschrieben haben, starten Sie Ihre Anwendung neu, doch sie funktioniert nicht mehr. Sie können die Änderung nicht rückgängig machen.
- Sie arbeiten an einer Berichtsdatei, deshalb ist sie für alle anderen schreibgeschützt, doch ein anderer Benutzer möchte sie bearbeiten. Der Benutzer kontaktiert Sie und bittet darum, die Arbeit daran zu beenden, damit er seine Aufgabe erledigen kann.
- Ihr Forschungsteam arbeitet an einer detaillierten Analyse, die Ihre zukünftige Arbeit prägen wird. Jemand hat versehentlich seine Einkaufsliste über den endgültigen Bericht gespeichert. Sie können die Änderung nicht rückgängig machen und müssen den Bericht neu erstellen.

Vorteile der Einführung dieser bewährten Methode: Durch die Verwendung von Versionskontrollfunktionen können Sie problemlos auf einen bekanntermaßen funktionierenden Status bzw. frühere Versionen zurücksetzen und so das Risiko von verlorenen Assets begrenzen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Versionskontrolle verwenden: Bewahren Sie Ressourcen in Repositorys mit Versionskontrolle auf. Dies ermöglicht die Nachvollziehung von Änderungen, die Bereitstellung neuer Versionen, die Erkennung von Änderungen an bestehenden Versionen und die Rückkehr zu vorherigen Versionen (zum Beispiel bei einem Fehler die Zurücksetzung auf einen bekanntermaßen funktionierenden Zustand). Integrieren Sie die Versionskontrollfunktionen Ihrer Konfigurationsverwaltungssysteme in Ihre Verfahren.
  - Einführung in AWS CodeCommit
  - Was ist AWS CodeCommit?

Ressourcen

Zugehörige Dokumente:

- Was ist AWS CodeCommit?

Relevante Videos:

- Einführung in AWS CodeCommit

OPS05-BP02 Testen und Validieren von Änderungen

Testen und validieren Sie Änderungen, um Fehler zu reduzieren und zu erkennen. Automatisieren Sie Tests, um Fehler aufgrund von manuellen Prozessen zu reduzieren und den Testaufwand zu verringern.

Viele AWS-Services bieten Versionskontrollfunktionen. Verwenden Sie ein Revisions- oder Quellcodeverwaltungssystem wie AWS CodeCommit um Code und andere Artefakte zu verwalten, z. B. versionsgesteuerte AWS CloudFormation -Vorlagen Ihrer Infrastruktur.

Gängige Antimuster:

- Sie stellen Ihren neuen Code für die Produktion bereit und Kunden rufen an, weil Ihre Anwendung nicht mehr funktioniert.
- Sie wenden neue Sicherheitsgruppen an, um Ihre Umgebungssicherheit zu verbessern. Es funktioniert, jedoch mit unbeabsichtigten Konsequenzen, denn Ihre Benutzer können nicht mehr auf Ihre Anwendungen zugreifen.
- Sie ändern eine Methode, die von Ihrer neuen Funktion aufgerufen wird. Eine andere Funktion war ebenfalls von dieser Methode abhängig und funktioniert nicht mehr. Das Problem bleibt unbemerkt und wird in die Produktion aufgenommen. Die andere Funktion wird für einige Zeit nicht aufgerufen und schlägt schließlich in der Produktion fehl, ohne dass die Ursache klar wäre.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie Änderungen frühzeitig testen und validieren, können Sie Probleme mit minimalen Kosten beheben und die Auswirkungen auf Ihre Kunden einschränken. Durch Tests vor der Bereitstellung minimieren Sie die Fehler.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Testen und Validieren von Änderungen: Sie sollten in allen Phasen des Lebenszyklus (zum Beispiel bei Entwicklung, Test und Produktion) die Änderungen testen und die Ergebnisse validieren. Prüfen Sie anhand der Testergebnisse neue Funktionen und minimieren Sie das Risiko und die Auswirkung fehlgeschlagener Bereitstellungen. Automatisieren Sie Testverfahren und Validierungen, um eine einheitliche Prüfung zu gewährleisten, Fehler aufgrund von manuellen Prozessen zu reduzieren und den Aufwand zu verringern.
  - Was ist AWS CodeBuild?
  - Lokale Build-Unterstützung für AWS CodeBuild

Ressourcen

Zugehörige Dokumente:

- AWS-Entwicklertools
- Lokale Build-Unterstützung für AWS CodeBuild
- Was ist AWS CodeBuild?

OPS05-BP03 Einsatz von Systemen zur Konfigurationsverwaltung

Verwenden Sie Systeme zur Konfigurationsverwaltung, um Änderungen vorzunehmen und zu verfolgen. Diese Systeme reduzieren Fehler aufgrund von manuellen Prozessen und verringern den Testaufwand.

Beim statischen Konfigurationsmanagement werden Werte festgelegt, wenn eine Ressource initialisiert wird, die erwartungsgemäß während der Lebensdauer der Ressource konsistent bleibt. Einige Beispiele sind die Konfiguration eines Web- oder Anwendungsservers auf einer Instance oder die Definition der Konfiguration eines AWS-Service innerhalb der AWS Management Console oder durch die AWS CLI.

Beim dynamischen Konfigurationsmanagement werden bei der Initialisierung Werte festgelegt, die sich während der Lebensdauer einer Ressource ändern können oder voraussichtlich ändern werden. So können Sie zum Beispiel durch eine Konfigurationsänderung eine Funktion in Ihrem Code aktivieren oder während eines Vorfalls den Detaillierungsgrad des Protokolls ändern, um mehr Daten zu erfassen, und dann nach dem Vorfall wieder zum Ursprungswert zurückkehren, um unnötige Protokolle und damit verbundene Kosten zu vermeiden.

Wenn Sie dynamische Konfigurationen in Ihren Anwendungen haben, die auf Instances, Containern, serverlosen Funktionen oder Geräten ausgeführt werden, können Sie AWS AppConfig zur Verwaltung und Bereitstellung in Ihren gesamten Umgebungen verwenden.

In AWS können Sie AWS Config zur kontinuierlichen Überwachung Ihrer AWS-Ressourcenkonfigurationen über Konten und Regionen hinweg verwenden. So können Sie den Konfigurationsverlauf verfolgen, nachvollziehen, wie sich eine Konfigurationsänderung auf andere Ressourcen auswirkt, und sie mit den erwarteten oder gewünschten Konfigurationen mithilfe von AWS-Config-Regeln und AWS Config Conformance Packsüberprüfen.

In AWS können Sie CI/CD-Pipelines (Continuous Integration/Continuous Deployment) unter Verwendung von Services wie den AWS-Entwicklertools (z. B. AWS CodeCommit, AWS CodeBuild, AWS CodePipeline, AWS CodeDeployund AWS CodeStar) erstellen.

Legen Sie einen Änderungskalender an und verfolgen Sie, wann wichtige geschäftliche oder betriebliche Aktivitäten oder Ereignisse geplant sind, die durch die Implementierung von Änderungen beeinträchtigt werden könnten. Passen Sie Aktivitäten an, um Risiken im Zusammenhang mit diesen Plänen zu verwalten. AWS Systems Manager Change Calendar bietet einen Mechanismus zum Dokumentieren von Zeitblöcken als offen oder geschlossen für Änderungen inklusive Grund und gibt diese Informationen an andere AWS-Konten weiter. AWS Systems Manager Automation-Skripts können so konfiguriert werden, dass sie den Status des Änderungskalenders einhalten.

AWS Systems Manager Maintenance Windows können verwendet werden, um die Leistung von AWS SSM Run Command- oder Automatisierungsskripts, AWS Lambda-Aufrufen oder AWS Step Functions-Aktivitäten zu bestimmten Zeiten zu planen. Markieren Sie diese Aktivitäten in Ihrem Kalender, damit sie in Ihre Auswertung aufgenommen werden können.

Gängige Antimuster:

- Sie aktualisieren die Konfigurationen aller Webserver manuell und eine Reihe von Servern reagiert aufgrund von Updatefehlern nicht mehr.
- Sie aktualisieren Ihre Anwendungsserver mehrere Stunden lang auf manuelle Weise. Die Inkonsistenz der Konfiguration während der Änderung führt zu unerwarteten Verhaltensweisen.
- Jemand hat Ihre Sicherheitsgruppen aktualisiert und auf Ihre Webserver kann nicht mehr zugegriffen werden. Sie wissen nicht, was geändert wurde, und verbringen viel Zeit mit der Suche nach dem Problem – die Zeit bis zur Wiederherstellung nimmt zu.

Vorteile der Einführung dieser bewährten Methode: Die Einführung von Konfigurationsverwaltungssystemen reduziert den Aufwand für die Durchführung und Nachverfolgung von Änderungen sowie die Häufigkeit der durch manuelle Verfahren verursachten Fehler.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Konfigurationsverwaltungssysteme verwenden: Verwenden Sie Systeme zur Konfigurationsverwaltung für die Nachverfolgung und Implementierung von Änderungen, Reduzierung von Fehlern, die durch manuelle Prozesse entstehen, und zur Verringerung des Aufwands.

- Verwaltung der Infrastrukturkonfiguration
- AWS Config
- Was ist AWS Config?
- Einführung in AWS CloudFormation
- Was ist AWS CloudFormation?
- AWS OpsWorks
- Was ist AWS OpsWorks?
- Einführung in AWS Elastic Beanstalk
- Was ist AWS Elastic Beanstalk?

Ressourcen

Zugehörige Dokumente:

- AWS AppConfig
- AWS-Entwicklertools
- AWS OpsWorks
- AWS Systems Manager Change Calendar
- AWS Systems Manager Maintenance Windows
- Verwaltung der Infrastrukturkonfiguration
- Was ist AWS CloudFormation?
- Was ist AWS Config?
- Was ist AWS Elastic Beanstalk?
- Was ist AWS OpsWorks?

Relevante Videos:

- Einführung in AWS CloudFormation
- Einführung in AWS Elastic Beanstalk

OPS05-BP04 Einsatz von Systemen zur Build- und Bereitstellungsverwaltung.

Verwenden Sie Systeme zur Build- und Bereitstellungsverwaltung. Diese Systeme reduzieren Fehler aufgrund von manuellen Prozessen und verringern den Testaufwand.

In AWS können Sie CI/CD-Pipelines (Continuous Integration/Continuous Deployment) unter Verwendung von Services wie den AWS-Entwicklertools (z. B. AWS CodeCommit, AWS CodeBuild, AWS CodePipeline, AWS CodeDeployund AWS CodeStar) erstellen.

Gängige Antimuster:

- Nachdem Sie Ihren Code auf Ihrem Entwicklungssystem kompiliert haben, kopieren Sie die ausführbare Datei auf Ihre Produktionssysteme und sie kann nicht gestartet werden. Die lokalen Protokolldateien zeigen an, dass die Ausführung aufgrund fehlender Abhängigkeiten fehlgeschlagen ist.
- Sie erstellen Ihre Anwendung erfolgreich mit neuen Funktionen in Ihrer Entwicklungsumgebung und übergeben den Code zur QA-Prüfung (Quality Assurance). Die QA-Prüfung schlägt fehl, da statische Komponenten fehlen.
- Am Freitag haben Sie Ihre Anwendung nach großem Aufwand manuell in Ihrer Entwicklungsumgebung erstellt, einschließlich der neu geschriebenen Funktionen. Am Montag können Sie die Schritte, mit denen Sie Ihre Anwendung erfolgreich erstellen konnten, nicht wiederholen.
- Sie führen die Tests durch, die Sie für den neuen Release erstellt haben. Sie verbringen die nächste Woche damit, eine Testumgebung einzurichten und alle vorhandenen Integrationstests durchzuführen, gefolgt von den Leistungstests. Der neue Code bewirkt eine inakzeptable Leistungsbeeinträchtigung und muss neu entwickelt und dann erneut getestet werden.

Vorteile der Einführung dieser bewährten Methode: Mithilfe von Mechanismen zur Verwaltung von Erstellungs- und Bereitstellungsaktivitäten reduzieren Sie den Aufwand für wiederholte Aufgaben, verschaffen Ihren Teammitgliedern die Zeit, sich auf ihre wichtigen Aufgaben zu konzentrieren, und begrenzen die Entstehung von Fehlern durch manuelle Verfahren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Einsatz von Systemen zur Build- und Bereitstellungsverwaltung: Verwenden Sie Systeme zur Build- und Bereitstellungsverwaltung für die Verfolgung und Implementierung von Änderungen, die

Reduzierung von Fehlern, die durch manuelle Prozesse entstehen, sowie zur Verringerung des Aufwands. Nutzen Sie eine vollständig automatisierte Integrations- und Bereitstellungs-Pipeline vom Einchecken des Codes über das Testen und die Bereitstellung bis hin zur Validierung. Dies verkürzt die Vorlaufzeit, ermöglicht häufigere Änderungen und verringert den Aufwand.

- Was ist AWS CodeBuild?
- Best Practices der fortlaufenden Integration bei der Softwareentwicklung
- Slalom: CI/CD für Serverless Anwendungen in AWS
- Einführung in AWS CodeDeploy – automatisierte Softwarebereitstellung mit Amazon Web Services
- Was ist AWS CodeDeploy?

Ressourcen

Zugehörige Dokumente:

- AWS-Entwicklertools
- Was ist AWS CodeBuild?
- Was ist AWS CodeDeploy?

Relevante Videos:

- Best Practices der fortlaufenden Integration bei der Softwareentwicklung
- Einführung in AWS CodeDeploy – automatisierte Softwarebereitstellung mit Amazon Web Services
- Slalom: CI/CD für Serverless Anwendungen in AWS

OPS05-BP05 Durchführen der Patch-Verwaltung

Führen Sie eine Patch-Verwaltung durch, um Funktionen zu erhalten, Probleme zu beheben und die Konformität mit der Governance zu gewährleisten. Automatisieren Sie die Patch-Verwaltung, um Fehler aufgrund von manuellen Prozessen zu reduzieren und den Aufwand für die Installation von Patches zu verringern.

Patch- und Schwachstellenmanagement sind Teil Ihrer Vorteile- und Risikomanagement-Aktivitäten. Es ist vorzuziehen, unveränderliche Infrastrukturen zu haben und Workloads in verifizierten bekannten guten Zuständen bereitzustellen. Wenn dies nicht realisierbar ist, ist das Patchen die verbleibende Option.

Das Aktualisieren von Computerabbildern, Container-Abbildern oder benutzerdefinierten Lambda-Laufzeiten und zusätzlichen Bibliotheken, um Schwachstellen zu entfernen, ist Teil der Patch-Verwaltung. Sie sollten Updates für Amazon Machine Images (AMIs) für Linux- oder Windows Server-Images mit EC2 Image Builderverwalten. Sie können Amazon Elastic Container Registry mit Ihrer vorhandenen Pipeline verwenden, um Amazon ECS Images und Amazon EKS Imageszu verwalten. AWS Lambda umfasst Versionsverwaltungsfunktionen .

Patches sollten nicht auf Produktionssystemen ohne erste Tests in einer sicheren Umgebung durchgeführt werden. Patches sollten nur angewendet werden, wenn sie ein betriebliches oder geschäftliches Ergebnis unterstützen. In AWS können Sie AWS Systems Manager Patch Manager verwenden, um das Patchen verwalteter Systeme zu automatisieren und die Aktivitäten mithilfe von AWS Systems Manager Maintenance Windows.

Gängige Antimuster:

- Sie erhalten den Auftrag, alle neuen Sicherheits-Patches innerhalb von zwei Stunden anzuwenden, was zu mehreren Ausfällen aufgrund der Anwendungsinkompatibilität mit bestimmten Patches führt.
- Eine ungepatchte Bibliothek hat unbeabsichtigte Folgen, weil unbekannte Personen Schwachstellen darin verwenden, um auf Ihren Workload zuzugreifen.
- Sie patchen die Entwicklerumgebungen automatisch, ohne die Entwickler zu benachrichtigen. Sie erhalten mehrere Beschwerden von den Entwicklern, dass ihre Umgebung nicht mehr wie erwartet funktioniert.
- Sie haben die kommerziell im Handel erhältliche Software auf einer persistenten Instance nicht gepatcht. Als ein Problem mit der Software auftritt und Sie sich an den Anbieter wenden, werden Sie darüber informiert, dass die Version nicht unterstützt wird und Sie bestimmte Patches installieren müssen, um Unterstützung zu erhalten.
- Ein kürzlich veröffentlichter Patch für Ihre verwendete Verschlüsselungssoftware bietet signifikante Leistungsverbesserungen. Ihr ungepatchtes System weist Leistungsprobleme auf, die bestehen bleiben, weil es nicht gepatcht ist.

Vorteile der Einführung dieser bewährten Methode: Durch die Einrichtung eines Patch-Verwaltungsprozesses, einschließlich Ihrer Patching-Kriterien und Bereitstellungsmethodik für Ihre Umgebungen, können Sie ihre Vorteile nutzen und ihre Auswirkungen kontrollieren. Dies ermöglicht das Übernehmen der gewünschten Merkmale und Funktionen, das Entfernen von Problemen und die kontinuierliche Compliance. Implementieren Sie Verwaltungssysteme und Automatisierung für

Patches, um den Aufwand für die Bereitstellung von Patches zu reduzieren und Fehler zu begrenzen, die durch manuelle Prozesse verursacht werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Patch-Verwaltung: Installieren Sie auf Ihren Systemen Patches zur Behebung von Problemen, zur Erlangung der gewünschten Funktionen oder Fähigkeiten sowie zur kontinuierlichen Einhaltung der Governance-Richtlinien und der Anforderungen des Lieferantensupport. Nehmen Sie in unveränderlichen Systemen eine Bereitstellung mit einer geeigneten Patch-Gruppe vor, um das gewünschte Ergebnis zu erzielen. Automatisieren Sie den Mechanismus der Patch-Verwaltung, um die Patch-Zeit zu verkürzen, Fehler aufgrund von manuellen Prozessen zu reduzieren und den Aufwand für die Installation von Patches zu verringern.
  - AWS Systems Manager Patch Manager

Ressourcen

Zugehörige Dokumente:

- AWS-Entwicklertools
- AWS Systems Manager Patch Manager

Relevante Videos:

- CI/CD für Serverless Anwendungen in AWS
- Design mit Blick auf die Ops

Zugehörige Beispiele:

- Well-Architected Labs – Bestands- und Patch-Verwaltung

OPS05-BP06 Gemeinsame Design-Standards

Tauschen Sie teamübergreifend bewährte Methoden aus, um das Bewusstsein zu schärfen und den Nutzen der Entwicklungsarbeit zu maximieren.

Auf AWS können Anwendung, Computing, Infrastruktur und Betrieb mit Hilfe von Code-Methoden definiert und verwaltet werden. Dadurch gestalten sich Veröffentlichungen, Freigaben und Einführungen ganz einfach.

Viele AWS-Services und -Ressourcen sind so ausgelegt, dass sie von mehreren Konten gemeinsam genutzt werden können. Dies ermöglicht es Ihnen, erstellte Assets und Erkenntnisse teamübergreifend freizugeben. Sie können beispielsweise CodeCommit -Repositorys, Lambda -Funktionen, Amazon S3-Bucketsund AMIs für bestimmte Konten freigeben.

Nutzen Sie beim Veröffentlichen neuer Ressourcen oder Aktualisierungen Amazon SNS, um Benachrichtigungen über Konten hinwegzu veröffentlichen. Abonnenten können Lambda verwenden, um neue Versionen zu erhalten.

Wenn gemeinsame Standards in Ihrem Unternehmen durchgesetzt werden, ist es wichtig, dass Mechanismen vorhanden sind, um Ergänzungen, Änderungen und Ausnahmen von Standards zur Unterstützung der Teamaktivitäten anzufordern. Ohne diese Option werden Standards zu einer Einschränkung der Innovation.

Gängige Antimuster:

- Sie haben wie jedes der anderen Entwicklungsteams in Ihrem Unternehmen Ihren eigenen Benutzerauthentifizierungsmechanismus erstellt. Ihre Benutzer müssen für jeden Teil des Systems, auf den sie zugreifen möchten, eigene Anmeldeinformationen verwenden.
- Sie haben wie jedes der anderen Entwicklungsteams in Ihrem Unternehmen Ihren eigenen Benutzerauthentifizierungsmechanismus erstellt. Ihr Unternehmen erhält eine neue Compliance-Anforderung, die erfüllt werden muss. Jedes einzelne Entwicklungsteam muss jetzt die erforderlichen Ressourcen investieren, um die neue Anforderung zu erfüllen.
- Sie haben wie jedes der anderen Entwicklungsteams in Ihrem Unternehmen ein eigenes Bildschirmlayout erstellt. Ihre Benutzer beschweren sich über die Schwierigkeit, durch die inkonsistenten Oberflächen zu navigieren.

Vorteile der Einführung dieser bewährten Methode: Verwenden Sie gemeinsame Standards, um die Übernahme bewährter Methoden zu unterstützen und die Vorteile der Entwicklungsbemühungen zu maximieren, wenn Standards die Anforderungen für mehrere Anwendungen oder Organisationen erfüllen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Gemeinsame Design-Standards: Tauschen Sie bestehende Best Practices, Design-Standards, Checklisten, Betriebsverfahren sowie Richtlinien und Governance-Anforderungen in Teams aus, um die Abläufe unkomplizierter zu gestalten und den Nutzen der Entwicklungsarbeit zu maximieren. Stellen Sie sicher, dass es Verfahren für die Beantragung von Änderungen, Ergänzungen und Ausnahmen von Design-Standards gibt. Auf diese Weise schaffen Sie Raum für kontinuierliche Verbesserungen und Innovationen. Sorgen Sie dafür, dass die Teams veröffentlichte Inhalte kennen, damit sie davon profitieren, weniger Überarbeitungen erforderlich sind und keine Ressourcen vergeudet werden.
  - Delegieren des Zugriffs auf Ihre AWS-Umgebung
  - Freigeben eines AWS CodeCommit-Repositorys
  - Unkomplizierte Autorisierung von AWS Lambda-Funktionen
  - Freigeben eines AMI mit bestimmten AWS-Konten
  - Schnelles Freigeben von Vorlagen mit einer AWS CloudFormation-Designer-URL
  - Verwenden von AWS Lambda mit Amazon SNS

Ressourcen

Zugehörige Dokumente:

- Unkomplizierte Autorisierung von AWS Lambda-Funktionen
- Freigeben eines AWS CodeCommit-Repositorys
- Freigeben eines AMI mit bestimmten AWS-Konten
- Schnelles Freigeben von Vorlagen mit einer AWS CloudFormation-Designer-URL
- Verwenden von AWS Lambda mit Amazon SNS

Relevante Videos:

- Delegieren des Zugriffs auf Ihre AWS-Umgebung

OPS05-BP07 Implementieren von Verfahren zur Verbesserung der Codequalität

Implementieren Sie Verfahren zur Verbesserung der Codequalität und Minimierung von Fehlern. Einige Beispiel sind testbasierte Entwicklungen, Codeprüfungen und die Einführung von Standards.

In AWS können Sie Services wie Amazon CodeGuru in Ihre Pipeline integrieren, um potenzielle Code- und Sicherheitsprobleme mithilfe von Programmanalyse und Machine Learning zu identifizieren. CodeGuru bietet Empfehlungen zur Implementierung der bewährten AWS-Methoden, um diese Probleme zu beheben.

Gängige Antimuster:

- Damit Sie Ihre Funktion früher testen können, haben Sie sich entschieden, Ihre standardmäßige Bibliothek für die Eingabekorrektur nicht zu integrieren. Nach dem Testen bestätigen Sie Ihren Code und vergessen dabei, die Bibliothek zu integrieren.
- Sie haben nur sehr wenig Erfahrung mit dem zu verarbeitenden Datensatz und wissen nicht, dass es eine Reihe von Grenzfällen gibt, die in diesem Datensatz vorhanden sein können. Diese Grenzfälle sind nicht mit dem Code kompatibel, den Sie implementiert haben.

Vorteile der Einführung dieser bewährten Methode: Durch die Übernahme von Methoden zur Verbesserung der Codequalität können Sie die Anzahl der Probleme minimieren, die bei der Produktion noch vorhanden sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Implementieren von Verfahren zur Verbesserung der Codequalität: Implementieren Sie Verfahren zur Verbesserung der Codequalität, um Fehler und das Risiko der Implementierung von Fehlern zu minimieren. Geeignete Maßnahmen sind zum Beispiel testbasierte Entwicklungen, Paarprogrammierung, Codeprüfungen und die Einführung von Standards.
  - Amazon CodeGuru

Ressourcen

Zugehörige Dokumente:

- Amazon CodeGuru

OPS05-BP08 Verwenden mehrerer Umgebungen

Verwenden Sie mehrere Umgebungen, um Ihren Workload auszuprobieren, zu entwickeln und zu testen. Verwenden Sie zunehmende Kontrollstufen, wenn Umgebungen sich der Produktion nähern, um sicherzustellen, dass Ihr Workload bei der Bereitstellung wie beabsichtigt funktioniert.

Gängige Antimuster:

- Sie führen die Entwicklung in einer gemeinsamen Entwicklungsumgebung durch und ein weiterer Entwickler überschreibt Ihre Codeänderungen.
- Die restriktiven Sicherheitskontrollen Ihrer gemeinsamen Entwicklungsumgebung verhindern, dass Sie mit neuen Services und Funktionen experimentieren können.
- Sie führen Belastungstests auf Ihren Produktionssystemen durch und verursachen einen Ausfall für Ihre Benutzer.
- In der Produktion ist ein kritischer Fehler aufgetreten, der zum Verlust von Daten geführt hat. In Ihrer Produktionsumgebung versuchen Sie, die Bedingungen, die zum Datenverlust geführt haben, nachzustellen, damit Sie die Ursache feststellen und beseitigen können. Um einen weiteren Datenverlust während des Testens zu verhindern, müssen Sie die Anwendung für Ihre Benutzer deaktivieren.
- Sie betreiben einen Mehrmandanten-Service und können eine Kundenanfrage nach einer eigenen Umgebung nicht erfüllen.
- Sie testen nicht immer, aber wenn, dann in der Produktion.
- Sie glauben, dass die Einfachheit einer einzelnen Umgebung die Auswirkungen von Änderungen innerhalb der Umgebung ausgleicht.

Vorteile der Einführung dieser bewährten Methode: Durch die Bereitstellung mehrerer Umgebungen können Sie gleichzeitig mehrere Entwicklungs-, Test- und Produktionsumgebungen unterstützen, ohne Konflikte zwischen Entwicklern oder User-Communities zu erzeugen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Verwenden mehrerer Umgebungen: Stellen Sie den Entwicklern Sandbox-Umgebungen mit weniger Kontrollen zur Verfügung, in denen sie experimentieren können. Richten Sie individuelle Entwicklungsumgebungen ein, damit parallele Arbeit möglich ist. Dadurch steigern Sie die Agilität der Entwicklung. Implementieren Sie strengere Kontrollen erst in den Umgebungen, die kurz vor

der Produktionsaufnahme stehen, damit Entwickler Innovationen schaffen können. Nutzen Sie die Infrastruktur als Code sowie Konfigurationsverwaltungssysteme, um Umgebungen bereitzustellen, die mit den in der Produktion vorhandenen Kontrollen einheitlich konfiguriert sind. Auf diese Weise können Sie sicherstellen, dass die Systeme bei der Bereitstellung wie erwartet funktionieren. Wenn Umgebungen nicht in Gebrauch sind, schalten Sie sie ab, um Kosten für ungenutzte Ressourcen zu vermeiden (z. B. Entwicklungssysteme am Abend und am Wochenende). Stellen Sie beim Belastungstest produktionsgleiche Umgebungen bereit, um stichhaltige Ergebnisse zu erzielen.

- Was ist AWS CloudFormation?
- Wie beende und starten ich Amazon EC2-Instances mit AWS Lambda in festgelegten Intervallen?

Ressourcen

Zugehörige Dokumente:

- Wie beende und starten ich Amazon EC2-Instances mit AWS Lambda in festgelegten Intervallen?
- Was ist AWS CloudFormation?

Häufige, kleine, umkehrbare Änderungen vornehmen:

Gängige Antimuster:

- Sie stellen vierteljährlich eine neue Version Ihrer Anwendung bereit.
- Sie nehmen häufig Änderungen an Ihrem Datenbankschema vor.
-

Vorteile der Einführung dieser bewährten Methode:

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird:

Implementierungsleitfaden

-

OPS05-BP10 Vollständige Automatisierung von Integration und Bereitstellung

Automatisieren Sie den Aufbau, die Bereitstellung und die Tests des Workloads. Dadurch werden Fehler aufgrund von manuellen Prozessen und der Aufwand für die Bereitstellung von Änderungen verringert.

Wenden Sie Metadaten mithilfe von Ressourcen-Tags und AWS Resource Groups nach einer konsistenten Markierungsstrategie an, um die Identifizierung Ihrer Ressourcen zu ermöglichen. Versehen Sie Ihre Ressourcen mit Tags für Organisation, Kostenkalkulation, Zugriffssteuerung und Zielrichtung der Ausführung von automatisierten Betriebsaktivitäten.

Gängige Antimuster:

- Am Freitag schreiben Sie den neuen Code für Ihren Funktionszweig fertig. Am Montag, nach dem Ausführen Ihrer Skripts für die Codequalitätstests und einzelnen Komponententests, werden Sie Ihren Code für den nächsten geplanten Release überprüfen.
- Sie erhalten die Aufgabe, eine Korrektur für ein kritisches Problem zu schreiben, das sich auf eine große Anzahl von Kunden in der Produktion auswirkt. Nachdem Sie die Korrektur getestet haben, übergeben Sie Ihren Code und fordern beim Änderungsmanagenement die Bereitstellungsgenehmigung zur Produktion an.

Vorteile der Einführung dieser bewährten Praxis: Durch die Implementierung automatisierter Build- und Bereitstellungsverwaltungssysteme reduzieren Sie Fehler von manuellen Prozessen und den Aufwand für die Bereitstellung von Änderungen, sodass sich Ihre Teammitglieder auf die Wertschöpfung konzentrieren können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwendung von Build- und Deployment-Management-Systemen: Verwenden Sie Build- und Deployment-Managementsysteme, um Änderungen zu verfolgen und zu implementieren, Fehler durch manuelle Prozesse zu reduzieren und den Aufwand zu verringern. Nutzen Sie eine vollständig automatisierte Integrations- und Bereitstellungs-Pipeline vom Einchecken des Codes über das Testen und die Bereitstellung bis hin zur Validierung. Dies verkürzt die Vorlaufzeit, ermöglicht häufigere Änderungen und verringert den Aufwand.
  - Was ist AWS CodeBuild?
  - Best Practices der fortlaufenden Integration bei der Softwareentwicklung

- Slalom: CI/CD für serverlose Anwendungen auf
- Einführung in die - automatische Softwareverteilung mit
- Was ist AWS CodeDeploy?

Ressourcen

Verbundene Dokumente:

- Was ist AWS CodeBuild?
- Was ist AWS CodeDeploy?

Verbundene Videos:

- Best Practices der fortlaufenden Integration bei der Softwareentwicklung
- Einführung in die - automatische Softwareverteilung mit
- Slalom: CI/CD für serverlose Anwendungen auf

OPS 6 Wie können Sie Bereitstellungsrisiken eindämmen?

Verwenden Sie Ansätze, die ein schnelles Feedback zur Qualität liefern und eine umgehende Wiederherstellung des vorherigen Zustands nach Änderungen ermöglichen, die nicht zu den gewünschten Ergebnissen führen. Mit diesen Verfahren können Sie die Auswirkung von Problemen eindämmen, die durch die Bereitstellung von Änderungen entstehen.

Bewährte Methoden

- OPS06-BP01 Einkalkulieren nicht erfolgreicher Änderungen
- OPS06-BP02 Testen und Validieren von Änderungen
- OPS06-BP03 Verwenden von Systemen zur Bereitstellungsverwaltung
- OPS06-BP04 Testen mit begrenzten Bereitstellungen
- OPS06-BP05 Bereitstellung unter Verwendung paralleler Umgebungen
- OPS06-BP06 Bereitstellen häufiger, kleiner und umkehrbarer Änderungen
- OPS06-BP07 Vollständige Automatisierung von Integration und Bereitstellung
- OPS06-BP08 Automatisieren von Tests und Rollback

OPS06-BP01 Einkalkulieren nicht erfolgreicher Änderungen

Planen Sie Maßnahmen für die Rückkehr zu einem bekanntermaßen funktionierenden Zustand oder die Korrektur in der Produktionsumgebung ein, falls eine Änderung nicht das gewünschte Ergebnis bewirkt. Dank dieser Vorbereitung verkürzt sich die Wiederherstellungszeit, da schneller reagiert werden kann.

Gängige Antimuster:

- Sie haben Code bereitgestellt und Ihre Anwendung ist instabil geworden, aber es befinden sich aktive Benutzer im System. Sie müssen entscheiden, ob Sie die Änderung rückgängig machen und Auswirkungen auf die aktiven Benutzer in Kauf nehmen möchten, oder ob Sie die Änderung erst später rückgängig machen möchten, wodurch möglicherweise trotzdem Auswirkungen auf die Benutzer entstehen könnten.
- Nachdem Sie eine Routing-Änderung vorgenommen haben, kann auf Ihre neuen Umgebungen zugegriffen werden, aber eines Ihrer Subnetze ist nicht mehr erreichbar. Sie müssen entscheiden, ob Sie die gesamte Änderung rückgängig machen oder versuchen, die Nichtverfügbarkeit des Subnetzes zu beheben. Während Sie diese Entscheidung abwägen, bleibt das Subnetz nicht erreichbar.

Vorteile der Einführung dieser bewährten Methode: Ein Plan verringert die mittlere Reparaturzeit (Mean Time to Recover, MTTR), um sich von Fehlschlägen bei Änderungen zu erholen. Dadurch verringern sich auch die Auswirkungen auf Endbenutzer.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Einkalkulieren nicht erfolgreicher Änderungen: Planen Sie Maßnahmen für die Rückkehr zu einem bekanntermaßen funktionierenden Zustand („Rollback“ der Änderung) oder die Korrektur in der Produktionsumgebung („Rollforward“ der Änderung) ein, falls eine Änderung nicht zum gewünschten Ergebnis führt. Falls Sie Änderungen finden, die im Fall eines Misserfolgs nicht zurückgesetzt werden können, seien Sie vor der Festschreibung der Änderung sehr vorsichtig.

OPS06-BP02 Testen und Validieren von Änderungen

Testen Sie Änderungen und validieren Sie die Ergebnisse in allen Phasen des Lebenszyklus. Auf diese Weise können Sie neue Funktionen prüfen und das Risiko und die Auswirkungen fehlgeschlagener Bereitstellungen minimieren.

In AWS können Sie temporäre parallele Umgebungen erstellen. Das senkt die Risiken, Mühen und Kosten, die mit dem Experimentieren und Testen verbunden sind. Automatisieren Sie die Bereitstellung dieser Umgebungen mithilfe von AWS CloudFormation um eine konsistente Implementierung Ihrer temporären Umgebungen sicherzustellen.

Gängige Antimuster:

- Sie stellen eine neue Funktion für Ihre Anwendung bereit. Sie funktioniert nicht. Sie wissen das nicht.
- Sie aktualisieren Ihre Zertifikate. Sie installieren die Zertifikate versehentlich für die falschen Komponenten. Sie wissen das nicht.

Vorteile der Einführung dieser bewährten Methode: Durch das Testen und Validieren von Änderungen nach der Bereitstellung können Sie Probleme frühzeitig identifizieren und so die Auswirkungen auf Ihre Kunden minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Testen und Validieren von Änderungen: Testen Sie Änderungen und validieren Sie die Ergebnisse in allen Phasen des Lebenszyklus, zum Beispiel in den Entwicklungs-, Test- und Produktionsphasen. Auf diese Weise können Sie neue Funktionen prüfen und das Risiko und die Auswirkungen fehlgeschlagener Bereitstellungen minimieren.
  - AWS Cloud9
  - Was ist AWS Cloud9?
  - Vorgehensweise für den lokalen Test und lokales Debugging von AWS CodeDeploy vor der Auslieferung Ihres Codes

Ressourcen

Zugehörige Dokumente:

- AWS Cloud9
- AWS-Entwicklertools
- Vorgehensweise für den lokalen Test und lokales Debugging von AWS CodeDeploy vor der Auslieferung Ihres Codes

- Was ist AWS Cloud9?

OPS06-BP03 Verwenden von Systemen zur Bereitstellungsverwaltung

Verwenden Sie Systeme zur Bereitstellungsverwaltung, um Änderungen zu verfolgen und zu implementieren. Dadurch werden Fehler aufgrund von manuellen Prozessen und der Aufwand für die Bereitstellung von Änderungen verringert.

In AWS können Sie CI/CD-Pipelines (Continuous Integration/Continuous Deployment) unter Verwendung von Services wie den AWS-Entwicklertools (z. B. AWS CodeCommit, AWS CodeBuild, AWS CodePipeline, AWS CodeDeployund AWS CodeStar) erstellen.

Gängige Antimuster:

- Sie stellen Updates manuell auf Ihren Anwendungsservern bereit und eine Reihe von Servern reagiert aufgrund von Updatefehlern nicht mehr.
- Sie verbringen viele Stunden damit, Änderungen manuell auf den Anwendungsservern bereitzustellen. Die Inkonsistenz bei den Versionen während der Änderung führt zu unerwarteten Verhaltensweisen.

Vorteile der Einführung dieser bewährten Methode: Die Einführung von Systemen zur Bereitstellungsverwaltung reduziert den Aufwand für die Bereitstellung von Änderungen und die Häufigkeit der durch manuelle Verfahren verursachten Fehler.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Bereitstellungsverwaltungssysteme verwenden: Verwenden Sie Bereitstellungsverwaltungssysteme, um Änderungen nachzuverfolgen und zu implementieren. Dadurch reduzieren Sie Fehler aufgrund von manuellen Prozessen und verringern den Aufwand für die Bereitstellung von Änderungen. Automatisieren Sie die Integrations- und Bereitstellungs-Pipeline vom Einchecken des Codes über das Testen und die Bereitstellung bis hin zur Validierung. Dies verkürzt die Vorlaufzeit, ermöglicht häufigere Änderungen und verringert den Aufwand noch weiter.
  - Einführung in AWS CodeDeploy – automatisierte Softwarebereitstellung mit Amazon Web Services
  - Was ist AWS CodeDeploy?

- Was ist AWS Elastic Beanstalk?
- Was ist Amazon API Gateway?

Ressourcen

Zugehörige Dokumente:

- AWS CodeDeploy-Benutzerhandbuch
- AWS-Entwicklertools
- Testen Sie eine Blau-/Grün-Beispielbereitstellung in AWS CodeDeploy
- Was ist AWS CodeDeploy?
- Was ist AWS Elastic Beanstalk?
- Was ist Amazon API Gateway?

Relevante Videos:

- Eingehende Informationen zu modernen Continuous Delivery-Verfahren mit AWS
- Einführung in AWS CodeDeploy – automatisierte Softwarebereitstellung mit Amazon Web Services

OPS06-BP04 Testen mit begrenzten Bereitstellungen

Führen Sie parallel zu den bestehenden Systemen Tests mit begrenzten Bereitstellungen durch, um vor der Gesamtbereitstellung zu prüfen, ob tatsächlich die gewünschten Ergebnisse erzielt werden. Führen Sie beispielsweise Tests mit Bereitstellungen in einer ausgewählten Gruppe oder in nur einem System durch.

Gängige Antimuster:

- Sie stellen eine nicht erfolgreiche Änderung für die gesamte Produktion gleichzeitig bereit. Sie wissen das nicht.

Vorteile der Einführung dieser bewährten Methode: Durch das Testen und Validieren von Änderungen nach einer eingeschränkten Bereitstellung können Sie Probleme frühzeitig mit minimalen Auswirkungen auf Ihre Kunden identifizieren und so die Auswirkungen auf Ihre Kunden weiter minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mit begrenzten Bereitstellungen testen: Führen Sie parallel zu den bestehenden Systemen Tests mit begrenzten Bereitstellungen durch, um vor der Gesamtbereitstellung zu prüfen, ob tatsächlich die gewünschten Ergebnisse erzielt werden. Führen Sie beispielsweise Tests mit Bereitstellungen in einer ausgewählten Gruppe oder in nur einem System durch.
  - AWS CodeDeploy-Benutzerhandbuch
  - Blau/Grün-Bereitstellungen mit AWS Elastic Beanstalk
  - Einrichten einer API Gateway-Canary-Bereitstellung
  - Testen Sie eine Blau/Grün-Beispielbereitstellung in AWS CodeDeploy
  - Arbeiten mit Bereitstellungskonfigurationen in AWS CodeDeploy

Ressourcen

Zugehörige Dokumente:

- AWS CodeDeploy-Benutzerhandbuch
- Blau-/Grün-Bereitstellungen mit AWS Elastic Beanstalk
- Einrichten einer API Gateway-Canary-Bereitstellung
- Testen Sie eine Blau-/Grün-Beispielbereitstellung in AWS CodeDeploy
- Arbeiten mit Bereitstellungskonfigurationen in AWS CodeDeploy

OPS06-BP05 Bereitstellung unter Verwendung paralleler Umgebungen

Implementieren Sie Änderungen in parallelen Umgebungen und führen Sie dann die Umstellung auf die neue Umgebung durch. Behalten Sie die bisherige Umgebung, bis die erfolgreiche Bereitstellung sichergestellt ist. Dadurch verkürzt sich die Wiederherstellungszeit, da Sie jederzeit zur vorherigen Umgebung zurückkehren können.

Gängige Antimuster:

- Sie führen eine veränderbare Bereitstellung durch, indem Sie Ihre vorhandenen Systeme ändern. Nachdem Sie festgestellt haben, dass die Änderung nicht erfolgreich war, müssen Sie die Systeme erneut ändern, um die alte Version wiederherzustellen, was die Wiederherstellungsdauer verlängert.

- Während eines Wartungszeitfensters nehmen Sie die alte Umgebung außer Betrieb und beginnen dann mit der Erstellung der neuen Umgebung. Nach vielen Stunden Arbeit entdecken Sie nicht korrigierbare Probleme mit der Bereitstellung. Ziemlich erschöpft müssen Sie nun den vorherigen Bereitstellungsablauf finden und mit der Neuerstellung der alten Umgebung beginnen.

Vorteile der Einführung dieser bewährten Methode: Durch die Verwendung von parallelen Umgebungen können Sie die neue Umgebung vorerst bereitstellen und bei Bedarf wechseln. Wenn die neue Umgebung nicht funktioniert, können Sie eine schnelle Wiederherstellung durchführen, indem Sie zurück zu Ihrer ursprünglichen Umgebung wechseln.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Unter Verwendung paralleler Umgebungen bereitstellen: Implementieren Sie Änderungen in parallelen Umgebungen und wechseln Sie dann in die neue Umgebung. Behalten Sie die bisherige Umgebung, bis die erfolgreiche Bereitstellung sichergestellt ist. Dadurch verkürzt sich die Wiederherstellungszeit, da Sie jederzeit zur vorherigen Umgebung zurückkehren können. Verwenden Sie beispielsweise unveränderliche Infrastrukturen mit Blau/Grün-Bereitstellungen.
  - Arbeiten mit Bereitstellungskonfigurationen in AWS CodeDeploy
  - Blau/Grün-Bereitstellungen mit AWS Elastic Beanstalk
  - Einrichten einer API Gateway-Canary-Bereitstellung
  - Testen Sie eine Blau/Grün-Beispielbereitstellung in AWS CodeDeploy

Ressourcen

Zugehörige Dokumente:

- AWS CodeDeploy-Benutzerhandbuch
- Blau-/Grün-Bereitstellungen mit AWS Elastic Beanstalk
- Einrichten einer API Gateway-Canary-Bereitstellung
- Testen Sie eine Blau-/Grün-Beispielbereitstellung in AWS CodeDeploy
- Arbeiten mit Bereitstellungskonfigurationen in AWS CodeDeploy

Relevante Videos:

- Eingehende Informationen zu modernen Continuous Delivery-Verfahren mit AWS

OPS06-BP06 Bereitstellen häufiger, kleiner und umkehrbarer Änderungen

Verringern Sie den Umfang einer Änderung durch häufige, kleine und umkehrbare Änderungen. Dies erleichtert die Fehlersuche und ermöglicht eine schnellere Korrektur, da die Möglichkeit besteht, eine Änderung zurückzusetzen.

Gängige Antimuster:

- Sie stellen vierteljährlich eine neue Version Ihrer Anwendung bereit.
- Sie nehmen häufig Änderungen an Ihrem Datenbankschema vor.
- Sie führen direkte manuelle Updates durch und überschreiben damit bestehende Installationen und Konfigurationen.

Vorteile der Einführung dieser bewährten Methode: Sie profitieren schneller von den Entwicklungsarbeiten, wenn Sie kleine Änderungen häufig bereitstellen. Wenn die Änderungen klein sind, ist es viel einfacher zu erkennen, ob sie unbeabsichtigte Folgen haben. Wenn die Änderungen rückgängig gemacht werden können, ist die Implementierung mit geringeren Risiken verbunden, da die Wiederherstellung vereinfacht wird.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Häufige, kleine, umkehrbare Änderungen vornehmen: Verwenden Sie häufige, kleine und umkehrbare Änderungen, um den Umfang und die Auswirkungen einer Änderung zu reduzieren. Dies erleichtert die Fehlersuche und ermöglicht eine schnellere Korrektur, da die Möglichkeit besteht, eine Änderung zurückzusetzen.

OPS06-BP07 Vollständige Automatisierung von Integration und Bereitstellung

Automatisieren Sie den Aufbau, die Bereitstellung und die Tests des Workloads. Dadurch werden Fehler aufgrund von manuellen Prozessen reduziert und der Aufwand für die Bereitstellung von Änderungen verringert.

Wenden Sie Metadaten mithilfe von Ressourcen-Tags und AWS Resource Groups nach einer konsistenten Markierungsstrategie an, um die Identifizierung Ihrer Ressourcen zu ermöglichen.

Versehen Sie Ihre Ressourcen mit Tags für Organisation, Kostenkalkulation, Zugriffssteuerung und Zielrichtung der Ausführung von automatisierten Betriebsaktivitäten.

Gängige Antimuster:

- Am Freitag schließen Sie die Erstellung des neuen Codes für Ihren Featurebranch ab. Am Montag, nach dem Ausführen Ihrer Skripts für die Codequalitätstests und einzelnen Komponententests, werden Sie Ihren Code für den nächsten geplanten Release überprüfen.
- Sie erhalten die Aufgabe, eine Korrektur für ein kritisches Problem zu schreiben, das sich auf eine große Anzahl von Kunden in der Produktion auswirkt. Nachdem Sie die Korrektur getestet haben, übergeben Sie Ihren Code und fordern beim Änderungsmanagenement die Bereitstellungsgenehmigung zur Produktion an.

Vorteile der Einführung dieser bewährten Praxis: Durch die Implementierung automatisierter Build- und Bereitstellungsverwaltungssysteme reduzieren Sie Fehler von manuellen Prozessen und den Aufwand für die Bereitstellung von Änderungen, sodass sich Ihre Teammitglieder auf die Wertschöpfung konzentrieren können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwendung von Build- und Deployment-Management-Systemen: Verwenden Sie Build- und Deployment-Managementsysteme, um Änderungen zu verfolgen und zu implementieren, Fehler durch manuelle Prozesse zu reduzieren und den Aufwand zu verringern. Nutzen Sie eine vollständig automatisierte Integrations- und Bereitstellungs-Pipeline vom Einchecken des Codes über das Testen und die Bereitstellung bis hin zur Validierung. Dies verkürzt die Vorlaufzeit, ermöglicht häufigere Änderungen und verringert den Aufwand.
  - Was ist AWS CodeBuild?
  - Best Practices der fortlaufenden Integration bei der Softwareentwicklung
  - Slalom: CI/CD für serverlose Anwendungen auf
  - Einführung in die - automatische Softwareverteilung mit
  - Was ist AWS CodeDeploy?
  - Eingehende Informationen zu modernen Continuous Delivery-Verfahren mit AWS

Ressourcen

Verbundene Dokumente:

- Testen Sie eine Blau/Grün-Beispielbereitstellung in AWS CodeDeploy
- Was ist AWS CodeBuild?
- Was ist AWS CodeDeploy?

Verbundene Videos:

- Best Practices der fortlaufenden Integration bei der Softwareentwicklung
- Eingehende Informationen zu modernen Continuous Delivery-Verfahren mit AWS
- Einführung in die - automatische Softwareverteilung mit
- Slalom: CI/CD für serverlose Anwendungen auf

OPS06-BP08 Automatisieren von Tests und Rollback

Automatisieren Sie die Tests von bereitgestellten Umgebungen, um die gewünschten Ergebnisse sicherzustellen. Automatisieren Sie die Zurücksetzung auf einen zuvor bekanntermaßen funktionierenden Zustand, wenn die gewünschten Ergebnisse nicht erzielt werden. So können Sie die Wiederherstellungszeit minimieren und verringern Fehler, die durch manuelle Prozesse entstehen.

Gängige Antimuster:

- Sie stellen Änderungen an Ihrem Workload bereit. Nachdem Sie sehen, dass die Änderung abgeschlossen ist, beginnen Sie mit den Tests, die auf die Bereitstellung folgen müssen. Nachdem sie abgeschlossen sind, bemerken Sie, dass Ihr Workload nicht mehr funktioniert und die Verbindung der Kunden getrennt wird. Sie starten das Rollback zur vorherigen Version. Nach einer langen Problemsuche verlängert sich die Wiederherstellungsdauer zusätzlich durch die neue manuelle Bereitstellung.

Vorteile der Einführung dieser bewährten Methode: Durch das Testen und Validieren von Änderungen nach der Bereitstellung können Sie Probleme sofort identifizieren. Durch das automatische Rollback zur vorherigen Version werden die Auswirkungen auf Ihre Kunden minimiert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Tests und Rollback automatisieren: Automatisieren Sie Tests von bereitgestellten Umgebungen, um die gewünschten Ergebnisse zu bestätigen. Automatisieren Sie die Zurücksetzung auf einen zuvor bekanntermaßen funktionierenden Zustand, wenn die gewünschten Ergebnisse nicht erzielt werden. So können Sie die Wiederherstellungszeit minimieren und verringern Fehler, die durch manuelle Prozesse entstehen. Führen Sie beispielsweise nach der Bereitstellung detaillierte synthetische Benutzertransaktionen durch, überprüfen Sie die Ergebnisse und nehmen Sie bei einem Fehler eine Zurücksetzung vor.
  - Erneutes Bereitstellen und Zurücksetzen einer Bereitstellung mit AWS CodeDeploy

Ressourcen

Zugehörige Dokumente:

- Erneutes Bereitstellen und Zurücksetzen einer Bereitstellung mit AWS CodeDeploy

OPS 7 Wie bringen Sie in Erfahrung, ob Sie für die Unterstützung eines Workloads bereit sind?

Bewerten Sie die betriebliche Bereitschaft Ihres Workloads, Prozesse und Verfahren sowie Ihrer Mitarbeiter, damit Sie die betrieblichen Risiken im Zusammenhang mit Ihrer Workload genau kennen.

Bewährte Methoden

- OPS07-BP01 Sicherstellen des Know-hows der Mitarbeiter
- OPS07-BP02 Sicherstellen einer konsistenten Prüfung der betrieblichen Bereitschaft
- OPS07-BP03 Verwenden von Runbooks zur Durchführung von Verfahren
- OPS07-BP04 Verwenden von Playbooks zum Untersuchen von Problemen
- OPS07-BP05 Treffen fundierter Entscheidungen für die Bereitstellung von Systemen und Änderungen

OPS07-BP01 Sicherstellen des Know-hows der Mitarbeiter

Stellen Sie einen Mechanismus bereit, mit dem Sie prüfen können, ob Sie über ausreichend geschulte Mitarbeiter verfügen, die die betrieblichen Anforderungen erfüllen können. Schulen Sie

Ihre Mitarbeiter und passen Sie die Mitarbeiterkapazität bei Bedarf an, damit Sie immer über genug Ressourcen verfügen.

Stellen Sie sicher, dass Sie über genügend Teammitglieder verfügen, um die operativen Aktivitäten abzudecken, einschließlich der Rufbereitschaftsaktivitäten. Stellen Sie sicher, dass Ihre Teams über die erforderlichen Fähigkeiten verfügen, um erfolgreich mit der Schulung Ihres Workloads, Ihrer operativen Tools und AWS zu arbeiten.

AWS stellt Ressourcen bereit, darunter das Erste Schritte – AWS Resource Center, AWS-Blogs, AWS Online Tech Talks, AWS-Veranstaltungen und -Webinaresowie die AWS Well-Architected Labs, die Anleitungen, Beispiele und detaillierte Walkthroughs zur Schulung Ihrer Teams bieten. Darüber hinaus bietet AWS Training und Zertifizierung bietet einige kostenlose Schulungen durch digitale Kurse im Selbststudium zu den Grundlagen von AWS. Sie können sich auch für eine Schulung registrieren, die von Dozenten geleitet wird, um die AWS-Fähigkeiten und -Fertigkeiten Ihres Teams auszubauen.

Gängige Antimuster:

- Bereitstellen eines Workload, ohne dass die Teammitglieder zum Umgang mit der verwendeten Plattform und den Services qualifiziert sind.
- Bereitstellen eines Workload, ohne dass die Teammitglieder während der geplanten Zeiten verfügbar sind.
- Bereitstellen eines Workload, ohne dass genug Teammitglieder verfügbar sind, wenn bestimmte Teammitglieder im Urlaub oder krank sind.
- Bereitstellen zusätzlicher Workloads, ohne dass die zusätzlichen Auswirkungen auf die Teammitglieder überprüft werden, die sich um diese und andere Workloads kümmern.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie über qualifizierte Teammitglieder verfügen, können sie Ihren Workload effektiv unterstützen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Know-how der Mitarbeiter: Überprüfen Sie, ob gut geschultes Personal für den Workload vorhanden ist, das effektiv für sie eingesetzt werden kann.
  - Teamgröße: Stellen Sie sicher, dass Sie über genügend Teammitglieder verfügen, um die operativen Aktivitäten abzudecken, einschließlich der Rufbereitschaftsaktivitäten.

- Qualifikationen des Teams: Stellen Sie sicher, dass Ihre Teammitglieder die erforderlichen Schulungen zu AWS, zum Workload und zu Ihren Einsatzmitteln zur Erfüllung der zugewiesenen Aufgaben erhalten.
  - AWS-Veranstaltungen und -Webinare
  - Willkommen bei AWS Training and Certification
- Überprüfen der Kompetenzen: Überprüfen Sie die Größe und Qualifikation des Teams bei sich ändernden Betriebsbedingungen und Workloads, um sicherzustellen, dass ausreichende Fähigkeiten zur Aufrechterhaltung der operativen Leistung vorhanden sind. Nehmen Sie Anpassungen vor, um sicherzustellen, dass Teamgröße und -fähigkeit den betrieblichen Anforderungen für die vom Team unterstützten Workloads entsprechen.

Ressourcen

Zugehörige Dokumente:

- AWS-Blogs
- AWS-Veranstaltungen und -Webinare
- Erste Schritte – AWS Resource Center
- AWS Online Tech Talks
- Willkommen bei AWS Training and Certification

Zugehörige Beispiele:

- Well-Architected Labs

OPS07-BP02 Sicherstellen einer konsistenten Prüfung der betrieblichen Bereitschaft

Verwenden Sie Operational Readiness Reviews (ORRs, Überprüfungen der Einsatzbereitschaft), um zu prüfen, ob Sie Ihren Workload betreiben können. ORR ist ein bei Amazon entwickelter Mechanismus zur Prüfung, ob Teams ihre Workloads in sicherer Weise betreiben können. ORR bezeichnet einen Prüfungs- und Inspektionsprozess anhand einer Checkliste mit Anforderungen. Dies ist ein Self-Service-Vorgang, mit dem Teams ihre Workloads zertifizieren. ORRs beinhalten bewährte Methoden aus unseren jahrelangen Erfahrungen bei der Erstellung von Software.

Eine ORR-Checkliste besteht aus Architekturempfehlungen, betrieblichen Prozessen, Ereignismanagement und Freigabequalität. Unser Correction of Error (CoE)-Prozess ist dafür eine

sehr wichtige Grundlage. Ihre eigene Analyse nach einem Vorfall sollte die Weiterentwicklung Ihrer eigenen ORR unterstützen. Bei einer ORR geht es nicht nur um die Umsetzung bewährter Methoden, sondern auch darum, das erneute Auftreten von Ereignissen zu verhindern. Schließlich können auch Sicherheit, Governance und Compliance zu einer ORR gehören.

Führen Sie eine ORR durch, bevor ein Workload zur allgemeinen Verfügbarkeit gestartet wird, und anschließend während des gesamten Softwareentwicklungslebenszyklus. Die Durchführung der ORR vor dem Start verbessert Ihre Fähigkeit zum sicheren Betrieb des Workloads. Führen Sie die ORR auf dem Workload regelmäßig erneut durch, um Abweichungen von bewährten Methoden zu erkennen. Sie können ORR-Checklisten für neue Serviceeinführungen oder für regelmäßige Prüfungen haben. So bleiben Sie hinsichtlich der neuen bewährten Methoden auf dem Laufenden und können Erfahrungen aus Analysen nach Vorfällen einarbeiten. Wenn Sie mit der Cloud immer vertrauter werden, können Sie ORR-Anforderungen als Standardelemente in Ihre Architektur einbauen.

Gewünschtes Ergebnis: Sie haben eine ORR-Checkliste mit bewährten Methoden für Ihre Organisation. ORRs werden vor dem Start von Workloads durchgeführt. ORR werden im Laufe des Workloadlebenszyklus regelmäßig durchgeführt.

Typische Anti-Muster:

- Sie starten einen Workload, ohne zu wissen, ob Sie diesen betreiben können.
- Governance- und Sicherheitsanforderungen gehören nicht zur Zertifizierung eines Workloads für den Start.
- Workloads werden nicht regelmäßig erneut bewertet.
- Workloads werden gestartet, ohne dass erforderliche Verfahren eingerichtet sind.
- Sie erleben die Wiederholung von Ausfällen mit der gleichen Ursache bei mehreren Workloads.

Vorteile der Nutzung dieser bewährten Methode:

- Ihre Workloads beinhalten bewährte Methoden für Architektur, Prozess und Management.
- Erkenntnisse werden in Ihren ORR-Prozess integriert.
- Workloads werden gestartet, wenn erforderliche Verfahren eingerichtet sind.
- ORRs werden über den gesamten Softwarelebenszyklus Ihrer Workloads hinweg ausgeführt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Eine ORR ist zweierlei: ein Verfahren und eine Checkliste. Ihr ORR-Verfahren sollte von ihrer Organisation übernommen und von der Unternehmensleitung unterstützt werden. ORRs müssen mindestens durchgeführt werden, bevor Workloads zur allgemeinen Verfügbarkeit gestartet werden. Führen Sie die ORR während des gesamten Lebenszyklus der Softwareentwicklung durch, um ihn bei bewährten Methoden oder neuen Anforderungen aktuell zu halten. Die ORR-Checkliste sollte Konfigurationselemente, Sicherheits- und Governance-Elemente sowie bewährte Methoden aus Ihrer Organisation enthalten. Mit der Zeit können Sie Services wie AWS Config, AWS Security Hubund AWS Control Tower Guardrailsverwenden, um bewährte Methoden aus der ORR in den Integritätsschutz für die automatische Erkennung optimaler Verfahrensweisen aufzunehmen.

Kundenbeispiel

Nach mehreren Produktionsvorfällen entschied sich AnyCompany Retail, einen ORR-Prozess zu implementieren. Das Unternehmen erstellte eine Checkliste mit bewährten Methoden sowie Governance- und Compliance-Anforderungen und Erfahrungen aus früheren Ausfällen. Für neue Workloads werden vor dem Start ORRs durchgeführt. Für jeden Workload wird eine jährliche ORR mit einer Teilmenge der bewährten Methoden durchgeführt, um neue bewährte Methoden und Anforderungen umzusetzen, die der ORR-Checkliste hinzugefügt werden. Mit der Zeit verwendete AnyCompany Retail AWS Config zur Aufdeckung einer bewährter Methoden, was den ORR-Prozess beschleunigte.

Implementierungsschritte

Weitere Informationen zu ORRs finden Sie im Whitepaper zur Überprüfung der betrieblichen Bereitschaft (ORR). Hier finden Sie ausführliche Informationen zur Geschichte des ORR-Verfahrens, zum Aufbau Ihrer eigenen ORR-Praxis und zur Erstellung Ihrer ORR-Checkliste. Die folgenden Schritte sind eine verkürzte Version dieses Dokuments. Für ein vertieftes Verständnis des ORR-Konzepts und der Erstellung eigener ORRs empfehlen wir, das Whitepaper zu lesen.

1. Bringen Sie die wichtigsten Beteiligten zusammen, darunter auch Vertreter aus den Bereichen Sicherheit, Operations und Entwicklung.
2. Lassen Sie alle Beteiligten mindestens eine Anforderung beisteuern. Versuchen Sie für den ersten Durchgang die Anzahl der Elemente auf höchstens dreißig zu beschränken.
   - Anhang B: Beispielfragen für ORRs aus dem ORR-Whitepaper enthält Beispielfragen, die Ihnen beim Start helfen können.
3. Fassen Sie Ihre Anforderungen in einer Tabelle zusammen.

    - Sie können Fokusbereiche in AWS Well-Architected Tool verwenden, um Ihre ORR zu entwickeln und an Ihre Konten und die AWS-Organisation weiterzugeben.
4. Identifizieren Sie einen Workload für die ORR. Ideal ist dafür ein Pre-Launch-Workload oder ein interner Workload.
5. Gehen Sie die ORR-Checkliste durch und notieren Sie alle Erkenntnisse. Diese sind möglicherweise nicht OK, wenn eine Behebung stattfindet. Fügen Sie alle Erkenntnisse ohne Behebung Ihrer Liste hinzu und implementieren Sie die Behebungen vor dem Start.
6. Fügen Sie Ihrer ORR-Checkliste stets weitere bewährte Methoden und Anforderungen hinzu.

AWS Support-Kunden mit Enterprise Support können den Operational Readiness Review Workshop bei ihrem Technical Account Manager anfordern. Der Workshop ist eine interaktive „Working Backwards“- Sitzung zur Entwicklung Ihrer eigenen ORR-Checkliste.

Aufwand für den Implementierungsplan: Hoch. Die Einführung einer ORR-Praxis in Ihrer Organisation erfordert die Unterstützung durch Führungskräfte und alle Beteiligten. Erstellen und aktualisieren Sie die Checkliste mit Beiträgen aus der gesamten Organisation.

Ressourcen

Zugehörige bewährte Methoden:

- OPS01-BP03 Bewerten der Governance-Anforderungen – Governance-Anforderungen passen perfekt zu einer ORR-Checkliste
- OPS01-BP04 Bewerten der Compliance-Anforderungen – Compliance-Anforderungen werden manchmal auf ORR-Checklisten berücksichtigt. Ansonsten sind sie ein separater Prozess.
- OPS03-BP07 Teams mit entsprechenden Ressourcen ausstatten – Die Team-Kapazität ist ein guter Kandidat für eine ORR-Anforderung.
- OPS06-BP01 Einkalkulieren nicht erfolgreicher Änderungen – Vor dem Start Ihres Workloads muss ein Rollback- oder Rollforward-Plan eingerichtet werden.
- OPS07-BP01 Sicherstellen des Know-hows der Mitarbeiter – Zur Unterstützung eines Workloads benötigen Sie das erforderliche Personal.
- SEC01-BP03 Identifizieren und Validieren von Kontrollzielen – Sicherheitskontrollziele sind hervorragende ORR-Anforderungen.
- REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten – Notfallwiederherstellungspläne sind eine gute ORR-Anforderung.

- COST02-BP01 Entwickeln von Richtlinien auf Basis Ihrer Organisationsanforderungen – Kostenmanagementrichtlinien sind für Ihre ORR-Checkliste gut geeignet.

Zugehörige Dokumente:

- AWS Control Tower - Integritätsschutz in AWS Control Tower
- AWS Well-Architected Tool - Fokusbereiche
- Operational Readiness Review Template von Adrian Hornsby
- Whitepaper zur Überprüfung der betrieblichen Bereitschaft (ORR)

Zugehörige Videos:

- AWS Supports You | Building an Effective Operational Readiness Review (ORR) (AWS Supports You | Entwickeln einer effektiven Überprüfung der betrieblichen Bereitschaft (ORR))

Zugehörige Beispiele:

- Sample Operational Readiness Review (ORR)-Fokusbereich

Zugehörige Services:

- AWS Config
- AWS Control Tower
- AWS Security Hub
- AWS Well-Architected Tool

OPS07-BP03 Verwenden von Runbooks zur Durchführung von Verfahren

A Runbooks ist ein dokumentierter Prozess für das Erreichen eines bestimmten Ergebnisses. Runbooks bestehen aus einer Reihe von Schritten, die befolgt werden sollen, um ein Ergebnis zu erzielen. Runbooks werden schon seit den frühen Tagen der Luftfahrt verwendet. Im Cloud-Bereich werden Runbooks verwendet, um die Risiken zu reduzieren und die gewünschten Ergebnisse zu erzielen. In der einfachsten Form ist ein Runbook eine Checkliste für die Durchführung einer Aufgabe.

Runbooks stellen einen kritischen Teil der Ausführung Ihres Workloads dar. Vom Onboarding eines neuen Teammitglieds bis zur Bereitstellung einer Hauptversion – Runbooks stellen kodifizierte

Prozesse dar, mit denen unabhängig von der ausführenden Person konsistente Ergebnisse erzielt werden können. Runbooks sollten an einer zentralen Stelle veröffentlicht werden. Wenn sich der Prozess verändert, sollten sie aktualisiert werden; dies stellt eine zentrale Komponente des Änderungsmanagements dar. Sie sollten auch Anleitungen für Fehlerbehandlung, Tools, Berechtigungen, Ausnahmen und Eskalationen enthalten, falls ein Problem auftritt.

Wenn sich Ihre Organisation entwickelt, sollten Sie mit der Automatisierung von Runbooks beginnen. Sie sollten zunächst Runbooks automatisieren, die kurz sind und häufig verwendet werden. Verwenden Sie Skriptsprachen, um Schritte zu automatisieren oder ihre Ausführung zu vereinfachen. Nach der Automatisierung der ersten Runbooks können Sie komplexere Runbooks automatisieren. Mit der Zeit sollten die meisten Ihrer Runbooks auf die eine oder andere Art automatisiert werden.

Gewünschtes Ergebnis: Ihr Team besitzt eine Sammlung von Schritt-für-Schritt-Anleitungen für die Ausführung von Workload-Aufgaben. Die Runbooks enthalten Angaben zum gewünschten Ergebnis sowie zu notwendigen Tools und Berechtigungen. Darüber hinaus stellen sie Anleitungen für die Fehlerbehandlung bereit. Sie sind an einer zentralen Stelle gespeichert und werden häufig aktualisiert.

Typische Anti-Muster:

- Verlassen auf das Gedächtnis, um die einzelnen Schritte in einem Prozess durchzuführen.
- Manuelle Bereitstellung von Änderungen ohne Checkliste.
- Verschiedene Teammitglieder führen den gleichen Prozess aus, aber mit unterschiedlichen Schritten oder Ergebnissen.
- Runbooks sind nicht mehr mit Systemänderungen und Automatisierungen synchronisiert.

Vorteile der Nutzung dieser bewährten Methode:

- Reduzierung der Fehlerquoten für manuelle Aufgaben.
- Prozess werden konsistent ausgeführt.
- Neue Teammitglieder können schneller mit der Ausführung von Aufgaben beginnen.
- Runbooks können automatisiert werden, um den Aufwand zu reduzieren.

Risikostufe bei fehlender Befolgung dieser Best Practice: Mittel

Implementierungsleitfaden

Runbooks können verschiedene Formen annehmen, abhängig vom Entwicklungsstand Ihrer Organisation. Sie sollten mindestens aus einem Schritt-für-Schritt-Textdokument bestehen. Das gewünschte Ergebnis sollte klar angegeben werden. Dokumentieren Sie klar die notwendigen Berechtigungen oder Tools. Stellen Sie für den Fall, dass etwas nicht funktioniert, detaillierte Anleitungen für Fehlerbehandlung und Eskalation bereit. Nennen Sie die Person, die für das Runbook verantwortlich ist, und veröffentlichen Sie es an einer zentralen Stelle. Validieren Sie das Runbook, nachdem Sie es dokumentiert haben, indem Sie es von einem Teammitglied ausführen lassen. Mit der weiteren Entwicklung der Verfahren sollten Sie Ihre Runbooks entsprechend Ihrem Prozess für das Änderungsmanagement aktualisieren.

Ihre textbasierten Runbooks sollten mit zunehmender Entwicklung Ihrer Organisation automatisiert werden. Mit Services wie AWS Systems Manager Automationkönnen Sie Textdateien zu Automatisierungen transformieren, die Sie für Ihren Workload ausführen können. Diese Automatisierungen können als Reaktion auf Ereignisse ausgeführt werden, was den operativen Aufwand für die Wartung des Workloads reduziert.

Kundenbeispiel

AnyCompany Retail muss während Softwarebereitstellungen die Datenbankschemata aktualisieren. Das Cloud Operations-Team entwickelt gemeinsam mit dem Datenbankverwaltungsteam ein Runbook für die manuelle Bereitstellung dieser Änderungen. In diesem Runbook werden die einzelnen Prozessschritte in Form einer Checkliste aufgelistet. Es enthält für den Fall, dass es ein Problem gibt, auch einen Abschnitt zur Fehlerbehandlung. Das Runbook wird wie die übrigen Runbooks im internen Wiki veröffentlicht. Das Cloud Operations-Team plant, das Runbook in der Zukunft zu automatisieren.

Implementierungsschritte

Wenn Sie noch kein Dokumenten-Repository besitzen, dann ist ein Repository für die Versionskontrolle hervorragend als Grundlage für Ihre Runbook-Bibliothek geeignet. Sie können Ihre Runbooks mithilfe von Markdown erstellen. Wir haben eine Runbook-Beispielvorlage bereitgestellt, die Sie für die Erstellung von Runbooks verwenden können.

```
# Runbook-Titel ## Runbook-Informationen | Runbook-ID | Beschreibung | Verwendete Tools
 | Spezielle Berechtigungen | Runbook-Autor | Letzte Aktualisierung | Eskalations-POC |
 |-------|-------|-------|-------|-------|-------|-------| | RUN001 | Wofür ist dieses
 Runbook bestimmt? Was ist das gewünschte Ergebnis? | Tools | Berechtigungen| Ihr Name
 | 2022-09-21 | Eskalationsname | ## Schritte 1. Schritt eins 2. Schritt zwei
```

1. Wenn Sie noch kein Dokumentations-Repository oder -Wiki besitzen, sollten Sie in Ihrem Versionskontrollsystem ein neues Versionskontroll-Repository erstellen.
2. Identifizieren Sie einen Prozess, für den es kein Runbook gibt. Ein idealer Prozess hierfür ist ein Prozess, der halbregelmäßig ausgeführt wird, nur wenige Schritte enthält und bei Fehlern nur geringe Auswirkungen hat.
3. Erstellen Sie in Ihrem Dokument-Repository ein neues Markdown-Entwurfsdokument auf der Basis der Vorlage. Geben Sie den `Runbook-Titel` ein und füllen Sie die erforderlichen Felder unter `Runbook-Informationen`aus.
4. Füllen Sie beginnend mit dem ersten Schritt den Abschnitt `Schritte` im Runbook aus.
5. Geben Sie das Runbook einem Teammitglied. Lassen Sie das Teammitglied das Runbook ausführen, um die Schritte zu validieren. Aktualisieren Sie das Runbook, wenn etwas fehlt oder unklar ist.
6. Veröffentlichen Sie das Runbook in Ihrem internen Dokumentationsspeicher. Informieren Sie Ihr Team und die übrigen Stakeholder über das Runbook, nachdem es veröffentlicht wurde.
7. Mit der Zeit werden Sie eine Bibliothek von Runbooks aufbauen. Beginnen Sie mit der Automatisierung von Runbooks, wenn diese Bibliothek wächst.

Aufwand für den Implementierungsplan: Niedrig. Eine Schritt-für-Schritt-Anleitung in Textform ist der Mindeststandard für ein Runbook. Die Automatisierung von Runbooks kann den Implementierungsaufwand erhöhen.

Ressourcen

Zugehörige bewährte Methoden:

- OPS02-BP02 Prozesse und Verfahren haben feste Besitzer: Es sollte eine verantwortliche Person für jedes Runbook geben, die das jeweilige Runbook verwaltet und aktualisiert.
- OPS07-BP04 Verwenden von Playbooks zum Untersuchen von Problemen: Runbooks und Playbooks sind sich zwar ähnlich, es gibt jedoch einen wichtigen Unterschied: Ein Runbook hat ein gewünschtes Ergebnis. Häufig werden Runbooks ausgelöst, wenn ein Playbook die Ursache für ein Problem identifiziert hat.
- OPS10-BP01 Verwenden eines Prozesses für die Bewältigung von Ereignissen, Vorfällen und Problemen: Runbooks sind Bestandteil guter Verfahren für die Verwaltung von Ereignissen, Vorfällen und Problemen.

- OPS10-BP02 Implementieren eines Prozesses für jeden Alarm: Runbooks und Playbooks sollten verwendet werden, um auf Warnungen zu reagieren. Mit der Zeit sollten diese Reaktionen automatisiert werden.
- OPS11-BP04 Wissensmanagement: Die Verwaltung und Aktualisierung ist ein wesentlicher Bestandteil des Wissensmanagement.

Zugehörige Dokumente:

- Operative Kompetenz durch automatisierte Playbooks und Runbooks
- AWS Systems Manager: Mit Runbooks arbeiten
- Migrations-Playbook für große AWS-Migrationen – Aufgabe 4: Verbesserung Ihrer Migrations-Runbooks
- Verwendung von AWS Systems Manager Automation-Runbooks zur Lösung operativer Aufgaben

Zugehörige Videos:

- AWS re:Invent 2019: DIY guide to runbooks, incident reports, and incident response (SEC318-R1)
- Automatisierung von IT-Abläufen in AWS | Amazon Web Services
- Integration von Skripts in AWS Systems Manager

Zugehörige Beispiele:

- AWS Systems Manager: Automation-Walkthroughs
- AWS Systems Manager: Runbook für die Wiederherstellung eines Root-Volumes anhand des letzten Snapshots
- Entwicklung eines Runbooks für Vorfallreaktionen in AWS mit Jupyter Notebooks und CloudTrail Lake
- Gitlab – Runbooks
- Rubix – eine Python-Bibliothek für die Erstellung von Runbooks in Jupyter Notebooks
- Verwendung von Document Builder für die Erstellung angepasster Runbooks
- Well-Architected Labs: Automatisieren von Vorgängen mit Playbooks und Runbooks

Zugehörige Services:

- AWS Systems Manager Automation

OPS07-BP04 Verwenden von Playbooks zum Untersuchen von Problemen

Playbooks sind Schritt-für-Schritt-Anleitungen zur Untersuchung von Vorfällen. Wenn Vorfälle auftreten, werden Playbooks verwendet, um sie zu untersuchen, die Auswirkungen abzuschätzen und Ursachen zu identifizieren. Playbooks werden für verschiedene Szenarien eingesetzt, von fehlgeschlagenen Bereitstellungen bis hin zu Sicherheitsvorfällen. In vielen Fällen identifizieren Playbooks Ursachen, die dann mithilfe eines Runbooks beseitigt werden. Playbooks sind eine sehr wichtige Komponente der Vorfallreaktionspläne Ihrer Organisation.

Ein gutes Playbook weist einige zentrale Merkmale auf. Es leitet den Nutzer Schritt für Schritt durch den Erkennungsprozess. Welche Schritte sollten befolgt werden, um einen Vorfall zu diagnostizieren? Legen Sie im Playbook klar fest, ob bestimmte Tools oder erhöhte Berechtigungen benötigt werden. Ein wichtiger Teil ist ein Kommunikationsplan, um alle Beteiligten über den Status der Untersuchung zu informieren. Für den Fall, dass die eigentliche Ursache des Vorfalls nicht identifiziert werden kann, sollte das Playbook einen Eskalationsplan enthalten. Wenn die Ursache identifiziert wurde, sollte das Playbook auf ein Runbook verweisen, das beschreibt, wie die Ursache zu beheben ist. Playbooks sollten zentral gespeichert und regelmäßig gepflegt werden. Wenn Playbooks für bestimmte Warnungen verwendet werden, sollte Ihr Team in den Warnungen auf das Playbook verwiesen werden.

Im Zuge der Weiterentwicklung Ihrer Organisation sollten Sie Ihre Playbooks automatisieren. Beginnen Sie mit Playbooks für Vorfälle mit geringem Risikograd. Automatisieren Sie die Erkennungsschritte mit Skripts. Stellen Sie sicher, dass Sie über begleitende Runbooks für die Behebung typischer Ursachen verfügen.

Gewünschtes Ergebnis: Ihre Organisation verfügt über Playbooks für typische Vorfälle. Die Playbooks werden an einem zentralen Ort gespeichert und sind für Ihre Teammitglieder verfügbar. Playbooks werden häufig aktualisiert. Für alle bekannten Ursachen werden begleitende Runbooks erstellt.

Typische Anti-Muster:

- Es gibt kein Standardverfahren für die Untersuchung von Vorfällen.
- Teammitglieder verlassen sich auf ihr Gedächtnis oder allgemein vorhandenes Wissen, um eine fehlgeschlagene Bereitstellung zu beheben.
- Neue Teammitglieder lernen die Untersuchung von Problemen durch Ausprobieren.

- Es werden keine bewährten Methoden für die Untersuchung von Problemen zwischen Teams ausgetauscht.

Vorteile der Nutzung dieser bewährten Methode:

- Playbooks verbessern Ihre Fähigkeit zum Umgang mit Vorfällen.
- Verschiedene Teammitglieder können dasselbe Playbook verwenden, um Ursachen in konsistenter Weise zu ermitteln.
- Für bekannte Ursachen können Runbooks entwickelt werden, um die Wiederherstellungszeit zu verkürzen.
- Mit Playbooks können Teammitglieder schneller Beiträge leisten.
- Mit wiederholbaren Playbooks können Teams ihre Prozesse skalieren.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Mittel

Implementierungsleitfaden

Wie Sie Ihre Playbooks aufbauen und verwenden, hängt vom Reifegrad Ihrer Organisation ab. Wenn Sie noch neu in der Cloud sind, erstellen Sie Playbooks in Textform in einem zentralen Dokumenten-Repository. Wenn sich Ihre Organisation weiterentwickelt, können Playbooks mit Skriptsprachen wie Python teilweise automatisiert werden. Diese Skripts können zur Beschleunigung der Untersuchung in einem Jupyter Notebook ausgeführt werden. Fortgeschrittene Organisationen haben vollständig automatisierte Playbooks für häufig auftretende Probleme, die dann mit Runbooks automatisch behoben werden.

Beginnen Sie die Arbeit an Ihren Playbooks mit der Auflistung typischer Vorfälle bei Ihren Workloads. Wählen Sie Playbooks zunächst für Vorfälle mit geringem Risiko, bei denen die Ursache eingegrenzt werden kann. Wenn Sie über Playbooks für einfachere Szenarien verfügen, gehen Sie zu Szenarien mit höheren Risiken oder zu Szenarien über, bei denen die Ursache nicht vollständig klar ist.

Ihre textbasierten Runbooks sollten mit zunehmender Entwicklung Ihrer Organisation automatisiert werden. Mit Services wie AWS Systems Manager Automationskann einfacher Text in Automatisierungen umgewandelt werden. Diese Automatisierungen können dann für Ihren Workload ausgeführt werden, um die Untersuchungen zu beschleunigen. Sie können in Reaktion auf Ereignisse aktiviert werden, wodurch sich der durchschnittliche Zeitaufwand für die Untersuchung und Behebung von Vorfällen reduziert.

Kunden können AWS Systems Manager Incident Manager zur Reaktion auf Vorfälle verwenden. Dieser Service bietet eine einzige Oberfläche für die Untersuchung von Vorfällen, die Information der Beteiligten über Untersuchung und Abhilfemaßnahmen und die Zusammenarbeit während des gesamten Vorgangs. Er verwendet AWS Systems Manager Automations zur Beschleunigung von Untersuchung und Wiederherstellung.

Kundenbeispiel

Ein Produktionsvorfall hat Auswirkungen auf AnyCompany Retail. Der zuständige Techniker untersuchte das Problem mithilfe eines Playbooks. Im Zuge der einzelnen Schritte wurden anhand des aktuellen Playbooks die Beteiligten identifiziert. Der Techniker ermittelte einen Race-Zustand in einem Backend-Service als Ursache für den Vorfall. Mithilfe eines Runbooks startete er den Service neu und brachte AnyCompany Retail so wieder online.

Implementierungsschritte

Wenn Sie noch kein Dokumenten-Repository besitzen, dann sollten Sie ein Versionskontroll-Repository für Ihre Runbook-Bibliothek erstellen. Sie können Ihre Playbooks mit Markdown erstellen, das mit den meisten Playbook-Automatisierungssystemen kompatibel ist. Wenn Sie neu beginnen, verwenden Sie die folgende Beispielvorlage für ein Playbook.

```
# Playbook-Titel ## Playbook-Info | Playbook-ID | Beschreibung |
 Verwendete Tools | Besondere Berechtigungen | Playbook-Autor | Letzte
 Aktualisierung | Eskalation-POC | Beteiligte | Kommunikationsplan |
 |-------|-------|-------|-------|-------|-------|-------|-------|-------| | RUN001
 | Wofür ist dieses Playbook? Für welchen Vorfall wird es verwendet? | Tools |
 Berechtigungen | Ihr Name | 21.09.2022 | Eskalationsname | Name des Beteiligten | Wie
 werden während der Untersuchung Aktualisierungen mitgeteilt? | ## Schritte 1. Schritt
 eins 2. Schritt zwei
```

1. Wenn Sie noch kein Dokumenten-Repository oder -Wiki besitzen, sollten Sie in Ihrem Versionskontrollsystem ein neues Versionskontroll-Repository für Ihre Playbooks erstellen.
2. Identifizieren Sie ein typisches Problem, das eine Untersuchung erfordert. Dies sollte ein Szenario sein, bei dem die Ursache auf wenige Probleme eingegrenzt werden kann und das Risiko insgesamt niedrig ist.
3. Füllen Sie anhand der Markdown-Vorlage den Abschnitt `Name des Playbooks` und die Felder unter `Playbook-Info`aus.
4. Geben Sie die Schritte zur Fehlerbehebung ein. Benennen Sie die zu treffenden Maßnahmen bzw. die zu untersuchenden Bereiche so klar wie möglich.

5. Geben Sie das Playbook einem Teammitglied zur Prüfung. Wenn darin etwas fehlt oder nicht klar ist, aktualisieren Sie das Playbook.
6. Veröffentlichen Sie Ihr Playbook in Ihrem Dokumenten-Repository und informieren Sie Ihr Team und alle Beteiligten darüber.
7. Diese Playbook-Bibliothek wächst mit der Zeit an. Sobald Sie mehrere Playbooks haben, beginnen Sie mithilfe von Tools wie AWS Systems Manager Automations mit ihrer Automatisierung.

Aufwand für den Implementierungsplan: Niedrig. Ihre Playbooks sollten an einem zentralen Ort gespeicherte Textdokumente sein. Ausgereiftere Organisationen gehen zu automatisierten Playbooks über.

Ressourcen

Zugehörige bewährte Methoden:

- OPS02-BP02 Prozesse und Verfahren haben feste Besitzer: Es sollte eine verantwortliche Person für jedes Runbook geben, die das jeweilige Runbook verwaltet und aktualisiert.
- OPS07-BP03 Verwenden von Runbooks zur Durchführung von Verfahren: Runbooks und Playbooks sind sich zwar ähnlich, es gibt jedoch einen wichtigen Unterschied: Ein Runbook hat ein gewünschtes Ergebnis. Häufig werden Runbooks verwendet, wenn ein Playbook die Ursache für ein Problem identifiziert hat.
- OPS10-BP01 Verwenden eines Prozesses für die Bewältigung von Ereignissen, Vorfällen und Problemen: Runbooks sind Bestandteil guter Verfahren für die Verwaltung von Ereignissen, Vorfällen und Problemen.
- OPS10-BP02 Implementieren eines Prozesses für jeden Alarm: Runbooks und Playbooks sollten verwendet werden, um auf Warnungen zu reagieren. Mit der Zeit sollten diese Reaktionen automatisiert werden.
- OPS11-BP04 Wissensmanagement: Die Verwaltung und Aktualisierung ist ein wesentlicher Bestandteil des Wissensmanagements.

Zugehörige Dokumente:

- Operative Kompetenz durch automatisierte Playbooks und Runbooks
- AWS Systems Manager: Mit Runbooks arbeiten
- Verwendung von AWS Systems Manager-Automation-Runbooks zur Lösung operativer Aufgaben

Zugehörige Videos:

- AWS re:Invent 2019: DIY guide to runbooks, incident reports, and incident response (SEC318-R1) (AWS re:Invent 2019: DIY-Leitfaden für Runbooks, Vorfallberichte und Vorfallreaktion (SEC318-R1))
- AWS Systems Manager Incident Manager - AWS Virtual Workshops (AWS Systems Manager Incident Manager – virtuelle AWS-Workshops)
- Integrate Scripts into AWS Systems Manager (Integration von Skripts in AWS Systems Manager)

Zugehörige Beispiele:

- AWS Customer Playbook Framework
- AWS Systems Manager: Walkthroughs zur Automatisierung
- Entwicklung eines Runbooks für Vorfallreaktionen in AWS mit Jupyter Notebooks und CloudTrail Lake
- Rubix – Eine Python-Bibliothek für die Erstellung von Runbooks in Jupyter Notebooks
- Verwendung von Document Builder für die Erstellung angepasster Runbooks
- Well-Architected Labs: Automatisieren von Vorgängen mit Playbooks und Runbooks
- Well-Architected Labs: Playbook für Vorfallreaktion mit Jupyter

Zugehörige Services:

- AWS Systems Manager-Automatisierung
- AWS Systems Manager Incident Manager

OPS07-BP05 Treffen fundierter Entscheidungen für die Bereitstellung von Systemen und Änderungen

Bewerten Sie die Fähigkeiten des Teams zur Unterstützung des Workloads und die Einhaltung der Governance durch den Workload. Wägen Sie diese Aspekte gegen die Vorteile der Bereitstellung ab, wenn Sie vor der Entscheidung stehen, ob Sie ein System umstellen oder eine Änderung in der Produktion vornehmen sollten. Beschäftigen Sie sich eingehend mit den Vorteilen und Risiken, damit Sie fundierte Entscheidungen treffen können.

Eine „Pre-mortem“-Übung ist eine Übung, bei der ein Team einen Fehler simuliert, um Strategien zur Behebung zu entwickeln. Beugen Sie Fehlern nach Möglichkeit vor und stellen Sie entsprechende

Abläufe auf. Wenn Sie an Checklisten, mit denen Sie Ihre Workloads beurteilen, Änderungen vornehmen, bedenken Sie auch, was mit live geschalteten Systemen geschehen soll, die mit den Änderungen nicht mehr kompatibel sind.

Gängige Antimuster:

- Die Entscheidung, einen Workload bereitzustellen, ohne die Sicherheitsrisiken durch den Workload zu verstehen.
- Die Entscheidung, einen Workload bereitzustellen, ohne zu wissen, ob er Ihre Governance und Ihre Standards erfüllt.
- Die Entscheidung, einen Workload bereitzustellen, ohne zu wissen, ob Ihr Team damit fertig wird.
- Die Entscheidung, einen Workload bereitzustellen, ohne zu verstehen, wie er dem Unternehmen nützt.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie über qualifizierte Teammitglieder verfügen, können sie Ihren Workload effektiv unterstützen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Fundierte Entscheidungen zur Bereitstellung von Workloads und Änderungen treffen: Bewerten Sie die Fähigkeiten des Teams, um den Workload und die Compliance des Workloads mit Governance zu unterstützen. Wägen Sie diese Aspekte gegen die Vorteile der Bereitstellung ab, wenn Sie vor der Entscheidung stehen, ob Sie ein System umstellen oder eine Änderung in der Produktion vornehmen sollten. Beschäftigen Sie sich eingehend mit den Vorteilen und Risiken, damit Sie fundierte Entscheidungen treffen können.

Betrieb

Fragen

- OPS 8 Wie können Sie den Zustand Ihres Workloads beurteilen?
- OPS 9 Wie können Sie den Zustand Ihrer Operationen beurteilen?
- OPS 10 Wie bewältigen Sie Workload- und operationsspezifische Ereignisse?

OPS 8 Wie können Sie den Zustand Ihres Workloads beurteilen?

Definieren, erfassen und analysieren Sie Workload-Metriken, um einen Einblick in Workload-Ereignisse zu erhalten. Dies ist wichtig, damit Sie bei Bedarf entsprechende Maßnahmen ergreifen können.

Bewährte Methoden

- OPS08-BP01 Ermitteln wichtiger Leistungskennzahlen
- OPS08-BP02 Definieren von Workload-Metriken
- OPS08-BP03 Erfassen und Analysieren von Workload-Metriken
- OPS08-BP04 Festlegen von Ausgangswerten für Workload-Metriken
- OPS08-BP05 Lernen erwarteter Aktivitätsmuster für den Workload
- OPS08-BP06 Alarm bei gefährdeten Workload-Ergebnissen
- OPS08-BP07 Alarm bei festgestellten Workload-Anomalien
- OPS08-BP08 Prüfen der Erreichung von angestrebten Ergebnissen und der Wirksamkeit von KPIs und Metriken

OPS08-BP01 Ermitteln wichtiger Leistungskennzahlen

Identifizieren Sie wichtige Leistungskennzahlen (KPIs) anhand der gewünschten Geschäftsergebnisse (z. B. Auftragsrate, Kundenbindungsrate und Gewinn im Vergleich zu Betriebsausgaben) und Kundenergebnisse (z. B. Kundenzufriedenheit). Bewerten Sie zur Messung des Workload-Erfolgs KPIs.

Gängige Antimuster:

- Sie werden von der Geschäftsleitung gefragt, wie erfolgreich ein Workload die Geschäftsanforderungen erfüllt, haben aber keinen Referenzrahmen, um den Erfolg zu bestimmen.
- Sie können nicht feststellen, ob die kommerzielle Standardanwendung, die Sie für Ihr Unternehmen betreiben, kostengünstig ist.

Vorteile der Einführung dieser bewährten Methode: Durch die Ermittlung wichtiger Leistungskennzahlen ermöglichen Sie das Erreichen von Geschäftsergebnissen als Test des Workload-Zustands und -Erfolgs.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Ermitteln wichtiger Leistungskennzahlen: Ermitteln Sie auf Basis der gewünschten geschäftlichen und kundenspezifischen Ergebnisse wichtige Leistungskennzahlen (Key Performance Indicators, KPIs). Bewerten Sie zur Messung des Workload-Erfolgs KPIs.

OPS08-BP02 Definieren von Workload-Metriken

Definieren Sie Workload-Metriken, um den Erfolg von KPIs zu messen (z. B. abgebrochene Einkaufsvorgänge, getätigte Bestellungen, Kosten, Preis und zugewiesene Workload-Ausgaben). Definieren Sie Workload-Metriken, um den Zustand des Workloads zu messen (z. B. Schnittstellenreaktionszeit, Fehlerrate, Anfragen, abgeschlossene Anfragen und Auslastung). Bewerten Sie Metriken, um festzustellen, ob der Workload die gewünschten Ergebnisse erzielt, und um den Zustand des Workloads zu beurteilen.

Sie sollten Protokolldaten an einen Service wie CloudWatch Logs senden und Metriken aus Beobachtungen der erforderlichen Protokollinhalte generieren.

CloudWatch verfügt über spezielle Funktionen wie Amazon CloudWatch Insights für .NET und SQL Server und Container Insights, die Sie bei der Identifizierung und Einrichtung von Schlüsselmetriken, Protokollen und Alarmen für Ihre speziell unterstützten Anwendungsressourcen und Technologiestapel unterstützen können.

Gängige Antimuster:

- Sie haben „Standard“-Metriken definiert, die nicht mit KPIs verknüpft oder auf Workloads zugeschnitten sind.
- In Ihren Metrikberechnungen liegen Fehler vor, die zu ungültigen Ergebnissen führen.
- Sie haben keine Metriken für Ihren Workload definiert.
- Sie messen nur die Verfügbarkeit.

Vorteile der Einführung dieser bewährten Praxis: Durch das Definieren und Auswerten von Workload-Metriken können Sie den Zustand Ihrer Workload bestimmen und den Fortschritt beim Erreichen der Geschäftsergebnisse messen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Definieren von Workload-Metriken: Definieren Sie Workload-Metriken für die Analyse der Erfüllung von KPIs. Definieren Sie Workload-Metriken für die Analyse des Zustands des Workloads und dessen einzelnen Komponenten. Bewerten Sie Metriken, um festzustellen, ob der Workload die gewünschten Ergebnisse erzielt, und um den Zustand des Workloads zu beurteilen.
  - Veröffentlichen von benutzerdefinierten Metriken
  - Suchen und Filtern von Protokolldaten
  - Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch

Ressourcen

Verbundene Dokumente:

- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- Veröffentlichen von benutzerdefinierten Metriken
- Suchen und Filtern von Protokolldaten

OPS08-BP03 Erfassen und Analysieren von Workload-Metriken

Unterziehen Sie die Metriken regelmäßigen proaktiven Überprüfungen, um Trends zu ermitteln und festzustellen, wo gegebenenfalls Maßnahmen ergriffen werden müssen.

Sie sollten Protokolldaten aus Ihrer Anwendung, Workload-Komponenten, Services und API-Aufrufen in einem Service wie CloudWatch Logs zusammenfassen. Generieren Sie Metriken aus Beobachtungen der erforderlichen Protokollinhalte, um Einblicke in die Leistung von Betriebsaktivitäten zu erhalten.

In AWS können Sie Workload-Metriken analysieren und betriebliche Probleme mithilfe der Machine-Learning-Funktionen von Amazon DevOps Guruidentifizieren. AWS DevOps Guru sendet Benachrichtigungen über betriebliche Probleme mit zielgerichteten und proaktiven Empfehlungen, um Probleme zu beheben und den Anwendungszustand beizubehalten.

Aufgrund der aufgeteilten Verantwortungen in AWS werden Teile der Überwachung über das AWS Health Dashboardan Sie übermittelt. Dieses Dashboard stellt Warnungen und Informationen zur Behebung von möglicherweise problematischen AWS-Ereignissen bereit. Kunden mit Business- und Enterprise Support-Abonnements erhalten auch Zugriff auf die AWS Health-API,was die Integration in deren Ereignisverwaltungssysteme ermöglicht.

In AWS können Sie Ihre Protokolldaten zu Amazon S3 exportieren oder Protokolle zur langfristigen Speicherung direkt für langfristige Speicherung an Amazon S3 senden. Mit AWS Gluekönnen Sie Ihre Protokolldaten in Amazon S3 zur Analyse erkunden und vorbereiten und die zugehörigen Metadaten im AWSAWS Glue Data Catalog. Amazon Athenakann dann durch eine native Integration mit AWS Glue zum Analysieren Ihrer Protokolldaten und für Abfragen mit Standard-SQL verwendet werden. Mit einem Business Intelligence-Tool wie Amazon QuickSight können Sie Ihre Daten visualisieren, untersuchen und analysieren.

Eine alternative Lösung wäre die Verwendung von Amazon OpenSearch Service und OpenSearch-Dashboards zum Erfassen, Analysieren und Anzeigen von Protokollen in AWS über mehrere Konten und AWS-Regionen hinweg.

Gängige Antimuster:

- Sie werden vom Netzwerkdesign-Team nach den aktuellen Auslastungsraten der Netzwerkbandbreite gefragt. Sie geben die aktuellen Metriken an, denen zufolge die Netzwerkauslastung bei 35 % liegt. Das Team reduziert die Netzkapazität, um Kosten zu sparen. Dies führt zu weitreichenden Verbindungsproblemen, da bei Ihrer zeitpunktbezogenen Messung keine Auslastungsraten-Trends berücksichtigt wurden.
- Ihr Router ist ausgefallen. Bis zum vollständigen Ausfall protokollierte das Gerät mit immer größerer Häufigkeit nicht kritische Speicherfehler. Sie haben diesen Trend nicht erkannt und den fehlerhaften Speicher deshalb nicht ausgetauscht, sodass der Router eine Serviceunterbrechung verursachen konnte.

Vorteile der Einführung dieser bewährten Methode: Durch das Sammeln und Analysieren Ihrer Workload-Metriken gewinnen Sie einen Überblick über den Zustand Ihres Workloads und erhalten Einblicke in Trends, die sich auf Ihren Workload oder Ihre Geschäftsergebnisse auswirken können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Erfassen und Analysieren von Workload-Metriken: Unterziehen Sie die Metriken regelmäßigen proaktiven Überprüfungen, um Trends zu ermitteln und festzustellen, wo gegebenenfalls Maßnahmen ergriffen werden müssen.
  - Verwenden von Amazon CloudWatch-Metriken
  - Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch

- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent

Ressourcen

Zugehörige Dokumente:

- Amazon Athena
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- Amazon DevOps Guru
- AWS Glue
- AWSAWS Glue Data Catalog
- Amazon OpenSearch Service
- AWS Health Dashboard
- Amazon QuickSight
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent
- Verwenden von Amazon CloudWatch-Metriken

OPS08-BP04 Festlegen von Ausgangswerten für Workload-Metriken

Legen Sie Ausgangswerte für Metriken fest, um erwartete Werte als Grundlage für den Vergleich und die Ermittlung von Komponenten mit unter- oder überdurchschnittlicher Leistung bereitzustellen. Bestimmen Sie Schwellenwerte für Verbesserung, Untersuchung und Intervention.

Gängige Antimuster:

- Ein Server wird mit einer CPU-Auslastung von 95 % ausgeführt. Sie werden gefragt, ob das gut oder schlecht ist. Für die CPU-Auslastung auf diesem Server wurden keine Ausgangswerte festgelegt, sodass sie diese Frage nicht beantworten können.

Vorteile der Einführung dieser bewährten Praxis: Durch die Definition von Metrikausgangswerten können Sie aktuelle Metrikwerte und Metriktrends auswerten, um festzustellen, ob Maßnahmen erforderlich sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Festlegen von Ausgangswerten für Workload-Metriken: Legen Sie Ausgangswerte für Workload-Metriken fest, um erwartete Werte als Vergleichsgrundlage bereitzustellen.
  - Erstellen von Amazon CloudWatch-Alarmen

Ressourcen

Verbundene Dokumente:

- Erstellen von Amazon CloudWatch-Alarmen

OPS08-BP05 Lernen erwarteter Aktivitätsmuster für den Workload

Zeichnen Sie Workload-Aktivitätsmuster auf, um außergewöhnliches Verhalten zu identifizieren, damit Sie bei Bedarf entsprechend reagieren können.

CloudWatch durch die Funktion CloudWatch Anomaly Detection wendet statistische und Machine Learning-Algorithmen an, um eine Reihe von erwarteten Werten zu generieren, die ein normales Metrikverhalten darstellen.

Amazon DevOps Guru kann verwendet werden, um außergewöhnliches Verhalten über die Korrelation von Ereignissen, Protokollanalysen und die Anwendung von Machine Learning zu identifizieren und Ihre Workload-Telemetrie zu analysieren. Wird unerwartetes Verhalten erkannt, erhalten die zugehörigen Metriken und Ereignisse Empfehlungen, um das Verhalten anzugehen.

Gängige Antimuster:

- Sie prüfen Netzwerkauslastungsprotokolle und stellen fest, dass die Netzwerkauslastung zwischen 11.30 und 13.30 Uhr und dann erneut zwischen 16.30 und 18.00 Uhr gestiegen ist. Sie wissen nicht, ob diese Werte als normal betrachtet werden können.
- Ihre Webserver werden jede Nacht um 3.00 Uhr neu gestartet. Sie wissen nicht, ob dies erwartetes Verhalten ist.

Vorteile der Einführung dieser bewährten Methode: Durch das Aufzeichnen von Verhaltensmustern können Sie unerwartetes Verhalten erkennen und bei Bedarf Maßnahmen ergreifen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mehr über erwartete Aktivitätsmuster für Workload erfahren: Legen Sie Muster für die Workload-Aktivität fest, um festzustellen, wann das Verhalten von den erwarteten Werten abweicht, so dass Sie bei Bedarf angemessen reagieren können.

Ressourcen

Zugehörige Dokumente:

- Amazon DevOps Guru
- Funktion CloudWatch Anomaly Detection

OPS08-BP06 Alarm bei gefährdeten Workload-Ergebnissen

Lösen Sie einen Alarm aus, wenn die Workload-Ergebnisse gefährdet sind, damit Sie bei Bedarf angemessen reagieren können.

Idealerweise haben Sie zuvor einen Metrikschwellenwert identifiziert, bei dem Sie Alarme senden können, oder ein Ereignis, das Sie verwenden können, um eine automatisierte Antwort auszulösen.

In AWS können Sie Amazon CloudWatch Synthetics verwenden, um Canary-Skripts zur Überwachung Ihrer Endpunkte und APIs zu erstellen, indem Sie dieselben Aktionen ausführen wie Ihre Kunden. Durch die generierte Telemetrie und die erhaltenen Einblicke können Sie Probleme identifizieren, bevor die Kunden davon betroffen sind.

Sie können CloudWatch Logs Insights verwenden, um Ihre Protokolldaten mithilfe einer speziell entwickelten Abfragesprache interaktiv zu durchsuchen und zu analysieren. CloudWatch Logs Insights entdeckt automatisch Felder in Protokollen von AWS-Services und benutzerdefinierte Protokollereignisse in JSON. Es skaliert mit Ihrem Protokollvolumen und der Komplexität Ihrer Abfrage und gibt Ihnen innerhalb von Sekunden Antworten, sodass Sie nach den beitragenden Faktoren eines Vorfalls suchen können.

Gängige Antimuster:

- Sie haben keine Netzwerkkonnektivität. Niemand weiß es. Niemand versucht die Ursache zu ermitteln oder ergreift Maßnahmen, um die Konnektivität wiederherzustellen.

- Nach einem Patch sind Ihre persistenten Instances nicht mehr verfügbar und sorgen für Unterbrechungen bei den Benutzern. Ihre Benutzer haben Supportanfragen gestellt. Niemand wurde benachrichtigt. Niemand ergreift Maßnahmen.

Vorteile der Einführung dieser bewährten Methode: Indem Sie feststellen, dass Geschäftsergebnisse gefährdet sind, und mit einem Alarm auf erforderliche Maßnahmen hinweisen, können Sie die Auswirkungen eines Vorfalls verhindern oder mindern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Alarm bei gefährdeten Workload-Ergebnissen auslösen: Lösen Sie einen Alarm aus, wenn Workload-Ergebnisse gefährdet sind, damit Sie bei Bedarf entsprechend reagieren können.
  - Was ist Amazon CloudWatch Events?
  - Erstellen von Amazon CloudWatch-Alarmen
  - Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch Synthetics
- CloudWatch Logs Insights
- Erstellen von Amazon CloudWatch-Alarmen
- Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen
- Was ist Amazon CloudWatch Events?

OPS08-BP07 Alarm bei festgestellten Workload-Anomalien

Lösen Sie einen Alarm aus, wenn Workload-Anomalien festgestellt werden, damit Sie bei Bedarf angemessen reagieren können.

Ihre Analyse Ihrer Workload-Metriken im Laufe der Zeit kann Verhaltensmuster bestimmen, die Sie ausreichend quantifizieren können, um ein Ereignis zu definieren oder als Reaktion einen Alarm auszulösen.

Nach der Schulung kann die Funktion Funktion CloudWatch Anomaly Detection verwendet werden, um bei erkannten Anomalien einen Alarm auszulösen oder überlagerte erwartete Werte in einem Diagramm mit Metrikdaten für einen laufenden Vergleich bereitzustellen.

Gängige Antimuster:

- Der Umsatz über Ihre Einzelhandelswebsite ist plötzlich und drastisch angestiegen. Niemand weiß es. Niemand versucht herauszufinden, was zu diesem Anstieg geführt hat. Niemand ergreift Maßnahmen, um angesichts der zusätzlichen Last ein hochwertiges Kundenerlebnis sicherzustellen.
- Nach der Anwendung eines Patches führen Ihre persistenten Server häufige Neustarts durch, was zu Unterbrechungen für die Benutzer führt. Ihre Server werden in der Regel bis zu drei Mal neu gestartet. Niemand weiß es. Niemand versucht, der Sache auf den Grund zu gehen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie mit Workload-Verhaltensmustern vertraut sind, können Sie unerwartetes Verhalten identifizieren und bei Bedarf Maßnahmen ergreifen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Alarm bei festgestellten Workload-Anomalien auslösen: Lösen Sie einen Alarm aus, wenn Workload-Anomalien erkannt werden, damit Sie bei Bedarf entsprechend reagieren können.
  - Was ist Amazon CloudWatch Events?
  - Erstellen von Amazon CloudWatch-Alarmen
  - Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen

Ressourcen

Zugehörige Dokumente:

- Erstellen von Amazon CloudWatch-Alarmen
- Funktion CloudWatch Anomaly Detection
- Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen
- Was ist Amazon CloudWatch Events?

OPS08-BP08 Prüfen der Erreichung von angestrebten Ergebnissen und der Wirksamkeit von KPIs und Metriken

Erstellen Sie eine Ansicht Ihrer Workload-Operationen auf Geschäftsebene, mit der Sie schnell feststellen können, ob Sie die Anforderungen erfüllen, und welche Bereiche verbessert werden müssen, um die Geschäftsziele zu erreichen. Prüfen Sie die Wirksamkeit von KPIs und Metriken und überarbeiten Sie diese gegebenenfalls.

AWS bietet über die AWS-Service-APIs und -SDKs auch Support für Protokollanalysesysteme und Business Intelligence-Tools von Drittanbietern (z. B. Grafana, Kibana und Logstash).

Gängige Antimuster:

- Die Seitenreaktionszeit wurde noch nie mit der Kundenzufriedenheit in Verbindung gebracht. Sie haben noch nie eine Metrik oder einen Schwellenwert für die Seitenreaktionszeit festgelegt. Ihre Kunden beschweren sich über langsame Ladevorgänge.
- Sie haben Ihre Zielwerte für die minimale Reaktionszeit nicht erreicht. Um die Reaktionszeit zu verbessern, haben Sie Ihre Anwendungsserver skaliert. Sie erzielen jetzt Reaktionszeiten, die weit über die Zielwerte hinausgehen, und haben erhebliche ungenutzte Kapazitäten, für die Sie zahlen.

Vorteile der Einführung dieser bewährten Praxis: Wenn Sie KPIs und Metriken überprüfen und überarbeiten, können Sie nachvollziehen, wie sich Ihr Workload auf die Geschäftsergebnisse auswirkt, und ermitteln, wo Verbesserungen erforderlich sind, um die Geschäftsziele zu erreichen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Erfolg von Ergebnissen und die Effektivität von KPIs und Metriken prüfen: Erstellen Sie eine Geschäftsansicht Ihrer Workload-Vorgänge, um festzustellen, ob Sie die Anforderungen erfüllen, und um Bereiche zu identifizieren, die verbessert werden müssen, um Geschäftsziele zu erreichen. Prüfen Sie die Wirksamkeit von KPIs und Metriken und überarbeiten Sie diese gegebenenfalls.
  - Verwendung von Amazon CloudWatch-Dashboards
  - Was ist Protokollanalytik?

Ressourcen

Verbundene Dokumente:

- Verwendung von Amazon CloudWatch-Dashboards
- Was ist Protokollanalytik?

OPS 9 Wie können Sie den Zustand Ihrer Operationen beurteilen?

Definieren, erfassen und analysieren Sie Metriken für Operationen, um einen Einblick in Ereignisse rund um Ihre operativen Abläufe zu erhalten. Dies ist wichtig, damit Sie bei Bedarf entsprechende Maßnahmen ergreifen können.

Bewährte Methoden

- OPS09-BP01 Ermitteln wichtiger Leistungskennzahlen
- OPS09-BP02 Definieren von Betriebsmetriken
- OPS09-BP03 Erfassen und Analysieren von Betriebsmetriken
- OPS09-BP04 Festlegen von Ausgangswerten für Betriebsmetriken
- OPS09-BP05 Aufzeichnen der erwarteten Aktivitätsmuster für den Betrieb
- OPS09-BP06 Alarm bei gefährdeten Ergebnissen von Operationen
- OPS09-BP07 Alarm bei festgestellten Betriebsanomalien
- OPS09-BP08 Prüfen der Erreichung von angestrebten Ergebnissen und der Wirksamkeit von KPIs und Metriken

OPS09-BP01 Ermitteln wichtiger Leistungskennzahlen

Ermitteln Sie wichtige Leistungskennzahlen (KPIs) anhand der gewünschten Geschäftsergebnisse (z. B. bereitgestellte neue Funktionen) und Kundenergebnisse (z. B. Kundenservice-Anfragen). Bewerten Sie zur Messung des Erfolgs von Operationen KPIs.

Gängige Antimuster:

- Sie werden von der Geschäftsleitung gefragt, wie erfolgreich der Betrieb die Geschäftsziele erreicht, aber haben keinen Referenzrahmen, um den Erfolg zu bestimmen.
- Sie können nicht feststellen, ob sich Ihre geplanten Wartungsarbeiten auf die Geschäftsergebnisse auswirken.

Vorteile der Einführung dieser bewährten Methode: Durch die Ermittlung wichtiger Leistungskennzahlen ermöglichen Sie das Erreichen von Geschäftsergebnissen als Test des Zustands und Erfolgs Ihrer Betriebsabläufe.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Ermitteln wichtiger Leistungskennzahlen: Ermitteln Sie auf Basis der gewünschten geschäftlichen und kundenspezifischen Ergebnisse wichtige Leistungskennzahlen (Key Performance Indicators, KPIs). Bewerten Sie zur Messung des Erfolgs von Operationen KPIs.

OPS09-BP02 Definieren von Betriebsmetriken

Definieren Sie Betriebsmetriken, um den Erfolg von KPIs zu messen (z. B. erfolgreiche und fehlgeschlagene Bereitstellungen). Definieren Sie Betriebsmetriken, um den Zustand von Betriebsaktivitäten zu messen (z. B. mittlere Zeit zur Erkennung eines Vorfalls (MTTD) und mittlere Reparaturzeit (MTTR) nach einem Vorfall). Bewerten Sie Metriken, um festzustellen, ob die Betriebsabläufe die gewünschten Ergebnisse erzielen, und um den Zustand der Betriebsaktivitäten zu beurteilen.

Gängige Antimuster:

- Ihre Betriebsmetriken basieren auf den Werten, die das Team für angemessen hält.
- In Ihren Metrikberechnungen liegen Fehler vor, die zu falschen Ergebnissen führen.
- Sie haben keine Metriken für Ihre Betriebsaktivitäten definiert.

Vorteile der Einführung dieser bewährten Methode: Durch das Definieren und Auswerten von Betriebsmetriken können Sie den Zustand Ihrer Betriebsaktivitäten bestimmen und den Fortschritt beim Erreichen der Geschäftsergebnisse messen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Definieren von Betriebsmetriken: Definieren Sie operationsspezifische Metriken für die Analyse der Erfüllung von KPIs. Definieren Sie operationsspezifische Metriken, um den Zustand der Operationen und ihrer Aktivitäten beurteilen zu können. Bewerten Sie Metriken, um festzustellen,

ob Operationen die gewünschten Ergebnisse erzielen, und um den Zustand der Operationen zu beurteilen.

- Veröffentlichen von benutzerdefinierten Metriken
- Suchen und Filtern von Protokolldaten
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch

Ressourcen

Zugehörige Dokumente:

- AWS-Antworten: zentralisierte Protokollierung
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- Erkennen von und Reagieren auf Änderungen im Pipeline-Zustand mit Amazon CloudWatch Events
- Veröffentlichen von benutzerdefinierten Metriken
- Suchen und Filtern von Protokolldaten

Relevante Videos:

- Erstellen eines Überwachungsplans

OPS09-BP03 Erfassen und Analysieren von Betriebsmetriken

Unterziehen Sie die Metriken regelmäßigen proaktiven Überprüfungen, um Trends zu ermitteln und festzustellen, wo gegebenenfalls Maßnahmen ergriffen werden müssen.

Sie sollten Protokolldaten aus der Ausführung Ihrer Betriebsaktivitäten und Betriebs-API-Aufrufe in einem Service wie CloudWatch Logs zusammenfassen. Generieren Sie Metriken aus Beobachtungen der erforderlichen Protokollinhalte, um Einblicke in die Leistung von Betriebsaktivitäten zu erhalten.

In AWS können Sie Ihre Protokolldaten zu Amazon S3 exportieren oder Protokolle zur langfristigen Speicherung direkt um Amazon S3 senden. Mit AWS Gluekönnen Sie Ihre Protokolldaten in Amazon S3 zur Analyse erkunden und vorbereiten und die zugehörigen Metadaten im AWSAWS Glue Data Catalog. Amazon Athenakann dann durch eine native Integration mit AWS Glue zum Analysieren Ihrer Protokolldaten und für Abfragen mit Standard-SQL verwendet werden. Mit einem Business Intelligence-Tool wie Amazon QuickSight können Sie Ihre Daten visualisieren, untersuchen und analysieren.

Gängige Antimuster:

- Die regelmäßige Bereitstellung neuer Funktionen gilt als wichtige Leistungskennzahl. Sie haben keine Möglichkeit, um die Häufigkeit von Bereitstellungen zu messen.
- Sie protokollieren Bereitstellungen, rückgängig gemachte Bereitstellungen, Patches und rückgängig gemachte Patches, um Ihre Betriebsaktivitäten zu verfolgen, aber die Metriken werden von niemandem überprüft.
- Sie haben ein Recovery Time Objective von 15 Minuten für die Wiederherstellung ausgefallener Datenbanken, das bei der Bereitstellung des Systems festgelegt wurde, als es noch nicht im Einsatz war. Heute haben Sie 10 000 Benutzer und Ihr System ist seit 2 Jahren in Betrieb. Eine kürzliche Wiederherstellung dauerte mehr als 2 Stunden. Dies wurde aber nicht aufgezeichnet, sodass niemand davon weiß.

Vorteile der Einführung dieser bewährten Praxis: Durch das Erfassen und Analysieren Ihrer Betriebsmetriken gewinnen Sie einen Überblick über den Zustand Ihrer Betriebsabläufe und erhalten Einblicke in Trends, die sich auf Ihren Betrieb oder Ihre Geschäftsergebnisse auswirken können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Betriebsmetriken erfassen und analysieren: Unterziehen Sie die Metriken regelmäßigen proaktiven Überprüfungen, um Trends ermitteln und feststellen zu können, wo gegebenenfalls geeignete Maßnahmen ergriffen werden müssen.
  - Verwenden von Amazon CloudWatch-Metriken
  - Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
  - Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent

Ressourcen

Verbundene Dokumente:

- Amazon Athena
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- Amazon QuickSight

- AWS Glue
- AWSAWS Glue Data Catalog
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und lokalen Servern mit dem CloudWatch Agent
- Verwenden von Amazon CloudWatch-Metriken

OPS09-BP04 Festlegen von Ausgangswerten für Betriebsmetriken

Legen Sie Ausgangswerte für Metriken fest, um erwartete Werte als Grundlage für den Vergleich und die Ermittlung von Betriebsaktivitäten mit unter- oder überdurchschnittlicher Leistung bereitzustellen.

Gängige Antimuster:

- Sie werden gefragt, wie viel Zeit die Bereitstellung voraussichtlich in Anspruch nimmt. Da Sie die Bereitstellungsdauer nicht gemessen haben, können Sie die voraussichtlich erforderliche Zeit nicht bestimmen.
- Sie werden gefragt, wie lange die Wiederherstellung nach einem Problem mit den Anwendungsservern dauert. Sie haben keine Informationen über die Wiederherstellungsdauer nach dem ersten Kundenkontakt. Sie haben keine Informationen über die Wiederherstellungsdauer ab der erstmaligen Ermittlung eines Problems im Rahmen der Überwachung.
- Sie werden gefragt, wie viele Supportmitarbeiter am Wochenende benötigt werden. Sie haben keine Ahnung, wie viele Supportanfragen üblicherweise an einem Wochenende eingehen und können keine geschätzte Anzahl nennen.
- Sie haben ein Recovery Time Objective von 15 Minuten für die Wiederherstellung ausgefallener Datenbanken, das bei der Bereitstellung des Systems festgelegt wurde, als es noch nicht im Einsatz war. Heute haben Sie 10 000 Benutzer und Ihr System ist seit 2 Jahren in Betrieb. Sie haben keine Informationen darüber, wie sich die Wiederherstellungsdauer für Ihre Datenbank geändert hat.

Vorteile der Einführung dieser bewährten Methode: Durch die Definition von Metrikausgangswerten können Sie aktuelle Metrikwerte und Metriktrends auswerten, um festzustellen, ob Maßnahmen erforderlich sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mehr über erwartete Aktivitätsmuster für den Betrieb erfahren: Legen Sie Muster für die betriebliche Aktivität fest, um festzustellen, wann das Verhalten von den erwarteten Werten abweicht, so dass Sie bei Bedarf angemessen reagieren können.

OPS09-BP05 Aufzeichnen der erwarteten Aktivitätsmuster für den Betrieb

Legen Sie Betriebsaktivitätsmuster fest, um außergewöhnliche Aktivitäten zu identifizieren, damit Sie bei Bedarf entsprechend reagieren können.

Gängige Antimuster:

- Ihre Bereitstellungsfehlerrate hat sich in letzter Zeit erheblich erhöht. Sie beheben die Fehler unabhängig voneinander. Ihnen fällt nicht auf, dass alle Fehler bei den Bereitstellungen eines neuen Mitarbeiters auftreten, der nicht mit dem System zur Bereitstellungsverwaltung vertraut ist.

Vorteile der Einführung dieser bewährten Methode: Durch das Aufzeichnen von Verhaltensmustern können Sie unerwartetes Verhalten erkennen und bei Bedarf Maßnahmen ergreifen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mehr über erwartete Aktivitätsmuster für den Betrieb erfahren: Legen Sie Muster für die betriebliche Aktivität fest, um festzustellen, wann das Verhalten von den erwarteten Werten abweicht, so dass Sie bei Bedarf angemessen reagieren können.

OPS09-BP06 Alarm bei gefährdeten Ergebnissen von Operationen

Wenn die Ergebnisse von Operationen in Gefahr sind, muss ein Alarm ausgegeben und darauf entsprechend reagiert werden. Dabei handelt es sich um alle Aktivitäten, die einen Workload in Produktion unterstützen. Dies umfasst alles von der Bereitstellung neuer Anwendungsversionen bis zur Wiederherstellung nach einem Ausfall. Die Ergebnisse von Operationen müssen als ähnlich wichtig behandelt werden wie Geschäftsergebnisse.

Softwareteams sollten die zentralen betrieblichen Metriken und Aktivitäten identifizieren und Alarme dafür einrichten. Alarme müssen zeitnah erfolgen und konkretes Handeln ermöglichen. Wenn ein

Alarm ausgegeben wird, sollte dazu ein Verweis zu einem entsprechenden Runbook oder Playbook gehören. Alarme ohne zugehörige Aktionen können zu Alarmermüdung führen.

Gewünschtes Ergebnis: Wenn Betriebsabläufe gefährdet sind, werden Alarme ausgesendet, um Maßnahmen auszulösen. Die Alarme enthalten Kontextinformationen dazu, warum der Alarm ausgegeben wurde, und verweisen auf ein Playbook für die Untersuchung oder ein Runbook für Abhilfemaßnahmen. Wo immer möglich, werden Runbooks automatisiert und Benachrichtigungen gesendet.

Typische Anti-Muster:

- Sie untersuchen einen Vorgang und registrieren Support-Fälle. Die Support-Fälle verstoßen gegen das Service Level Agreement (SLA), es werden aber keine Alarme ausgegeben.
- Eine für Mitternacht geplante Produktionsbereitstellung verzögert sich aufgrund von Code-Änderungen in letzter Minute. Es wird kein Alarm ausgegeben und die Bereitstellung steht still.
- Es tritt ein Produktionsausfall auf, es werden aber keine Alarme gesendet.
- Ihre Bereitstellungszeit fällt konsistent hinter den Schätzungen zurück. Es wird nichts unternommen, um dies zu untersuchen.

Vorteile der Nutzung dieser bewährten Methode:

- Ein Alarm bei einer Gefährdung der Ergebnisse von Operationen verbessert Ihre Fähigkeit, Ihren Workload zu unterstützen, da Sie Problemen immer einen Schritt voraus sind.
- Die geschäftlichen Ergebnisse werden dank korrekter Ergebnisse von Operationen verbessert.
- Erkennung und Korrektur von Betriebsproblemen werden verbessert.
- Insgesamt wird der Betriebszustand verbessert.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Mittel

Implementierungsleitfaden

Ergebnisse von Operationen müssen definiert werden, bevor Sie damit beginnen können, Alarme dafür einzurichten. Legen Sie zunächst fest, welche betrieblichen Aktivitäten für Ihre Organisation die wichtigsten sind. Ist es die Bereitstellung zur Produktion in weniger als zwei Stunden oder die Reaktion auf einen Support-Fall innerhalb eines festgelegten Zeitraums? Ihre Organisation muss ihre zentralen betrieblichen Aktivitäten und deren Messung definieren, damit diese überwacht, verbessert

und Gegenstand von Alarmen sein können. Sie benötigen einen zentralen Ort für die Speicherung und Analyse von Workload- und Betriebstelemetriedaten. Dieser Mechanismus sollte auch einen Alarm ausgeben können, wenn das Ergebnis einer Operation in Gefahr ist.

Kundenbeispiel

Während einer Routine-Bereitstellung bei AnyCompany Retail wurde ein CloudWatch-Alarm ausgelöst. Die Durchlaufzeit für die Bereitstellung wurde nicht eingehalten. Amazon EventBridge erstellte ein OpsItem in AWS Systems Manager OpsCenter. Das Cloud-Operations-Team untersuchte das Problem anhand eines Playbooks und fand heraus, dass ein Schemawechsel länger dauerte als erwartet. Das Team benachrichtigte den zuständigen Entwickler und beobachtete die Bereitstellung weiter. Nach Abschluss der Bereitstellung löste das Cloud-Operations-Team das OpsItem. Das Team analysiert den Vorfall im Rahmen eines Postmortem-Gesprächs.

Implementierungsschritte

1. Wenn Sie keine Betriebs-KPIs, Metriken und Aktivitäten identifiziert haben, arbeiten Sie an der Implementierung der obigen bewährten Methoden für diese Frage (OPS09-BP01 bis OPS09-BP05).
   - AWS Support-Kunden mit Enterprise Support können den Operations KPI Workshop bei ihrem Technical Account Manager anfordern. Dieser auf Zusammenarbeit ausgerichtete Workshop hilft Ihnen bei der Definition von betrieblichen KPIs und Metriken unter Berücksichtigung Ihrer geschäftlichen Ziele und ist ohne zusätzliche Kosten verfügbar. Wenden Sie sich an Ihren Technical Account Manager, um weitere Informationen zu erhalten.
2. Sobald Sie betriebliche Aktivitäten, KPIs und Metriken eingerichtet haben, konfigurieren Sie Alarme in Ihrer Beobachtungsplattform. Alarmen sollte eine konkrete Maßnahme zugeordnet sein, etwa ein Playbook oder ein Runbook. Alarme ohne Maßnahmen sollten vermieden werden.
3. Mit der Zeit sollten Sie Ihre betrieblichen Metriken, KPIs und Aktivitäten evaluieren, um Bereiche für mögliche Verbesserungen zu identifizieren. Erfassen Sie Feedback von Bedienern in Runbooks und Playbooks, um in Reaktion auf Alarme Bereiche für mögliche Verbesserungen zu identifizieren.
4. Alarme sollten einen Mechanismus enthalten, der es erlaubt, sie als falsche positiv zu markieren. Dies sollte zu einer Überprüfung der Metrik-Schwellenwerte führen.

Aufwand für den Implementierungsplan: Mittel. Es gibt verschiedene bewährte Methoden, die vor der Implementierung dieser Methode eingerichtet werden müssen. Sobald betriebliche Aktivitäten identifiziert und betriebliche KPIs eingerichtet wurden, sollten die Alarme eingerichtet werden.

Ressourcen

Zugehörige bewährte Methoden:

- OPS02-BP03 Betriebsaktivitäten haben feste Besitzer, die für ihre Leistung verantwortlich sind: Jede betriebliche Aktivität und jedes betriebliche Ergebnis sollte einen identifizierten Eigentümer haben, der dafür verantwortlich ist. Diese Person ist zu benachrichtigen, wenn Ergebnisse in Gefahr sind.
- OPS03-BP02 Teammitglieder sind befugt, Maßnahmen zu ergreifen, wenn Ergebnisse gefährdet sind:: Wenn Alarme ausgegeben werden, sollte Ihr Team in der Lage sein, Maßnahmen zu ergreifen, um das Problem zu beheben.
- OPS09-BP01 Ermitteln wichtiger Leistungskennzahlen: Die Alarmierung zu Ergebnissen von Operationen beginnt mit der Identifizierung der betrieblichen KPIs.
- OPS09-BP02 Definieren von Betriebsmetriken: Richten Sie diese bewährte Methode ein, bevor Sie mit der Generierung von Alarmen beginnen.
- OPS09-BP03 Erfassen und Analysieren von Betriebsmetriken: Zum Aufbau von Alarmen ist die zentrale Erfassung betrieblicher Metriken erforderlich.
- OPS09-BP04 Festlegen von Ausgangswerten für Betriebsmetriken: Baselines für betriebliche Metriken ermöglichen die Feineinstellung von Alarmen, um Alarmermüdung zu vermeiden.
- OPS09-BP05 Aufzeichnen der erwarteten Aktivitätsmuster für den Betrieb: Sie können die Korrektheit Ihrer Alarme verbessern, wenn Sie die Aktivitätsmuster für betriebliche Ereignisse verstehen.
- OPS09-BP08 Prüfen der Erreichung von angestrebten Ergebnissen und der Wirksamkeit von KPIs und Metriken: Evaluieren Sie das Erreichen der Ergebnisse von Operationen, um sicherzustellen, dass Ihre KPIs und Metriken korrekt sind.
- OPS10-BP02 Implementieren eines Prozesses für jeden Alarm: Jedem Alarm sollte ein Playbook oder Runbook zugeordnet sein und er muss Kontext für die alarmierte Person enthalten.
- OPS11-BP02 Durchführen von Analysen nach Vorfällen: Führen Sie nach dem Alarm eine Analyse durch, um Bereiche für Verbesserungen zu identifizieren.

Zugehörige Dokumente:

- AWS-Bereitstellungspipeline-Referenzarchitektur: Anwendungspipelinearchitektur
- GitLab: Erste Schritte mit Agile/DevOps Metrics

Zugehörige Videos:

- Aggregate and Resolve Operational Issues Using AWS Systems Manager OpsCenter (Aggregieren und Beheben betrieblicher Probleme mit AWS Systems Manager OpsCenter)
- Integrate AWS Systems Manager OpsCenter with Amazon CloudWatch Alarms (Integrieren von AWS Systems Manager OpsCenter in Amazon CloudWatch-Alarme)
- Integrate Your Data Sources into AWS Systems Manager OpsCenter Using Amazon EventBridge (Integrieren Ihrer Datenquellen in AWS Systems Manager OpsCenter mit Amazon EventBridge)

Zugehörige Beispiele:

- Automatisieren von Behebungsaktionen für Amazon EC2-Benachrichtigungen und mehr mithilfe von Amazon EC2 Systems Manager Automation und AWS Health
- AWS Management and Governance Tools Workshop - Operations 2022
- Aufnahme, Analyse und Visualisierung von Metriken mit dem DevOps Monitoring Dashboard auf AWS

Zugehörige Services:

- Amazon EventBridge
- AWS Support Proactive Services - Operations KPI Workshop
- AWS Systems Manager OpsCenter
- CloudWatch-Ereignisse

OPS09-BP07 Alarm bei festgestellten Betriebsanomalien

Lösen Sie einen Alarm aus, wenn Betriebsanomalien festgestellt werden, damit Sie bei Bedarf angemessen reagieren können.

Die Analyse Ihrer Betriebsmetriken im Laufe der Zeit kann Verhaltensmuster feststellen, die Sie ausreichend quantifizieren können, um ein Ereignis zu definieren oder als Reaktion einen Alarm auszulösen.

Nach der Schulung kann die Funktion Funktion CloudWatch Anomaly Detection verwendet werden, um bei erkannten Anomalien einen Alarm auszulösen oder überlagerte erwartete Werte in einem Diagramm mit Metrikdaten für einen laufenden Vergleich bereitzustellen.

Amazon DevOps Guru kann verwendet werden, um außergewöhnliches Verhalten über die Korrelation von Ereignissen, Protokollanalysen und die Anwendung von Machine Learning zu identifizieren und Ihre Workload-Telemetrie zu analysieren. Die erhaltenen Einblicke werden mit den relevanten Daten und Empfehlungen dargestellt.

Gängige Antimuster:

- Sie wenden einen Patch auf Ihre Instance-Flotte an. In der Testumgebung haben Sie den Patch erfolgreich getestet. Für einen hohen Anteil der Instances in Ihrer Flotte schlägt der Patch fehl. Sie unternehmen nichts.
- Sie stellen fest, dass Freitag am Ende des Tages Bereitstellungen anstehen. Die Wartungsfenster Ihres Unternehmens sind auf dienstags und donnerstags festgelegt. Sie unternehmen nichts.

Vorteile der Einführung dieser bewährten Praxis: Wenn Sie mit Betriebsverhaltensmustern vertraut sind, können Sie unerwartetes Verhalten identifizieren und bei Bedarf Maßnahmen ergreifen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Alarm bei festgestellten Betriebsanomalien auslösen: Lösen Sie einen Alarm aus, wenn Betriebsanomalien erkannt werden, damit Sie bei Bedarf entsprechend reagieren können.
  - Was ist Amazon CloudWatch Events?
  - Erstellen von Amazon CloudWatch-Alarmen
  - Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen

Ressourcen

Verbundene Dokumente:

- Amazon DevOps Guru
- Funktion CloudWatch Anomaly Detection
- Erstellen von Amazon CloudWatch-Alarmen
- Erkennen von und Reagieren auf Änderungen im Pipeline-Zustand mit Amazon CloudWatch Events
- Auslösen von Lambda-Funktionen mit Amazon SNS-Benachrichtigungen
- Was ist Amazon CloudWatch Events?

OPS09-BP08 Prüfen der Erreichung von angestrebten Ergebnissen und der Wirksamkeit von KPIs und Metriken

Erstellen Sie eine Ansicht Ihrer operationsspezifischen Aktivitäten auf Geschäftsebene, mit der Sie schnell feststellen können, ob Sie die Anforderungen erfüllen, und welche Bereiche verbessert werden müssen, um die Geschäftsziele zu erreichen. Prüfen Sie die Wirksamkeit von KPIs und Metriken und überarbeiten Sie diese gegebenenfalls.

AWS bietet über die AWS-Service-APIs und -SDKs auch Support für Protokollanalysesysteme und Business-Intelligence-Tools von Drittanbietern (z. B. Grafana, Kibana und Logstash).

Gängige Antimuster:

- Die Häufigkeit Ihrer Bereitstellungen ist mit der wachsenden Anzahl von Entwicklerteams gestiegen. Ursprünglich hatten sie festgelegt, dass einmal pro Woche bereitgestellt wird. Mittlerweile führen Sie jeden Tag Bereitstellungen durch. Wenn ein Problem mit Ihrem Bereitstellungssystem auftritt und keine Bereitstellungen möglich sind, kann es mehrere Tage dauern, bis das Problem erkannt wird.
- Bis vor Kurzem war der Support Ihres Unternehmens nur in den Kerngeschäftszeiten von Montag bis Freitag erreichbar. Als Reaktionszeit für Vorfälle galt dabei „am nächsten Werktag“. Jetzt bieten Sie Support rund um die Uhr mit einer Reaktionszeit von 2 Stunden. Die Mitarbeiter der Nachtschicht sind überfordert und die Kunden sind unzufrieden. Es liegen keine Hinweise darauf vor, dass die Reaktionszeiten bei Vorfällen nicht eingehalten werden, da weiterhin das Ziel „am nächsten Werktag“ gilt.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie KPIs und Metriken überprüfen und überarbeiten, können Sie nachvollziehen, wie sich Ihr Workload auf die Geschäftsergebnisse auswirkt, und ermitteln, wo Verbesserungen erforderlich sind, um die Geschäftsziele zu erreichen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Erfolg von Ergebnissen und die Effektivität von KPIs und Metriken prüfen: Erstellen Sie eine Geschäftsansicht Ihrer Betriebsaktivitäten, um festzustellen, ob Sie die Anforderungen erfüllen, und um Bereiche zu identifizieren, die verbessert werden müssen, um Geschäftsziele zu erreichen. Prüfen Sie die Wirksamkeit von KPIs und Metriken und überarbeiten Sie diese gegebenenfalls.
  - Verwendung von Amazon CloudWatch-Dashboards
  - Was ist Protokollanalytik?

Ressourcen

Zugehörige Dokumente:

- Verwendung von Amazon CloudWatch-Dashboards
- Was ist Protokollanalytik?

OPS 10 Wie bewältigen Sie Workload- und operationsspezifische Ereignisse?

Erarbeiten und prüfen Sie Verfahren für die Reaktion auf Ereignisse, um Beeinträchtigungen für Ihren Workload zu minimieren.

Bewährte Methoden

- OPS10-BP01 Verwenden eines Prozesses für die Bewältigung von Ereignissen, Vorfällen und Problemen
- OPS10-BP02 Implementieren eines Prozesses für jeden Alarm
- OPS10-BP03 Priorisieren von betrieblichen Ereignissen auf Basis der Auswirkung auf das Unternehmen
- OPS10-BP04 Definieren von Eskalationspfaden
- OPS10-BP05 Aktivieren von Push-Benachrichtigungen
- OPS10-BP06 Bekanntgeben des Status über Dashboards
- OPS10-BP07 Automatisieren von Reaktionen auf Ereignisse

OPS10-BP01 Verwenden eines Prozesses für die Bewältigung von Ereignissen, Vorfällen und Problemen

Ihre Organisation hat Prozesse für die Bewältigung von Ereignissen, Vorfällen und Problemen. Ereignisse sind Dinge, die in Ihrem Workload auftreten, aber möglicherweise kein Eingreifen erfordern. Vorfälle sind Ereignisse, die ein Eingreifen erfordern. Probleme sind wiederkehrende Ereignisse, die ein Eingreifen erfordern oder nicht behoben werden können. Sie benötigen Prozesse, um die Auswirkungen solcher Ereignisse auf Ihr Unternehmen zu mindern und um sicherzustellen, dass Sie in angemessener Weise darauf reagieren.

Wenn Ihr Workload von Vorfällen und Problemen betroffen ist, benötigen Sie Prozesse, um diese zu bewältigen. Wie informieren Sie Stakeholder über den Status des Ereignisses? Wer leitet die Reaktion? Welche Tools verwenden Sie, um das Ereignis abzumildern? Dies sind Beispiele für Fragen, die Sie beantworten müssen, um einen fundierten Reaktionsprozess einführen zu können.

Prozesse müssen an zentraler Stelle dokumentiert werden und allen am Workload Beteiligten zur Verfügung stehen. Wenn Sie nicht über ein zentrales Wiki oder einen zentralen Dokumentenspeicher verfügen, können Sie dafür ein Repository für die Versionskontrolle verwenden. Sie halten diese Pläne aktuell, wenn sich die Prozesse weiterentwickeln.

Probleme sind Kandidaten für eine Automatisierung. Diese Ereignisse nehmen Zeit in Anspruch, die Sie eigentlich für Innovationen benötigen. Beginnen Sie mit der Entwicklung eines wiederholbaren Prozesses, um das Problem abzumildern. Konzentrieren Sie sich im Laufe der Zeit darauf, die Abmilderung zu automatisieren oder das zugrunde liegende Problem zu beheben. Dadurch sparen Sie Zeit ein, die Sie für Verbesserungen an Ihrem Workload aufwenden können.

Gewünschtes Ergebnis: Ihre Organisation hat einen Prozess für die Bewältigung von Ereignissen, Vorfällen und Problemen. Diese Prozesse werden dokumentiert und an zentraler Stelle gespeichert. Sie werden aktualisiert, wenn sich die Prozesse ändern.

Typische Anti-Muster:

- Ein Vorfall tritt am Wochenende ein und der Entwickler, der Rufbereitschaft hat, weiß nicht, was zu tun ist.
- Ein Kunde sendet Ihnen eine E-Mail, dass die Anwendung nicht verfügbar ist. Sie starten den Server neu, um das Problem zu beheben. Dies kommt häufig vor.
- Es gibt einen Vorfall und mehrere Teams arbeiten unabhängig voneinander daran, das Problem zu beheben.
- Es kommt zu Bereitstellungen in Ihrem Workload, die nicht dokumentiert werden.

Vorteile der Nutzung dieser bewährten Methode:

- Es gibt einen Prüfpfad der Ereignisse in Ihrem Workload.
- Die erforderliche Zeit für die Wiederherstellung nach einem Vorfall verringert sich.
- Die Teammitglieder können Vorfälle und Probleme einheitlich beheben.
- Bei der Untersuchung eines Vorfalls sind die Anstrengungen stärker miteinander verbunden.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Wenn Sie diese Best Practice implementieren, bedeutet dies, dass Sie Workload-Ereignisse nachverfolgen. Sie haben Prozesse für den Umgang mit Vorfällen und Problemen. Die Prozesse

werden dokumentiert, geteilt und oft aktualisiert. Probleme werden identifiziert, priorisiert und behoben.

Kundenbeispiel

AnyCompany Retail verwendet einen Teil seines internen Wikis für Prozesse zur Verwaltung von Ereignissen, Vorfällen und Problemen. Alle Ereignisse werden an Amazon EventBridgegesendet. Probleme werden in AWS Systems Manager OpsCenter als OpsItems identifiziert und zur Behebung priorisiert, sodass undifferenzierter Arbeitsaufwand reduziert wird. Wenn die Prozesse sich ändern, werden sie im internen Wiki aktualisiert. Das Unternehmen nutzt AWS Systems Manager Incident Manager für die Verwaltung von Vorfällen und das Koordinieren von Maßnahmen zur Abmilderung.

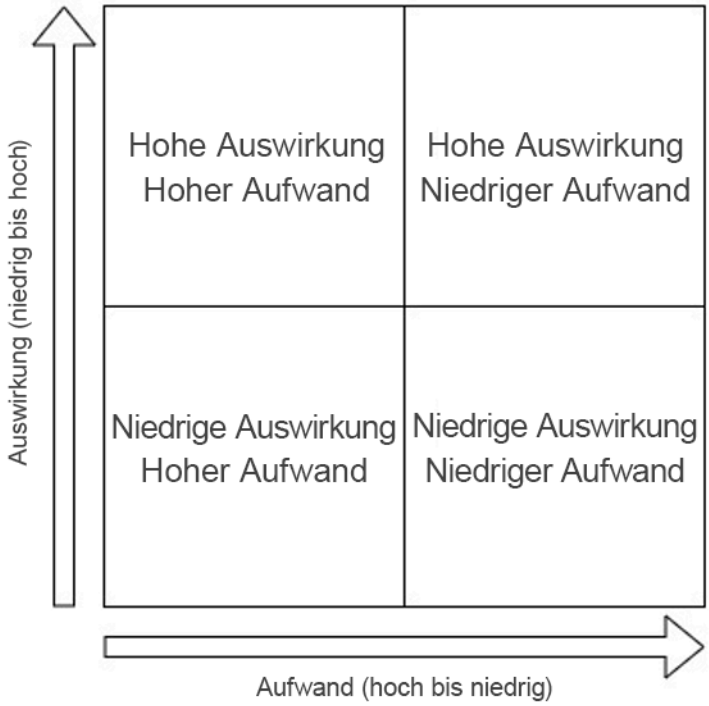
Implementierungsschritte

1. Ereignisse
   - Verfolgen Sie Ereignisse in Ihrem Workload nach, auch wenn kein menschliches Eingreifen erforderlich ist.
   - Entwickeln Sie gemeinsam mit den Workload-Stakeholdern eine Liste der Ereignisse, die nachverfolgt werden sollten. Beispiele sind abgeschlossene Bereitstellungen oder erfolgreiche Patches.
   - Sie können Services wie Amazon EventBridge oder Amazon Simple Notification Service nutzen, um benutzerdefinierte Ereignisse für die Nachverfolgung zu generieren.
2. Vorfälle
   - Definieren Sie zunächst den Kommunikationsplan für Vorfälle. Welche Stakeholder müssen informiert werden? Wie werden Sie sie auf dem Laufenden halten? Wer leitet die Koordination der Arbeiten? Wir empfehlen, einen internen Chat-Kanal für die Kommunikation und Koordination einzurichten.
   - Definieren Sie Eskalationspfade für die Teams, die Ihren Workload unterstützen, insbesondere wenn es im Team keine Rufbereitschaft gibt. Basierend auf Ihrem Support-Level können Sie auch einen Fall beim AWS Support öffnen.
   - Erstellen Sie ein Playbook, um den Vorfall zu untersuchen. Dieses sollte den Kommunikationsplan sowie detaillierte Maßnahmen zur Untersuchung beinhalten. Nehmen Sie in Ihre Untersuchung auch die Überprüfung von AWS Health Dashboard auf.
   - Dokumentieren Sie Ihren Reaktionsplan für Vorfälle. Kommunizieren Sie den Plan für das Vorfallmanagement, damit interne und externe Kunden die Regeln der Interaktion verstehen und wissen, was von ihnen erwartet wird. Schulen Sie die Teammitglieder hinsichtlich der Verwendung.

- Kunden können Incident Manager nutzen, um ihren Reaktionsplan für Vorfälle einzurichten und zu verwalten.
- Kunden mit Enterprise Support können den Workshop zum Vorfallmanagement bei ihrem Technical Account Manager anfordern. Dieser angeleitete Workshop testet Ihren vorhandenen Reaktionsplan für Vorfälle und hilft Ihnen, Verbesserungsmöglichkeiten zu identifizieren.

3. Probleme

- Probleme müssen identifiziert und in Ihrem ITSM-System nachverfolgt werden.
- Identifizieren Sie alle bekannten Probleme und priorisieren Sie sie nach Aufwand der Behebung und Auswirkungen auf den Workload.



- Beheben Sie zunächst Probleme, die mit erheblichen Auswirkungen und geringem Aufwand verbunden sind. Sobald diese behoben sind, wechseln Sie zu Problemen, die in den Quadranten der Probleme mit geringen Auswirkungen und geringem Aufwand fallen.
- Sie können Systems Manager OpsCenter verwenden, um diese Probleme zu identifizieren, Runbooks daran anzufügen und sie nachzuverfolgen.

Aufwand für den Implementierungsplan: Mittel. Sie benötigen einen Prozess und Tools, um diese Best Practice zu implementieren. Dokumentieren Sie Ihre Prozesse und stellen Sie sie allen am Workload Beteiligten zur Verfügung. Aktualisieren Sie sie häufig. Sie haben einen Prozess für die Verwaltung und Abmilderung oder Behebung von Problemen.

Ressourcen

Zugehörige bewährte Methoden:

- OPS07-BP03 Verwenden von Runbooks zur Durchführung von Verfahren: Bekannte Probleme benötigen ein angefügtes Runbook, damit die Maßnahmen zur Abmilderung einheitlich sind.
- OPS07-BP04 Verwenden von Playbooks zum Untersuchen von Problemen: Vorfälle müssen mithilfe von Playbooks untersucht werden.
- OPS11-BP02 Durchführen von Analysen nach Vorfällen: Führen Sie nach der Wiederherstellung nach einem Vorfall stets eine Post-Mortem-Analyse durch.

Zugehörige Dokumente:

- Atlassian - Incident management in the age of DevOps
- Leitfaden für AWS Security Incident Response
- Incident Management in the Age of DevOps and SRE
- PagerDuty - What is Incident Management?

Zugehörige Videos:

- AWS re:Invent 2020: Incident management in a distributed organization
- AWS re:Invent 2021 - Building next-gen applications with event-driven architectures
- AWS Supports You | Exploring the Incident Management Tabletop Exercise
- AWS Systems Manager Incident Manager - AWS Virtual Workshops
- AWS What's Next ft. Incident Manager | AWS Events

Zugehörige Beispiele:

- AWS Management and Governance Tools Workshop - OpsCenter
- AWS Proactive Services – Incident Management Workshop
- Building an event-driven application with Amazon EventBridge
- Building event-driven architectures on AWS

Zugehörige Services:

- Amazon EventBridge
- Amazon SNS
- AWS Health Dashboard
- AWS Systems Manager Incident Manager
- AWS Systems Manager OpsCenter

OPS10-BP02 Implementieren eines Prozesses für jeden Alarm

Legen Sie für jedes Ereignis, für das Sie einen Alarm auslösen, eine klar definierte Reaktion (Runbook oder Playbook) mit einem eigens dafür angegebenen Besitzer fest. Dies gewährleistet eine effektive und schnelle Reaktion auf Betriebsereignisse und verhindert, dass aktionsrelevante Ereignisse aufgrund weniger wichtiger Benachrichtigungen übersehen werden.

Gängige Antimuster:

- Ihr Überwachungssystem präsentiert Ihnen einen Stream genehmigter Verbindungen zusammen mit anderen Nachrichten. Die Menge der Nachrichten ist so groß, dass Sie regelmäßig Fehlermeldungen verpassen, die eigentlich Ihren Eingriff erfordern würden.
- Sie erhalten eine Warnung, dass die Website nicht verfügbar ist. Es gibt keinen definierten Prozess dafür, wann dies geschieht. Sie müssen das Problem mit einem Ad-hoc-Ansatz diagnostizieren und lösen. Durch die individuelle Fehlerbehebung ohne vorgefertigte Prozesse verlängert sich die Zeit bis zur Wiederherstellung.

Vorteile der Einführung dieser bewährten Praxis: Indem Sie nur benachrichtigt werden, wenn tatsächlich eine Aktion erforderlich ist, verhindern Sie, dass wichtige Warnungen in einer Flut unwichtiger Informationen untergehen. Durch einen Prozess, der nur aktionsrelevante Warnungen ausgibt, ermöglichen Sie eine konsistente und schnelle Reaktion auf die Ereignisse in Ihrer Umgebung.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Prozess pro Alarm: Jedem Ereignis, für das Sie eine Warnung auslösen, sollte eine klar definierte Reaktion (Runbook oder Playbook) mit einem speziellen Besitzer (z. B. eine Person, ein Team oder eine Rolle) zugewiesen sein, der für die erfolgreiche Ausführung verantwortlich ist. Die Reaktion kann zwar automatisiert oder von einem anderen Team übernommen werden, aber der Besitzer

trägt die Verantwortung dafür, dass der Prozess die erwarteten Ergebnisse liefert. Diese Prozesse gewährleisten eine effektive und schnelle Reaktion auf Betriebsereignisse und verhindern, dass aktionsrelevante Ereignisse aufgrund weniger wichtiger Benachrichtigungen übersehen werden. Beispielsweise kann eine automatische Skalierung zur Skalierung eines Web-Front-End-Systems verwendet werden, aber das Team des operativen Bereichs könnte dafür verantwortlich sein, dass die Regeln und Limits der automatischen Skalierung den Anforderungen des Workloads entsprechen.

Ressourcen

Verbundene Dokumente:

- Amazon CloudWatch-Funktionen
- Was ist Amazon CloudWatch Events?

Verbundene Videos:

- Erstellen eines Überwachungsplans

OPS10-BP03 Priorisieren von betrieblichen Ereignissen auf Basis der Auswirkung auf das Unternehmen

Stellen Sie sicher, dass bei mehreren Ereignissen, die eine Intervention erfordern, zuerst diejenigen angegangen werden, die für das Unternehmen die größte Tragweite haben. Zu den Auswirkungen können Todesfälle oder Verletzungen, finanzielle Verluste oder Rufschädigung bzw. Vertrauensverlust gehören.

Gängige Antimuster:

- Sie erhalten eine Supportanfrage, in der Sie für einen Benutzer eine Druckerkonfiguration hinzufügen sollen. Während der Arbeit an dem Problem erhalten Sie eine Supportanfrage, dass Ihre Website für den Einzelhandel nicht mehr aufrufbar ist. Nachdem Sie die Druckerkonfiguration für den Benutzer abgeschlossen haben, beginnen Sie mit der Arbeit am Problem mit der Website.
- Sie werden benachrichtigt, dass sowohl Ihre Einzelhandelswebsite als auch Ihr System für die Lohn- und Gehaltsabrechnung ausgefallen sind. Sie wissen nicht, welches Problem Priorität haben sollte.

Vorteile der Einführung dieser bewährten Methode: Durch die Priorisierung von Reaktionen auf Vorfälle mit der größten Auswirkung auf das Unternehmen kommen Sie mit den Auswirkungen leichter zurecht.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Priorisieren von operativen Ereignissen basierend auf den Auswirkungen auf das Geschäft: Wenn mehrere Ereignisse Eingriffe erfordern, stellen Sie sicher, dass diejenigen, die für das Geschäft am wichtigsten sind, zuerst behandelt werden. Zu den Auswirkungen können Todesfälle oder Verletzungen, finanzielle Verluste, Verstöße gegen Vorschriften oder Rufschädigung bzw. Vertrauensverlust gehören.

OPS10-BP04 Definieren von Eskalationspfaden

Definieren Sie Eskalationspfade in Ihren Runbooks und Playbooks und legen Sie auch fest, was eine Eskalation auslöst. Erarbeiten Sie zudem Verfahren für die Eskalation. Weisen Sie jeder Aktion explizit Besitzer zu, um effektive und schnelle Reaktionen auf betriebliche Ereignisse zu gewährleisten.

Legen Sie fest, wann jemand eine Entscheidung treffen muss, bevor eine Aktion durchgeführt wird. Arbeiten Sie mit Entscheidungsträgern zusammen, um diese Entscheidung im Voraus treffen und die Aktion vorab genehmigen zu lassen, damit MTTR nicht auf eine Antwort wartet.

Gängige Antimuster:

- Ihre Einzelhandelswebsite ist nicht mehr aufrufbar. Sie verstehen das Runbook für die Wiederherstellung der Website nicht. Sie rufen Kollegen in der Hoffnung an, dass Ihnen jemand helfen kann.
- Sie erhalten eine Supportanfrage zu einer nicht erreichbaren Anwendung. Sie haben keine Berechtigungen für die Systemverwaltung. Sie wissen nicht, wer die Berechtigungen dafür hat. Sie versuchen, sich an den Besitzer des Systems zu wenden, der die Anfrage gestellt hat, und erhalten keine Antwort. Sie haben keine Kontakte für das System und Ihre Kollegen kennen sich damit nicht aus.

Vorteile der Einführung dieser bewährten Methode: Durch das Definieren von Eskalationen sowie von Auslösern und Verfahren für die Eskalation können Ressourcen einem Vorfall systematisch mit einer für die Auswirkungen geeigneten Menge hinzugefügt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Eskalationspfade definieren: Definieren Sie Eskalationspfade in Ihren Runbooks und Playbooks und legen Sie auch fest, was eine Eskalation auslöst. Erarbeiten Sie zudem Verfahren für die Eskalation. Beispielsweise kann ein Problem von den Support-Technikern eine Stufe höher an leitende Support-Techniker eskaliert werden, wenn das Problem nicht durch Runbooks gelöst werden kann oder wenn eine vordefinierte Zeitspanne verstrichen ist. Ein weiteres Beispiel für einen geeigneten Eskalationspfad bei einem Workload ist die Weiterleitung von den leitenden Support-Technikern an das Entwicklungsteam, wenn die Playbooks keinen Korrekturpfad ermitteln können oder wenn eine vordefinierte Zeitspanne verstrichen ist. Weisen Sie jeder Aktion explizit Besitzer zu, um effektive und schnelle Reaktionen auf betriebliche Ereignisse zu gewährleisten. Eskalationen können auch Dritte beinhalten. Beispiele hierfür sind Anbieter von Netzwerkkonnektivität oder Software. Eskalationen können festgelegte autorisierte Entscheidungsträger für betroffene Systeme einbeziehen.

OPS10-BP05 Aktivieren von Push-Benachrichtigungen

Kommunizieren Sie direkt mit Ihren Benutzern (beispielsweise per E-Mail oder SMS), wenn die von ihnen genutzten Services betroffen sind oder wenn die Services wieder ordnungsgemäß funktionieren, damit die Benutzer entsprechende Maßnahmen ergreifen können.

Gängige Antimuster:

- Ihre Anwendung wird von einem Distributed Denial of Service angegriffen und reagiert seit Tagen nicht mehr. Es gibt keine Fehlermeldung. Sie haben keine E-Mail-Benachrichtigung gesendet. Sie haben keine Textbenachrichtigungen gesendet. Sie haben keine Informationen in den sozialen Medien veröffentlicht. Ihre Kunden sind frustriert und suchen nach anderen Anbietern, die sie tatsächlich unterstützen können.
- Am Montag hatte Ihre Anwendung Probleme nach einem Patch und war mehrere Stunden nicht verfügbar. Am Dienstag hatte Ihre Anwendung Probleme nach einer Codebereitstellung und funktionierte einige Stunden lang nicht zuverlässig. Am Mittwoch hatte Ihre Anwendung Probleme nach einer Codebereitstellung, mit der eine Schwachstelle im Zusammenhang mit dem fehlgeschlagenen Patch geschlossen werden sollte, und war mehrere Stunden nicht verfügbar. Am Donnerstag begannen Ihre frustrierten Kunden mit der Suche nach einem anderen Anbieter, der sie tatsächlich unterstützen kann.

- Ihre Anwendung wird dieses Wochenende aufgrund von Wartungsarbeiten nicht verfügbar sein. Sie informieren Ihre Kunden nicht darüber. Einige Ihrer Kunden hatten Aktivitäten im Zusammenhang mit der Nutzung Ihrer Anwendung geplant. Sie sind sehr frustriert, als Sie feststellen, dass die Anwendung nicht verfügbar ist.

Vorteile der Einführung dieser bewährten Methode: Durch das Definieren von Benachrichtigungen sowie von Auslösern und Verfahren für Benachrichtigungen werden Ihre Kunden informiert und können reagieren, wenn sich Probleme bei Ihrem Workload auf sie auswirken.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Push-Benachrichtigungen aktivieren: Kommunizieren Sie direkt mit Ihren Benutzern (beispielsweise per E-Mail oder SMS), wenn die von ihnen genutzten Services betroffen sind oder wenn die Services wieder ordnungsgemäß funktionieren, damit die Benutzer entsprechende Maßnahmen ergreifen können.
  - Amazon SES-Funktionen
  - Was ist Amazon SES?
  - Einrichten von Amazon SNS-Benachrichtigungen

Ressourcen

Zugehörige Dokumente:

- Amazon SES-Funktionen
- Einrichten von Amazon SNS-Benachrichtigungen
- Was ist Amazon SES?

OPS10-BP06 Bekanntgeben des Status über Dashboards

Stellen Sie Dashboards zur Verfügung, die auf die jeweilige Zielgruppe zugeschnitten sind (z. B. interne technische Teams, Führungskräfte und Kunden), um diese über den aktuellen Betriebsstatus des Unternehmens zu informieren und interessante Metriken bereitzustellen.

Sie können Dashboards mithilfe von Amazon CloudWatch Dashboards auf anpassbaren Homepages in der CloudWatch-Konsole erstellen. Mit Business-Intelligence-Services wie Amazon QuickSight

können Sie interaktive Dashboards für Ihren Workload und den Betriebszustand (z. B. Bestellraten, verbundene Benutzer und Transaktionszeiten) erstellen und veröffentlichen. Erstellen Sie Dashboards, die Ihre Metriken auf System- und Geschäftsebene anzeigen.

Gängige Antimuster:

- Auf Anfrage führen Sie für die Verwaltung einen Bericht über die aktuelle Nutzung Ihrer Anwendung aus.
- Während eines Vorfalls werden Sie alle 20 Minuten von einem besorgten Besitzer eines Systems mit der Frage kontaktiert, ob der Fehler bereits behoben wurde.

Vorteile der Einführung dieser bewährten Methode: Durch das Erstellen von Dashboards aktivieren Sie den Self-Service-Zugriff auf Informationen. Dadurch können Ihre Kunden sich selbst informieren und feststellen, ob sie Maßnahmen ergreifen müssen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Status über Dashboards kommunizieren: Stellen Sie Dashboards zur Verfügung, die auf die jeweilige Zielgruppe zugeschnitten sind (z. B. interne technische Teams, Führungskräfte und Kunden), um diese über den aktuellen Betriebsstatus des Unternehmens zu informieren und interessante Metriken bereitzustellen. Die Bereitstellung einer Self-Service-Option für Statusinformationen reduziert Störungen aufgrund von gezielten Statusanfragen durch das Team des operativen Bereichs. Zu den Beispielen gehören Amazon CloudWatch-Dashboards und AWS Health Dashboard.
  - CloudWatch-Dashboards erstellen und nutzen benutzerdefinierte Metrikansichten

Ressourcen

Zugehörige Dokumente:

- Amazon QuickSight
- CloudWatch-Dashboards erstellen und nutzen benutzerdefinierte Metrikansichten

OPS10-BP07 Automatisieren von Reaktionen auf Ereignisse

Automatisieren Sie Reaktionen auf Ereignisse, um Fehler zu reduzieren, die durch manuelle Prozesse entstehen, und um schnelle und konsistente Reaktionen zu gewährleisten.

Es gibt mehrere Möglichkeiten, um Runbook- und Playbook-Aktionen auf AWS zu automatisieren. Um auf ein Ereignis aufgrund einer Statusänderung in Ihren AWS-Ressourcen oder von Ihren eigenen benutzerdefinierten Ereignissen zu reagieren, sollten Sie CloudWatch Events-Regeln erstellen, um Antworten über CloudWatch-Ziele (zum Beispiel Lambda-Funktionen, Amazon Simple Notification Service-Themen (Amazon SNS), Amazon ECS-Aufgaben und AWS Systems Manager Automation) auszulösen.

Für Reaktionen auf eine Metrik, die einen Schwellenwert für eine Ressource überschreitet (z. B. eine Wartezeit), sollten Sie CloudWatch-Alarme erstellen, um mittels Amazon EC2 oder Auto Scaling-Aktionen eine oder mehrere Aktionen durchzuführen oder um eine Benachrichtigung an ein Amazon SNS-Thema zu senden. Wenn als Reaktion auf einen Alarm benutzerdefinierte Aktionen durchgeführt werden sollen, rufen Sie Lambda per Amazon SNS-Benachrichtigung auf. Veröffentlichen Sie Ereignisbenachrichtigungen und Eskalationsmitteilungen per Amazon SNS, um alle Betroffenen zu informieren.

AWS unterstützt über die AWS-Service-APIs und -SDKs auch Systeme von Drittanbietern. Es gibt eine Reihe von Überwachungs-Tools, die von AWS-Partnern und Dritten zur Verfügung gestellt werden und die Überwachung, Benachrichtigungen und Reaktionen ermöglichen. Dazu gehören zum Beispiel New Relic, Splunk, Loggly, SumoLogic und Datadog.

Für den Fall, dass bei wichtigen Vorgängen automatisierte Verfahren fehlschlagen, sollten Sie manuelle Verfahren bereithalten.

Gängige Antimuster:

- Ein Entwickler überprüft seinen Code. Aufgrund des Ereignisses hätte ein Build gestartet und Tests hätten durchgeführt werden können, aber stattdessen passiert nichts.
- Ihre Anwendung protokolliert einen bestimmten Fehler, bevor sie nicht mehr funktioniert. Das Verfahren zum Neustarten der Anwendung ist bekannt und könnte skriptbasiert ausgeführt werden. Sie können das Protokollereignis verwenden, um ein Skript aufzurufen und die Anwendung neu zu starten. Stattdessen werden Sie am Sonntagmorgen um 3 Uhr geweckt, da Sie als verantwortliche Person für die Behebung von Problemen des Systems Bereitschaftsdienst haben, als der Fehler auftritt.

Vorteile der Einführung dieser bewährten Methode: Dank automatisierter Reaktionen auf Ereignisse reduzieren Sie die Reaktionszeit und begrenzen das Fehlerpotenzial manueller Aktivitäten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Reaktionen auf Ereignisse automatisieren: Automatisieren Sie Reaktionen auf Ereignisse, um Fehler zu reduzieren, die durch manuelle Prozesse entstehen, und um schnelle und konsistente Reaktionen zu gewährleisten.
  - Was ist Amazon CloudWatch Events?
  - Erstellen einer CloudWatch Events-Regel, die nach einem Ereignis ausgelöst wird
  - Erstellen einer CloudWatch Events-Regel, die nach einem AWS-API-Aufruf mithilfe von AWS CloudTrail ausgelöst wird
  - CloudWatch Events-Ereignisbeispiele aus unterstützten Services

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch-Funktionen
- CloudWatch Events-Ereignisbeispiele aus unterstützten Services
- Erstellen einer CloudWatch Events-Regel, die nach einem AWS-API-Aufruf mithilfe von AWS CloudTrail ausgelöst wird
- Erstellen einer CloudWatch Events-Regel, die nach einem Ereignis ausgelöst wird
- Was ist Amazon CloudWatch Events?

Relevante Videos:

- Erstellen eines Überwachungsplans

Zugehörige Beispiele:

Weiterentwicklung

Frage

- OPS 11 Wie können Sie Arbeitsvorgänge weiterentwickeln?

OPS 11 Wie können Sie Arbeitsvorgänge weiterentwickeln?

Kalkulieren Sie Zeit und Ressourcen für kontinuierliche schrittweise Verbesserungen ein, damit sich die Effektivität und Effizienz Ihrer Operationen ständig weiterentwickeln.

Bewährte Methoden

- OPS11-BP01 Implementieren eines Prozesses für die kontinuierliche Verbesserung
- OPS11-BP02 Durchführen von Analysen nach Vorfällen
- OPS11-BP03 Implementieren von Feedbackschleifen
- OPS11-BP04 Wissensmanagement
- OPS11-BP05 Definieren von Verbesserungsfaktoren:
- OPS11-BP06 Prüfen von Erkenntnissen
- OPS11-BP07 Prüfung von Betriebsmetriken
- OPS11-BP08 Dokumentieren und Weitergeben von Erkenntnissen
- OPS11-BP09 Einplanen von Zeit für Verbesserungen

OPS11-BP01 Implementieren eines Prozesses für die kontinuierliche Verbesserung

Bewerten und priorisieren Sie regelmäßig Verbesserungsmöglichkeiten, um die Maßnahmen dort zu intensivieren, wo sie den größten Nutzen bringen.

Gängige Antimuster:

- Sie haben die erforderlichen Verfahren zum Erstellen einer Entwicklungs- oder Testumgebung dokumentiert. Sie könnten den Prozess mit CloudFormation automatisieren, nutzen dafür stattdessen aber manuell die Konsole.
- Ihre Tests zeigen, dass der Großteil der CPU-Auslastung innerhalb Ihrer Anwendung von einer kleinen Gruppe ineffizienter Funktionen verursacht wird. Sie könnten sich darauf konzentrieren, diese zu verbessern und Ihre Kosten zu senken, aber Sie wurden beauftragt, eine neue Funktion für die Benutzerfreundlichkeit zu erstellen.

Vorteile der Einführung dieser bewährten Methode: Kontinuierliche Verbesserung bietet einen Mechanismus zur regelmäßigen Bewertung von Verbesserungsmöglichkeiten, Priorisierung von Geschäftschancen und Intensivierung von Maßnahmen, wo diese den größten Nutzen bringen können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Prozess für die kontinuierliche Verbesserung definieren: Bewerten und priorisieren Sie regelmäßig Verbesserungsmöglichkeiten, um die Maßnahmen dort zu intensivieren, wo sie den größten Nutzen bringen. Implementieren Sie Änderungen, die zu Verbesserungen führen sollen, und beurteilen Sie deren Ergebnisse. Wenn die Ergebnisse die Ziele nicht erfüllen und die Verbesserung immer noch Priorität hat, wiederholen Sie den Versuch mit alternativen Vorgehensweisen. In Ihren betrieblichen Prozessen sollten auch Zeit und Ressourcen genutzt werden, um kontinuierliche, schrittweise Verbesserungen zu ermöglichen.

OPS11-BP02 Durchführen von Analysen nach Vorfällen

Überprüfen Sie die Ereignisse mit Auswirkungen auf Kunden und bestimmen Sie die beitragenden Faktoren und Präventivmaßnahmen. Entwickeln Sie anhand dieser Informationen Abhilfemaßnahmen, um Wiederholungen einzuschränken oder zu verhindern. Entwickeln Sie Verfahren für schnelle und effektive Reaktionen. Informieren Sie nach Bedarf auf zielgruppengerechte Weise über beitragende Faktoren und Korrekturmaßnahmen.

Gängige Antimuster:

- Sie verwalten einen Anwendungsserver. Ungefähr alle 23 Stunden und 55 Minuten werden alle Ihre aktiven Sitzungen beendet. Sie haben versucht, festzustellen, wo der Fehler auf Ihrem Anwendungsserver liegt. Sie vermuten, dass es sich um ein Netzwerkproblem handeln könnte, das Netzwerkteam zeigt sich jedoch unkooperativ, da es für Ihr Anliegen zu beschäftigt ist. Sie haben keinen vordefinierten Prozess, den Sie befolgen könnten, um Support zu erhalten und die nötigen Informationen zu sammeln, um dem Problem auf den Grund zu gehen.
- Bei Ihrem Workload kam es zu Datenverlust. Dies ist das erste Mal, dass dieses Problem aufgetreten ist, und die Ursache ist nicht klar. Sie entscheiden, dass es nicht wichtig ist, da Sie die Daten wiederherstellen können. Datenverluste beginnen mit größerer Häufigkeit aufzutreten und wirken sich auf Ihre Kunden aus. Dadurch steigt auch der betriebliche Aufwand, wenn Sie die fehlenden Daten wiederherstellen.

Vorteile der Einführung dieser bewährten Methode: Durch vordefinierte Prozesse zur Bestimmung der Komponenten, Bedingungen, Maßnahmen und Ereignisse, die zu einem Vorfall beigetragen haben, können Sie Verbesserungsmöglichkeiten ermitteln.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden eines Prozesses zur Ermittlung beitragender Faktoren: Überprüfen Sie alle Vorfälle, die sich auf Kunden auswirken. Erarbeiten Sie ein Verfahren, um die beitragenden Faktoren eines Vorfalls zu ermitteln und zu dokumentieren. Damit können Sie Abhilfemaßnahmen entwickeln, um ein erneutes Auftreten einzudämmen oder gänzlich zu verhindern, und Verfahren für eine rasche und wirksame Reaktion erstellen. Kommunizieren Sie die Ursache, soweit erforderlich, auf die jeweiligen Zielgruppen zugeschnitten.

OPS11-BP03 Implementieren von Feedbackschleifen

Feedbackschleifen bieten umsetzbare Einblicke zur Unterstützung der Entscheidungsfindung. Integrieren Sie Feedbackschleifen in Ihre Verfahren und Workloads. Damit können Sie Probleme und Bereiche identifizieren, für die Verbesserungen erforderlich sind. Diese validieren auch Investitionen für Verbesserungen. Diese Feedbackschleifen sind die Grundlage für die kontinuierliche Verbesserung Ihres Workloads.

Feedbackschleifen können in zwei Kategorien unterteilt werden: Sofortiges Feedback und nachträgliche Analyse. Sofortiges Feedback wird durch Prüfung der Leistung und der Ergebnisse betrieblicher Aktivitäten eingeholt. Dieses Feedback kommt von Teammitgliedern, Kunden oder der automatisierten Ausgabe der Aktivität. Sofortiges Feedback kommt von Dingen wie A/B-Tests und der Auslieferung neuer Funktionen und ist für das „Schnell scheitern“-Konzept von entscheidender Bedeutung.

Nachträgliche Analysen werden regelmäßig durchgeführt, um Feedback aus der Überprüfung betrieblicher Ergebnisse und Metriken in der Vergangenheit zu erhalten. Dies geschieht am Ende einer Phase, in regelmäßigem Rhythmus oder nach größeren Releases oder Veranstaltungen. Diese Art von Feedbackschleife validiert Investitionen in Betriebsabläufe oder Ihren Workload. Dies hilft Ihnen beim Messen des Erfolgs und bei der Validierung Ihrer Strategie.

Gewünschtes Ergebnis: Sie nutzen sofortiges Feedback und nachträgliche Analysen für weitere Verbesserungen. Es gibt einen Mechanismus zur Erfassung des Feedbacks von Benutzern und Teammitgliedern. Nachträgliche Analysen identifizieren Trends, die Verbesserungen unterstützen können.

Typische Anti-Muster:

- Sie starten einige Funktionen, haben aber keine Möglichkeit, Feedback von den Kunden dazu zu erhalten.
- Nach einer Investition in verbesserte Betriebsabläufe führen Sie keine nachträgliche Analyse für deren Validierung durch.
- Sie holen das Feedback von Kunden ein, überprüfen dies jedoch nicht regelmäßig.
- Feedbackschleifen führen zu vorgeschlagenen Maßnahmen, werden jedoch nicht in den Softwareentwicklungsprozess einbezogen.
- Kunden erhalten kein Feedback zu Verbesserungen, die sie vorgeschlagen haben.

Vorteile der Nutzung dieser bewährten Methode:

- Sie können vom Kunden aus rückwärts arbeiten, um neue Funktionen zu unterstützen.
- Ihre Organisationskultur kann schneller auf Änderungen reagieren.
- Trends dienen zur Identifizierung von Verbesserungsmöglichkeiten.
- Nachträgliche Analysen validieren in Ihre Workloads und Betriebsabläufe getätigte Investitionen.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Die Implementierung dieser bewährten Methode bedeutet, dass Sie sofortiges Feedback und nachträgliche Analysen verwenden. Diese Feedbackschleifen erleichtern Verbesserungen. Es gibt zahlreiche Mechanismen für sofortiges Feedback, z. B. Umfragen, Kundenbefragungen oder Feedbackformulare. Ihre Organisation nutzt nachträgliche Analysen auch, um Möglichkeiten für Verbesserungen zu identifizieren und Initiativen zu validieren.

Kundenbeispiel

AnyCompany Retail hat ein Webformular erstellt, über das Kunden Feedback abgeben oder Probleme melden können. Bei der wöchentlichen Scrum-Sitzung evaluiert das Softwareentwicklungsteam das Benutzerfeedback. Das Feedback wird regelmäßig genutzt, um die Weiterentwicklung der Plattform zu steuern. Am Ende jeder Etappe wird eine nachträgliche Analyse durchgeführt, um Punkte zu identifizieren, bei denen Verbesserungsbedarf besteht.

Implementierungsschritte

1. Sofortiges Feedback

- Sie benötigen einen Mechanismus für den Erhalt von Feedback von Kunden und Teammitgliedern. Ihre betrieblichen Aktivitäten können auch so konfiguriert werden, dass Sie automatisiertes Feedback erhalten.
- Ihre Organisation benötigt einen Prozess zur Prüfung dieses Feedbacks, zum Feststellen der Verbesserungsbereiche und zur Planung der Verbesserungen.
- Das Feedback muss in Ihren Softwareentwicklungsprozess integriert werden.
- Wenn Sie Verbesserungen durchführen, informieren Sie die Personen, die dazu Feedback gegeben haben.
  - Sie können AWS Systems Manager OpsCenter verwenden, um diese Verbesserungen als OpsItems nachzuverfolgen.

2. Nachträgliche Analyse
   - Führen Sie nachträgliche Analysen am Ende eines Entwicklungszyklus, in regelmäßigen Abständen oder nach einem größeren Release durch.
   - Laden Sie an dem Workload beteiligte Personen zu einer Nachbesprechung ein.
   - Erstellen Sie auf einem Whiteboard oder in einem Spreadsheet drei Spalten: Beenden, Starten und Beibehalten.
     - Beenden gilt für alles, mit dem Ihr Team aufhören soll.
     - Starten gilt für Ideen, die ab sofort umgesetzt werden sollen.
     - Beibehalten gilt für Elemente, die weiterhin durchgeführt werden sollen.
   - Holen Sie das Feedback aller anwesenden beteiligten Personen ein.
   - Priorisieren Sie das Feedback. Weisen Sie allen „Starten“- oder „Beibehalten“-Elementen Aktionen und Beteiligte zu.
   - Fügen Sie die Aktionen Ihrem Softwareentwicklungsprozess hinzu und halten Sie die Beteiligten bei Ihren Verbesserungen über den Status auf dem Laufenden.

Aufwand für den Implementierungsplan: Mittel. Zur Implementierung dieser bewährten Methode benötigen Sie ein Verfahren zum Einholen und zur Analyse sofortigen Feedbacks. Dazu müssen Sie auch einen Prozess für die nachträgliche Analyse einrichten.

Ressourcen

Zugehörige bewährte Methoden:

- OPS01-BP01 Bedürfnisse externer Kunden bewerten: Feedbackschleifen sind ein Mechanismus zum Ermitteln der Anforderungen externer Kunden.
- OPS01-BP02 Bedürfnisse interner Kunden bewerten: Interne Beteiligte können Feedbackschleifen nutzen, um Bedürfnisse und Anforderungen zu kommunizieren.
- OPS11-BP02 Durchführen von Analysen nach Vorfällen: Analysen nach einem Vorfall sind eine wichtige Form nachträglicher Analyse nach Vorfällen.
- OPS11-BP07 Prüfung von Betriebsmetriken: Durch die Prüfung betrieblicher Metriken können Sie Trends und Bereiche für Verbesserungen identifizieren.

Zugehörige Dokumente:

- 7 Fehler, die Sie bei der Einrichtung eines CCOE vermeiden sollten
- Atlassian Team Playbook - Retrospectives
- E-Mail-Definitionen: Feedbackschleifen
- Einrichten von Feedbackschleifen mit der AWS Well-Architected Framework Review
- IBM Garage Methodology – Nachträgliche Analysen
- Investopedia – The PDCS Cycle
- Maximizing Developer Effectiveness von Tim Cochran
- Operations Readiness Reviews (ORR) Whitepaper - Iteration
- TIL CSI - Continual Service Improvement
- Toyota und E-Commerce: Lean bei Amazon

Zugehörige Videos:

- Building Effective Customer Feedback Loops (Aufbau effektiver Kundenfeedbackschleifen)

Zugehörige Beispiele:

- Astuto - Open-Source-Tool für Kundenfeedback
- AWS-Lösungen – QnABot auf AWS
- Fider – Eine Plattform zur Organisation von Kundenfeedback

Zugehörige Services:

- AWS Systems Manager OpsCenter

OPS11-BP04 Wissensmanagement

Es gibt Mechanismen, mit denen Ihre Teammitglieder die gesuchten Informationen rechtzeitig erkennen, darauf zugreifen und feststellen können, dass sie aktuell und vollständig sind. Mechanismen sind vorhanden, um benötigte Inhalte, zu aktualisierende Inhalte und zu archivierende Inhalte zu identifizieren, damit sie nicht mehr referenziert werden.

Gängige Antimuster:

- Ein einzelner frustrierter Kunde eröffnet eine Supportanfrage und fordert eine neue Produktfunktion für ein wahrgenommenes Problem an. Sie wird zur Liste der Verbesserungen mit Priorität hinzugefügt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Wissensmanagement: Es gibt Mechanismen, mit denen Ihre Teammitglieder die gesuchten Informationen rechtzeitig erkennen, darauf zugreifen und feststellen können, dass sie aktuell und vollständig sind. Mechanismen müssen vorhanden sein, um benötigte Inhalte, zu aktualisierende Inhalte und zu archivierende Inhalte zu identifizieren, damit sie nicht mehr referenziert werden.

OPS11-BP05 Definieren von Verbesserungsfaktoren:

Ermitteln Sie Verbesserungsfaktoren, um das Potenzial besser bewerten und priorisieren zu können.

In AWS können Sie die Protokolle all Ihrer betrieblichen Aktivitäten, Workloads und Infrastruktur zusammenstellen, um einen detaillierten Aktivitätsverlauf zu erstellen. Anschließend können Sie AWS-Tools verwenden, um Ihren Betrieb und den Workload-Zustand im Laufe der Zeit zu analysieren (z. B. Trends zu identifizieren, Ereignisse und Aktivitäten mit Ergebnissen zu korrelieren und zwischen Umgebungen und systemübergreifend zu vergleichen), um Verbesserungsmöglichkeiten basierend auf den auslösenden Faktoren aufzudecken.

Sie sollten API-Aktivitäten mithilfe von CloudTrail verfolgen (per AWS Management Console, Befehlszeilenschnittstelle, SDKs und APIs), um immer zu wissen, was sich bei Ihren Konten tut. Verfolgen Sie Bereitstellungsaktivitäten der AWS Developer Tools mit CloudTrail und CloudWatch

nach. Dadurch wird Ihren CloudWatch Logs-Protokolldaten ein detaillierter Aktivitätsverlauf Ihrer Bereitstellungen und deren Ergebnisse hinzugefügt.

Exportieren Sie Ihre Protokolldaten zur langfristigen Speicherung inAmazon S3 . Mit AWS Gluekönnen Sie Ihre Protokolldaten in Amazon S3 für Analysen erkunden und vorbereiten. Verwendung Sie Amazon Athenadurch die native Integration mit AWS Glue, um Ihre Protokolldaten zu analysieren. Verwenden Sie ein Business Intelligence-Tool wie Amazon QuickSight , um Ihre Daten zu visualisieren, zu untersuchen und zu analysieren.

Gängige Antimuster:

- Sie haben ein Skript, das zwar funktioniert, aber optisch nicht viel hermacht. Sie investieren Zeit in das Umschreiben. Es ist jetzt ein wahres Kunstwerk.
- Ihr Start-up versucht, weitere Finanzierung von einem Risikokapitalgeber zu erhalten. Dieser möchte, dass Sie die Compliance mit PCI DSS nachweisen. Sie möchten diesem Wunsch entsprechen und Ihre Compliance dokumentieren. Dabei übersehen Sie jedoch ein Lieferdatum für einen Kunden und verlieren diesen. Vom Grundgedanken her war das nicht verkehrt, Sie fragen sich allerdings, ob Sie richtig gehandelt haben.

Vorteile der Einführung dieser bewährten Methode: Durch die Bestimmung der Kriterien, die Sie für die Verbesserung verwenden möchten, können Sie die Auswirkungen ereignisbasierter Motivationen oder emotionaler Investitionen minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Kenntnis der Verbesserungsfaktoren: Sie sollten ein System nur dann ändern, wenn das gewünschte Ergebnis auch unterstützt wird.
  - Gewünschte Fähigkeiten: Prüfen Sie bei der Bewertung von Verbesserungsmöglichkeiten die gewünschten Funktionen und Fähigkeiten.
    - Neuerungen bei AWS
  - Nicht akzeptable Probleme: Prüfen Sie bei der Bewertung von Verbesserungsmöglichkeiten nicht akzeptable Probleme, Fehler und Schwachstellen.
    - Aktuelle AWS-Sicherheitsmitteilungen
    - AWS Trusted Advisor

- Compliance-Anforderungen: Prüfen Sie bei der Bewertung von Verbesserungsmöglichkeiten, welche Updates und Änderungen erforderlich sind, um Vorschriften bzw. Richtlinien einzuhalten oder weiterhin den Support eines Drittanbieters nutzen zu können.
  - AWS-Compliance
  - AWS-Compliance-Programme
  - Aktuelle Neuigkeiten zur AWS-Compliance

Ressourcen

Zugehörige Dokumente:

- Amazon Athena
- Amazon QuickSight
- AWS-Compliance
- Aktuelle Neuigkeiten zur AWS-Compliance
- AWS-Compliance-Programme
- AWS Glue
- Aktuelle AWS-Sicherheitsmitteilungen
- AWS Trusted Advisor
- Exportieren Sie Ihre Protokolldaten zur langfristigen Speicherung inAmazon S3
- Neuerungen bei AWS

OPS11-BP06 Prüfen von Erkenntnissen

Überprüfen Sie Ihre Analyseergebnisse und Reaktionen mit fachbereichsübergreifenden Teams und Geschäftsverantwortlichen. Schaffen Sie mithilfe dieser Prüfungen ein allgemeines Verständnis, ermitteln Sie weitere Auswirkungen und legen Sie einen Maßnahmenkatalog fest. Passen Sie die Reaktionen bei Bedarf an.

Gängige Antimuster:

- Sie sehen, dass die CPU-Auslastung auf einem System 95 % beträgt, und möchten mit Priorität eine Möglichkeit finden, die Auslastung dieses Systems zu reduzieren. Die beste Vorgehensweise ist die Skalierung nach oben. Das System wird als Transcoder verwendet und so skaliert, dass es jederzeit mit 95 % CPU-Auslastung ausgeführt wird. Der Besitzer des Systems hätte Ihnen die

Situation erklären können, wenn Sie sich an ihn gewandt hätten. Sie haben Ihre Zeit nicht sinnvoll genutzt.

- Der Besitzer eines Systems behauptet, dass sein System geschäftskritisch sei. Das System wird nicht in einer Umgebung betrieben, die für hohe Sicherheit ausgelegt ist. Zur Verbesserung der Sicherheit implementieren Sie zusätzliche Erkennungs- und Präventivfunktionen, die für geschäftskritische Systeme erforderlich sind. Sie benachrichtigen den Besitzer des Systems, dass die Arbeit abgeschlossen ist und ihm die zusätzlichen Ressourcen in Rechnung gestellt werden. In der Diskussion nach dieser Benachrichtigung erfährt der Besitzer des Systems, dass es eine offizielle Definition für geschäftskritische Systeme gibt, die sein System nicht erfüllt.

Vorteile der Einführung dieser bewährten Methode: Durch die Prüfung von Erkenntnissen zusammen mit Geschäftsinhabern und Fachexperten können Sie ein gemeinsames Verständnis aufbauen und effektiver für Verbesserungen sorgen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Prüfen von Erkenntnissen: Wenden Sie sich an die Geschäftsinhaber und Fachexperten, um sicherzustellen, dass die Bedeutung der von Ihnen gesammelten Daten allgemein verstanden und vereinbart ist. Ermitteln Sie zusätzliche Bedenken, potenzielle Auswirkungen und bestimmen Sie eine Vorgehensweise.

OPS11-BP07 Prüfung von Betriebsmetriken

Führen Sie regelmäßig teamübergreifend mit Teilnehmern aus verschiedenen Unternehmensbereichen nachträgliche Analysen der operationsspezifischen Metriken durch. Ermitteln Sie mithilfe dieser Prüfungen Verbesserungspotenziale sowie mögliche Maßnahmen und teilen Sie diese Erkenntnisse auch anderen mit.

Berücksichtigen Sie bei Ihrer Suche nach Verbesserungsmöglichkeiten all Ihre Umgebungen (z. B. Entwicklungs-, Test- und Produktionsumgebung).

Gängige Antimuster:

- Eine wichtige Verkaufsaktion wurde durch Ihr Wartungsfenster unterbrochen. Das Unternehmen weiß weiterhin nicht, dass es ein Standard-Wartungsfenster gibt, das verzögert werden könnte, wenn sich andere wichtige Ereignisse auf das Geschäft auswirken.

- Sie erlitten einen längeren Ausfall, weil Sie eine fehlerhafte Bibliothek verwendet hatten, die häufig in Ihrem Unternehmen genutzt wird. Seitdem sind Sie zu einer zuverlässigen Bibliothek migriert. Die anderen Teams in Ihrem Unternehmen wissen nicht, dass diese Gefahr besteht. Wenn Sie sich regelmäßig treffen und diesen Vorfall besprechen würden, wüssten sie über das Risiko Bescheid.
- Die Leistung Ihres Transcoders ist stetig gesunken und beeinträchtigt das Medienteam. Die Leistung ist noch nicht ganz schlimm. Sie haben aber keine Gelegenheit, von dem Problem zu erfahren, bis es so schlimm ist, dass daraus ein Vorfall entsteht. Würden Sie Ihre Betriebsmetriken gemeinsam mit dem Medienteam überprüfen, bestünde die Möglichkeit, die Metriken zu ändern, den vom Team spürbaren Leistungseinbruch zu erkennen und das Problem zu beheben.
- Sie prüfen nicht, wie zufrieden Kunden mit der Erfüllung Ihrer SLAs sind. Sie laufen Gefahr, die mit Kunden vereinbarten SLAs nicht zu erfüllen. Es gibt Geldstrafen im Zusammenhang mit der Nichteinhaltung von mit Kunden vereinbarten SLAs. Würden Sie die Metriken für diese SLAs bei regelmäßigen Treffen überprüfen, hätten Sie die Gelegenheit, das Problem zu erkennen und zu beheben.

Vorteile der Einführung dieser bewährten Methode: Durch regelmäßige Besprechungen zur Überprüfung von Betriebsmetriken, Ereignissen und Vorfällen schaffen Sie ein gemeinsames teamübergreifendes Verständnis, teilen gewonnene Erkenntnisse mit und können Verbesserungen priorisieren und gezielt in Angriff nehmen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Prüfungen von Betriebsmetriken: Führen Sie regelmäßig teamübergreifend mit Teilnehmern aus verschiedenen Unternehmensbereichen nachträgliche Analysen der operationsspezifischen Metriken durch. Binden Sie alle Beteiligten, einschließlich der Teams aus den Bereichen Betriebswirtschaft, Entwicklung und Operationen, ein, indem Sie Ihre Erkenntnisse aus dem sofortigen Feedback und der nachträglichen Analyse und gewonnene Erkenntnisse austauschen. Machen Sie sich deren Informationen zunutze, um Verbesserungspotenziale und mögliche Maßnahmen ausfindig zu machen.
  - Amazon CloudWatch
  - Verwenden von Amazon CloudWatch-Metriken
  - Veröffentlichen von benutzerdefinierten Metriken
  - Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch
- Referenzinformationen zu Metriken und Dimensionen von Amazon CloudWatch
- Veröffentlichen von benutzerdefinierten Metriken
- Verwenden von Amazon CloudWatch-Metriken

OPS11-BP08 Dokumentieren und Weitergeben von Erkenntnissen

Dokumentieren Sie die Erkenntnisse aus den betrieblichen Aktivitäten und geben Sie diese weiter, damit Sie sie sowohl intern als auch teamübergreifend nutzen können.

Die Erkenntnisse Ihres Teams sollten Sie an andere weitergeben in Ihrem Unternehmen, damit alle davon profitieren. Informationen und Ressourcen sollten Sie weitergeben, um vermeidbare Fehler zu verhindern und Entwicklungsbemühungen zu unterstützen. Dies wird es Ihnen ermöglichen, sich auf die Bereitstellung gewünschter Funktionen zu konzentrieren.

Definieren Sie mithilfe von AWS Identity and Access Management (IAM) Berechtigungen, die den gesteuerten Zugriff auf die Ressourcen ermöglichen, die Sie innerhalb von Konten und kontenübergreifend freigeben möchten. Anschließend sollten Sie versionsgesteuerte AWS CodeCommit verwenden, um Anwendungsbibliotheken, skriptbasierte Verfahren, Verfahrens- und andere Systemdokumentationen freizugeben. Geben Sie Ihre Computing-Standards für andere frei, indem Sie den Zugriff auf Ihre AMIs freigeben und die Verwendung Ihrer Lambda-Funktionen kontenübergreifend erlauben. Auch Ihre Infrastrukturstandards sollten Sie als AWS CloudFormation-Vorlagen freigeben.

Über die AWS-APIs und -SDKs können Sie externe und von Drittanbietern stammende Tools und Repositorys integrieren (z. B. GitHub, BitBucket und SourceForge). Achten Sie bei der Freigabe Ihrer Erkenntnisse und Entwicklungen sorgfältig darauf, Berechtigungen so zu strukturieren, dass die Integrität freigegebener Repositorys nicht gefährdet wird.

Gängige Antimuster:

- Sie erlitten einen längeren Ausfall, weil Sie eine fehlerhafte Bibliothek verwendet hatten, die häufig in Ihrem Unternehmen genutzt wird. Seitdem sind Sie zu einer zuverlässigen Bibliothek migriert. Die anderen Teams in Ihrem Unternehmen wissen nicht, dass diese Gefahr besteht. Würden Sie

Ihre Erfahrungen mit dieser Bibliothek dokumentieren und weitergeben, wüssten die anderen Teams über das Risiko Bescheid.

- Sie haben einen Grenzfall in einem intern gemeinsam genutzten Microservice ermittelt, der dazu führt, dass Sitzungen unterbrochen werden. Sie rufen den Service jetzt anders auf, um diesen Grenzfall zu vermeiden. Die anderen Teams in Ihrem Unternehmen wissen nicht, dass diese Gefahr besteht. Würden Sie Ihre Erfahrungen mit dieser Bibliothek dokumentieren und weitergeben, wüssten die anderen Teams über das Risiko Bescheid.
- Sie haben eine Möglichkeit gefunden, die Anforderungen an die CPU-Auslastung eines Ihrer Microservices deutlich zu reduzieren. Sie wissen nicht, ob andere Teams auch von diesem Verfahren profitieren könnten. Würden Sie Ihre Erfahrungen mit dieser Bibliothek dokumentieren und weitergeben, könnten auch andere davon profitieren.

Vorteile der Einführung dieser bewährten Methode: Gemeinsame Erkenntnisse unterstützen Verbesserungen und ermöglichen, erfahrungsbasierte Vorteile zu maximieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Dokumentieren und Weitergeben von Erkenntnissen: Implementieren Sie Verfahren zur Dokumentation der aus der Durchführung von betrieblichen Aktivitäten und nachträglichen Analysen gewonnenen Erkenntnisse, damit auch andere Teams davon profitieren.
  - Weitergeben von Erkenntnissen: Nutzen Sie Verfahren für den teamübergreifenden Austausch gewonnener Erkenntnisse und zugehöriger Nebenprodukte. Veröffentlichen Sie beispielsweise aktualisierte Verfahren, Richtlinien, Governance und Best Practices in einem allgemein zugänglichen Wiki oder teilen Sie Skripte, Code und Bibliotheken über ein gemeinsames Repository.
    - Delegieren des Zugriffs auf Ihre AWS-Umgebung
    - Freigeben eines AWS CodeCommit-Repositorys
    - Unkomplizierte Autorisierung von AWS Lambda-Funktionen
    - Freigeben eines AMI mit bestimmten AWS-Konten
    - Schnelles Freigeben von Vorlagen mit einer AWS CloudFormation-Designer-URL
    - Verwenden von AWS Lambda mit Amazon SNS

Ressourcen

Zugehörige Dokumente:

- Unkomplizierte Autorisierung von AWS Lambda-Funktionen
- Freigeben eines AWS CodeCommit-Repositorys
- Freigeben eines AMI mit bestimmten AWS-Konten
- Schnelles Freigeben von Vorlagen mit einer AWS CloudFormation-Designer-URL
- Verwenden von AWS Lambda mit Amazon SNS

Relevante Videos:

- Delegieren des Zugriffs auf Ihre AWS-Umgebung

OPS11-BP09 Einplanen von Zeit für Verbesserungen

Reservieren Sie Zeit und Ressourcen innerhalb Ihrer Prozesse, um kontinuierliche, schrittweise Verbesserungen zu ermöglichen.

In AWS können Sie temporäre Duplikate von Umgebungen erstellen. Das senkt die Risiken, Mühen und Kosten, die mit dem Experimentieren und Testen verbunden sind. Diese duplizierten Umgebungen können Sie nutzen, um die aus Ihren Analysen gezogenen Rückschlüsse zu testen, Verbesserungen zu entwickeln und geplante Verbesserungen zu testen.

Gängige Antimuster:

- Es besteht ein bekanntes Leistungsproblem auf Ihrem Anwendungsserver. Es wird im Backlog hinter jeder geplanten Funktionsimplementierung priorisiert. Bleibt die Rate der hinzugefügten geplanten Funktionen konstant, wird das Leistungsproblem niemals behoben.
- Um kontinuierliche Verbesserungen zu unterstützen, genehmigen Sie den Administratoren und Entwicklern, dass sie ihre Überstunden zur Auswahl und Implementierung von Verbesserungen nutzen können. Es werden niemals Verbesserungen vorgenommen.

Vorteile der Einführung dieser bewährten Methode: Indem Sie Zeit und Ressourcen innerhalb Ihrer Prozesse reservieren, ermöglichen Sie kontinuierliche, schrittweise Verbesserungen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Zeit für Verbesserungen einplanen: Reservieren Sie Zeit und Ressourcen innerhalb Ihrer Prozesse, um kontinuierliche, schrittweise Verbesserungen zu ermöglichen. Implementieren Sie Änderungen, die zu Verbesserungen führen sollen, und beurteilen Sie deren Ergebnisse. Wenn die Ergebnisse die Ziele nicht erfüllen und die Verbesserung immer noch Priorität hat, versuchen Sie alternative Vorgehensweisen.

Sicherheit

Themen

Sicherheitsgrundlagen

Frage

SICH 1 Wie können Sie Ihren Workload sicher betreiben?

Um Ihre Workload sicher zu betreiben, müssen Sie auf jeden Sicherheitsbereich übergreifende bewährte Methoden anwenden. Nutzen Sie Anforderungen und Prozesse, die Sie in Operational Excellence definiert haben, auf Organisations- und Workload-Ebene, und wenden Sie sie auf alle Bereiche an. Bleiben Sie auf dem Laufenden mit AWS- und Branchenempfehlungen sowie Bedrohungsinformationen, um Ihr Bedrohungsmodell und Ihre Kontrollziele weiterzuentwickeln. Durch die Automatisierung von Sicherheitsprozessen, Tests und Validierung können Sie Ihre Sicherheitsvorgänge skalieren.

Bewährte Methoden

- SEC01-BP02 Sicheres AWS-Konto
- SEC01-BP03 Identifizieren und Validieren von Kontrollzielen
- SEC01-BP04 Sicherstellen der Aktualität von Informationen zu Sicherheitsbedrohungen
- SEC01-BP05 Aktuelle Informationen dank Sicherheitsempfehlungen
- SEC01-BP06 Automatisieren von Tests und Validierung von Sicherheitskontrollen in Pipelines
- SEC01-BP07 Identifizieren und Priorisieren von Risiken anhand eines Bedrohungsmodells
- SEC01-BP08 Regelmäßiges Bewerten und Implementieren neuer Sicherheitsservices und -funktionen

SEC01-BP01 Trennen von Workloads mithilfe von Konten

Richten Sie zunächst Ihr Augenmerk auf die Sicherheit und Infrastruktur, damit Ihr Unternehmen bei zunehmenden Workloads bewährte Schutzvorkehrungen einrichten kann. Dieser Ansatz bietet Grenzen und Kontrollen zwischen Workloads. Die Trennung auf Kontoebene wird dringend empfohlen, um Produktionsumgebungen von Entwicklungs- und Testumgebungen zu isolieren oder eine starke logische Grenze zwischen Workloads bereitzustellen, die Daten unterschiedlicher Vertraulichkeitsstufen verarbeiten, wie durch externe Compliance-Anforderungen definiert (z. B. PCI-DSS oder HIPAA), und Workloads, die dies nicht tun.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden Sie AWS Organizations: Mit AWS Organizations können Sie zentral eine richtlinienbasierten Verwaltung für mehrere AWS-Konten erzwingen.
  - Erste Schritte mit AWS Organizations
  - Verwenden von Service-Kontrollrichtlinien zum Festlegen eines kontenübergreifenden Integritätsschutzes für Berechtigungen in AWS Organizations
- Erwägen Sie die den Einsatz von AWS Control Tower: AWS Control Tower bietet eine einfache Möglichkeit, eine neue, sichere, kontenübergreifende AWS-Umgebung basierend auf bewährten Methoden einzurichten und zu verwalten.
  - AWS Control Tower

Ressourcen

Zugehörige Dokumente:

- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Sicherheitsberichte
- Richtlinien zur AWS-Sicherheitsprüfung

Relevante Videos:

- Managing Multi-Account AWS Environments Using AWS Organizations (Verwalten von AWS-Umgebungen mit mehreren Konten mithilfe von AWS Organizations)
- Security Best Practices the Well-Architected Way
- Verwenden von AWS Control Tower zur Steuerung von AWS-Umgebungen mit mehreren Konten

SEC01-BP02 Sicheres AWS-Konto

Es gibt eine Reihe von Aspekten für die Sicherung Ihrer AWS-Konten-Konten, einschließlich der Sicherung und nicht der Verwendung des Stammbenutzers und der Aktualisierung der Kontaktinformationen. Mit AWS Organizations können Sie Ihre Konten zentral verwalten, während Sie Ihre Workloads in AWS vergrößern und skalieren. AWS Organizations unterstützt Sie bei der Verwaltung von Konten, der Einrichtung von Kontrollen und der Konfiguration von Services für Ihre Konten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden Sie AWS Organizations: Mit AWS Organizations können Sie zentral eine richtlinienbasierten Verwaltung für mehrere AWS-Konten erzwingen.
  - Erste Schritte mit AWS Organizations
  - Verwenden von Service-Kontrollrichtlinien zum Festlegen eines kontenübergreifenden Integritätsschutzes für Berechtigungen in AWS Organizations
- Grenzen Sie die Verwendung des AWS-Root-Benutzers ein: Verwenden Sie den Root-Benutzer nur, um Aufgaben auszuführen, für die dies explizit erforderlich ist.
  - AWS Tasks That Require AWS Account Root User Credentials (AWS-Aufgaben, für die Anmeldeinformationen für das AWS-Root-Benutzerkonto erforderlich sind)
- Aktivieren Sie die Multi-Faktor-Authentifizierung (MFA) für den Root-Benutzer: Aktivieren Sie MFA für den AWS-Konto-Root-Benutzer, wenn Root-Benutzer nicht durch AWS Organizations verwaltet werden.

  - Root-Benutzer
- Ändern Sie das Passwort des Root-Benutzers regelmäßig: Das Ändern des Root-Benutzerpassworts reduziert das Risiko der Verwendung eines gespeicherten Passworts. Dies ist besonders wichtig, wenn Sie AWS Organizations nicht verwenden und jeder physischen Zugriff hat.
  - Changing the AWS-Konto root user password (Ändern des Root-Benutzerpassworts für das AWS-Konto)
- Aktivieren Sie Benachrichtigungen, wenn das AWS-Konto-Root-Benutzerkonto verwendet wird: Die automatische Benachrichtigung reduziert das Risiko.
  - Empfang von Benachrichtigungen, wenn die Root-Zugriffschlüssel Ihres AWS-Konto-Kontos verwendet werden
- Schränken Sie den Zugriff auf neu hinzugefügte Regionen ein: Für neue AWS-Regionen werden IAM-Ressourcen, z. B. Benutzer und Rollen, nur an die von Ihnen aktivierten Regionen weitergegeben.
  - Festlegen von Berechtigungen zum Aktivieren von Konten für neue AWS-Regionen
- Erwägen Sie die Verwendung von AWS CloudFormation StackSets: CloudFormation StackSets können verwendet werden, um Ressourcen wie IAM-Richtlinien, -Rollen und -Gruppen aus einer genehmigten Vorlage in verschiedenen AWS-Konten bereitzustellen.
  - Verwenden von CloudFormation StackSets

Ressourcen

Zugehörige Dokumente:

- AWS Control Tower
- Richtlinien zur AWS-Sicherheitsprüfung
- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Sicherheitsberichte

Relevante Videos:

- Enable AWS adoption at scale with automation and governance (AWS-Übernahme in großem Umfang mit Automatisierung und Governance)
- Security Best Practices the Well-Architected Way

Zugehörige Beispiele:

- Übung: AWS-Konto und Root-Benutzer

SEC01-BP03 Identifizieren und Validieren von Kontrollzielen

Entsprechend Ihren Compliance-Anforderungen und Risiken, die aus Ihrem Bedrohungsmodell identifiziert werden, können Sie die Kontrollziele und Kontrollen ableiten und validieren, die Sie für Ihren Workload benötigen. Die laufende Validierung von Kontrollzielen und Kontrollen hilft Ihnen, die Effektivität der Risikominderung zu messen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Identifizieren Sie die Compliance-Anforderungen: Ermitteln Sie die organisatorischen, rechtlichen und Compliance-bezogenen Anforderungen, die Ihr Workload erfüllen muss.
- Identifizieren Sie AWS-Compliance-Ressourcen: Ermitteln Sie die Ressourcen, die AWS zur Verfügung stellt, um Sie bei der Compliance zu unterstützen.
  - https://aws.amazon.com/compliance/
  - https://aws.amazon.com/artifact/

Ressourcen

Zugehörige Dokumente:

- Richtlinien zur AWS-Sicherheitsprüfung
- Sicherheitsberichte

Relevante Videos:

- AWS Security Hub: Manage Security Alerts and Automate Compliance (Verwalten von Sicherheitsbenachrichtigungen und Automatisieren der Compliance)
- Security Best Practices the Well-Architected Way

SEC01-BP04 Sicherstellen der Aktualität von Informationen zu Sicherheitsbedrohungen

Um geeignete Kontrollen zu definieren und zu implementieren, müssen Sie Angriffsvektoren erkennen, indem Sie stets über die neuesten Sicherheitsbedrohungen auf dem Laufenden bleiben. Nutzen Sie AWS Managed Services, um einfacher über unerwartetes oder ungewöhnliches Verhalten in Ihren AWS-Konten benachrichtigt zu werden. Verwenden Sie für Untersuchungen im Rahmen Ihrer Abläufe zu Sicherheitsinformationen AWS-Partner-Tools oder Feeds mit Risikoinformationen von Drittanbietern. Die Liste der Common Vulnerabilities and Exposures (CVE) enthält öffentlich bekannte Cybersicherheitsrisiken, sodass Sie immer auf dem aktuellen Stand sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Abonnieren Sie Informationsquellen zu Bedrohungen: Überprüfen Sie regelmäßig Informationen aus mehreren Quellen zu Bedrohungen, die für die in Ihrem Workload verwendeten Technologien relevant sind.
  - Liste der Common Vulnerabilities and Exposures (CVE)
- Verwenden Sie den AWS Shield Advanced -Service: So erhalten Sie nahezu in Echtzeit Einblicke in Informationsquellen, wenn Ihr Workload über das Internet zugänglich ist.

Ressourcen

Zugehörige Dokumente:

- Richtlinien zur AWS-Sicherheitsprüfung
- AWS Shield
- Sicherheitsberichte

Relevante Videos:

- Security Best Practices the Well-Architected Way

SEC01-BP05 Aktuelle Informationen dank Sicherheitsempfehlungen

Bleiben Sie mit AWS- und Branchensicherheitsempfehlungen auf dem Laufenden, um die Sicherheitsstrategie für Ihren Workload zu entwickeln. AWS-Sicherheitsmitteilungen enthalten wichtige Informationen zur Sicherheit und zum Datenschutz.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verfolgen Sie AWS-Updates: Abonnieren Sie diese oder überprüfen Sie sie regelmäßig in Bezug auf neue Empfehlungen, Tipps und Tricks.
  - AWS Well-Architected Labs
  - AWS-Sicherheitsblog
  - AWS-Servicedokumentation
- Abonnieren Sie Branchennachrichten: Überprüfen Sie regelmäßig Newsfeeds aus verschiedenen Quellen, die für die in Ihrem Workload verwendeten Technologien relevant sind.
  - Beispiel: Liste der Common Vulnerabilities and Exposures (CVE)

Ressourcen

Zugehörige Dokumente:

- Sicherheitsberichte

Relevante Videos:

- Security Best Practices the Well-Architected Way

SEC01-BP06 Automatisieren von Tests und Validierung von Sicherheitskontrollen in Pipelines

Erstellen Sie sichere Ausgangswerte und Vorlagen für Sicherheitsmechanismen, die im Rahmen Ihres Builds, Ihrer Pipelines und Prozesse getestet und validiert werden. Verwenden Sie Tools und Automatisierung, um alle Sicherheitskontrollen kontinuierlich zu testen und zu validieren. Scannen Sie beispielsweise Elemente wie Machine Images und Infrastruktur als Codevorlagen in jeder Phase auf Sicherheitslücken, Unregelmäßigkeiten und Abweichungen von einer etablierten Ausgangsbasis. Mit AWS CloudFormation Guard können Sie sicherstellen, dass CloudFormation-Vorlagen sicher sind, Sie dadurch Zeit sparen und das Risiko von Konfigurationsfehlern verringert wird.

Wichtig ist, die Zahl der fehlerhaften Sicherheitskonfigurationen in einer Produktionsumgebung zu reduzieren. Je mehr Qualitätskontrollen Sie während des Entwicklungsprozesses durchführen und je mehr Fehler Sie vorab eliminieren können, desto besser. Entwickeln Sie Continuous Integration und Continuous Deployment-Pipelines (CI/CD), um kontinuierlich Sicherheitsprobleme zu erkennen. CI/

CD-Pipelines bieten die Möglichkeit, die Sicherheit in jeder Phase der Erstellung und Bereitstellung zu erhöhen. CI/CD-Sicherheitstools müssen kontinuierlich aktuell gehalten werden, um sie den sich ständig verändernden Bedrohungen anzupassen.

Verfolgen Sie Änderungen an der Workload-Konfiguation nach. Dies hilft Ihnen bei Compliance-Auditing, Änderungsverwaltung und ggf. bei Untersuchungen. Sie können mit AWS Config Ihre AWS- und Drittanbieterressourcen aufzeichnen und evaluieren. So können Sie die allgemeine Compliance mit Regeln und Conformance Packs, d. h. Regelsammlungen mit Maßnahmen zur Problembehebung, kontinuierlich prüfen und bewerten.

Die Änderungsverfolgung sollte geplante Änderungen einschließen, die Teil des Änderungskontrollprozesses Ihrer Organisation sind (manchmal als „MACD“ bezeichnet – Move/Add/Change/Delete), außerdem ungeplante Änderungen und unerwartete Änderungen, beispielsweise Vorfälle. Änderungen können sowohl bei der Infrastruktur als auch im Zusammenhang mit anderen Kategorien auftreten, z. B. Änderungen an Code-Repositorys, Machine Images oder beim Anwendungsinventar, sowie Prozess- und Richtlinienänderungen oder auch Änderungen an der Dokumentation.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren Sie die Konfigurationsverwaltung: Legen Sie fest, dass sichere Konfigurationen automatisch erzwungen und validiert werden. Verwenden Sie dazu einen Service oder ein Tool zur Konfigurationsverwaltung.
  - AWS Systems Manager
  - AWS CloudFormation
  - Einrichten einer CI/CD-Pipeline in AWS

Ressourcen

Zugehörige Dokumente:

- Verwenden von Service-Kontrollrichtlinien zum Festlegen eines kontenübergreifenden Integritätsschutzes für Berechtigungen in AWS Organizations

Relevante Videos:

- Managing Multi-Account AWS Environments Using AWS Organizations (Verwalten von AWS-Umgebungen mit mehreren Konten mithilfe von AWS Organizations)
- Security Best Practices the Well-Architected Way

SEC01-BP07 Identifizieren und Priorisieren von Risiken anhand eines Bedrohungsmodells

Verwenden Sie ein Bedrohungsmodell, um ein aktuelles Register potenzieller Bedrohungen zu erstellen und zu pflegen. Priorisieren Sie Bedrohungen und passen Sie Ihre Sicherheitskontrollen an, um zu verhindern, zu erkennen und zu reagieren. Wiederholen und warten Sie dies im Kontext der sich entwickelnden Sicherheitslandschaft.

Die Modellierung von Bedrohungen ermöglicht einen systematischen Ansatz, um Sicherheitsprobleme früh im Designprozess zu finden und zu beheben. Dabei gilt „je früher desto besser“, weil die Kosten für solche Problembehebungen am Anfang geringer sind als zu einem späteren Zeitpunkt.

Die Modellierung von Bedrohungen umfasst im Allgemeinen die folgenden Schritte:

1. Identifizieren von Ressourcen, Akteuren, Einstiegspunkten, Komponenten, Anwendungsfällen und Vertrauensstufen und Darstellen dieser Elemente in einem Designdiagramm.
2. Erstellen einer Liste der Risiken.
3. Identifizieren von Fehlerbehebungen für jede Bedrohung, ggf. mit Implementierung von Sicherheitskontrollen.
4. Erstellen und Prüfen einer Risikomatrix, um zu ermitteln, ob die Maßnahmen gegen die Bedrohung angemessen sind.

Die Modellierung von Bedrohungen ist am effektivsten, wenn sie auf Workload-Ebene (oder auf Ebene von Workload-Funktionen) durchgeführt wird. So ist sicherstellt, dass der gesamte Kontext bei der Bewertung berücksichtigt werden kann. Überprüfen und aktualisieren Sie diese Matrix regelmäßig, um die Entwicklung Ihrer Sicherheitsumgebung abzubilden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Erstellen Sie ein Bedrohungsmodell: Ein Bedrohungsmodell kann Ihnen helfen, potenzielle Sicherheitsbedrohungen zu identifizieren und zu beheben.

- NIST: Guide to Data-Centric System Threat Modeling (Handbuch zur datenzentrischen Modellierung von Systembedrohungen)

Ressourcen

Ähnliche Dokumente:

- Richtlinien zur AWS-Sicherheitsprüfung
- Sicherheitsberichte

Ähnliche Videos:

- Security Best Practices the Well-Architected Way

SEC01-BP08 Regelmäßiges Bewerten und Implementieren neuer Sicherheitsservices und -funktionen

Bewerten und implementieren Sie Sicherheitsservices und -funktionen von AWS und AWS-Partnern, mit denen Sie die Sicherheitsstrategie für Ihren Workload weiterentwickeln können. Das AWS-Sicherheitsblog bietet Informationen zu neuen AWS-Services und -Funktionen, Implementierungshandbücher und allgemeine Hinweise zur Sicherheit. Neuerungen bei AWS ist eine gute Möglichkeit, einen Überblick über alle neuen Funktionen, Services und Ankündigungen im Zusammenhang mit AWS zu erhalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Planen Sie regelmäßige Überprüfungen: Erstellen Sie einen Kalender mit Überprüfungsaktivitäten. Darin sollte Folgendes enthalten sein: Prüfung von Compliance-Anforderungen, Bewertung neuer AWS-Sicherheitsfunktionen und -services und Information über aktuelle Branchennachrichten.
- Informieren Sie sich über AWS-Services und -Funktionen: Erkunden Sie die für die von Ihnen genutzten Services verfügbaren Sicherheitsfunktionen und prüfen Sie neue Funktionen, sobald sie veröffentlicht werden.
  - AWS-Sicherheitsblog
  - AWS-Sicherheitsmitteilungen
  - AWS-Servicedokumentation

- Definieren Sie einen Onboarding-Prozess für AWS-Services: Definieren Sie Prozesse für das Onboarding neuer AWS-Services. Berücksichtigen Sie dabei, wie neue AWS-Services auf ihre Funktionalität hin bewertet werden sollen, und die Compliance-Anforderungen für Ihren Workload.
- Testen Sie neue Services und Funktionen: Testen Sie neue Services und Funktionen nach ihrer Veröffentlichung in einer Nicht-Produktionsumgebung, die Ihre Produktionsumgebung möglichst genau repliziert.
- Implementieren Sie andere Verteidigungsmechanismen: Implementieren Sie automatisierte Mechanismen zum Schutz Ihres Workloads und prüfen Sie die verfügbaren Optionen.
  - Korrigieren von nicht konformen AWS-Ressourcen mit AWS-Config-Regeln

Ressourcen

Relevante Videos:

- Security Best Practices the Well-Architected Way

Identity and Access Management

Fragen

- SICH 2 Was ist bei der Verwaltung der Authentifizierung für Personen und Rechner zu beachten?
- SICH 3 Wie verwalten Sie Berechtigungen für Personen und Maschinen?

SICH 2 Was ist bei der Verwaltung der Authentifizierung für Personen und Rechner zu beachten?

Es gibt zwei Arten von Identitäten, die Sie beim Betrieb sicherer AWS-Workloads verwalten müssen. Wenn Sie wissen, welche Art von Identität Sie verwalten und wie Sie Zugriff gewähren müssen, können Sie sicherstellen, dass die richtigen Identitäten unter den richtigen Bedingungen Zugriff auf die richtigen Ressourcen haben.

Menschliche Identitäten: Ihre Administratoren, Entwickler, Bediener und Endbenutzer benötigen eine Identität für den Zugriff auf Ihre AWS-Umgebungen und -Anwendungen. Dies sind Mitglieder Ihrer Organisation oder externe Benutzer, mit denen Sie zusammenarbeiten, und die mit Ihren AWS-Ressourcen über einen Webbrowser, eine Client-Anwendung oder interaktive Befehlszeilen-Tools interagieren.

Maschinenidentitäten: Ihre Service-Anwendungen, betrieblichen Tools und Workloads benötigen eine Identität, um Anforderungen an AWS-Services zu stellen, z. B. um Daten zu lesen. Zu diesen Identitäten gehören Maschinen, die in Ihrer AWS-Umgebung ausgeführt werden, z. B. Amazon EC2-Instances oder AWS-Funktionen. Sie können auch Maschinenidentitäten für externe Parteien verwalten, die Zugriff benötigen. Darüber hinaus verfügen Sie möglicherweise auch über Maschinen außerhalb von AWS, die Zugriff auf Ihre AWS-Umgebung benötigen.

Bewährte Methoden

- SEC02-BP01 Verwenden von starken Anmeldemechanismen
- SEC02-BP02 Verwenden von temporären Anmeldeinformationen
- SEC02-BP03 Sicheres Speichern und Verwenden von Secrets
- SEC02-BP04 Verlassen auf einen zentralen Identitätsanbieter
- SEC02-BP05 Regelmäßiges Überprüfen und Rotieren von Anmeldeinformationen
- SEC02-BP06 Nutzen von Benutzergruppen und Attributen

SEC02-BP01 Verwenden von starken Anmeldemechanismen

Erzwingen Sie die Mindestlänge des Passworts und informieren Sie Benutzer, dass sie gängige oder wiederverwendete Passwörter vermeiden sollen. Erzwingen Sie die Multi-Factor Authentication (MFA) mit Software- oder Hardwaremechanismen und stellen Sie so eine zusätzliche Verifizierungsstufe bereit. Wenn Sie beispielsweise IAM Identity Center als Identitätsquelle verwenden, konfigurieren Sie die MFA-Einstellung „context-aware“ oder „always-on“ und erlauben Sie Benutzern, ihre eigenen MFA-Geräte zu registrieren, um die Akzeptanz zu beschleunigen. Wenn Sie einen externen Identitätsanbieter (IdP) verwenden, konfigurieren Sie Ihren Identitätsanbieter für MFA.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Erstellen Sie eine AWS Identity and Access Management-Richtlinie (IAM), um die MFA-Anmeldung zu erzwingen: Erstellen Sie eine vom Kunden verwaltete IAM-Richtlinie, die alle IAM-Aktionen untersagt, außer die, mit denen die Benutzer Rollen annehmen, ihre eigenen Anmeldeinformationen ändern und ihre MFA-Geräte verwalten können (auf der Seite „My Security Credentials“).
- Aktivieren Sie MFA beim Identitätsanbieter: Aktivieren Sie MFA bei dem Identitätsanbieter oder Single Sign-on-Service, den Sie verwenden, z. B. AWS IAM Identity Center.

- Konfigurieren Sie eine sichere Passwortrichtlinie: Konfigurieren Sie eine sichere Passwortrichtlinie in IAM- und Identitätsverbundsystemen, um sich vor Brute-Force-Angriffen zu schützen.

- Regelmäßiges Ändern von Anmeldeinformationen: Sorgen Sie dafür, dass Administratoren Ihres Workloads die eigenen Passwörter und ggf. Zugriffsschlüssel regelmäßig ändern.

Ressourcen

Zugehörige Dokumente:

- Erste Schritte mit AWS Secrets Manager
- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Identitätsanbieter und Verbund
- Stammbenutzer des AWS-Kontos
- Erste Schritte mit AWS Secrets Manager
- Temporäre Sicherheits-Anmeldeinformationen
- Partnerlösungen im Bereich Sicherheit: Zugriff und Zugriffssteuerung
- Temporäre Sicherheits-Anmeldeinformationen
- Stammbenutzer des AWS-Kontos

Relevante Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)
- Managing user permissions at scale with IAM Identity Center (Verwalten von Benutzerberechtigungen in großem Umfang mit IAM Identity Center)
- Mastering identity at every layer of the cake

SEC02-BP02 Verwenden von temporären Anmeldeinformationen

Erzwingen Sie, dass Identitäten temporäre Anmeldeinformationen dynamisch. Verwenden Sie für Identitäten von Arbeitskräften AWS IAM Identity Center oder einen Verbund mit AWS Identity and Access Management-Rollen (IAM), um auf AWS-Konten zuzugreifen. Für Maschinenidentitäten, wie Amazon Elastic Compute Cloud-Instances (Amazon EC2) oder AWS Lambda-Funktionen, fordern Sie die Verwendung von IAM-Rollen anstelle von IAM-Benutzern mit langfristigen Zugriffsschlüsseln.

Für Identitäten von Personen, die die AWS Management Console verwenden, müssen Benutzer temporäre Anmeldeinformationen anfordern und einen Verbund mit AWS erstellen. Dies kann über das AWS IAM Identity Center-Benutzerportal erfolgen. Für Benutzer, die CLI-Zugriff benötigen, stellen Sie sicher, dass sie die AWS CLI v2verwenden, die die direkte Integration mit IAM Identity Center unterstützt. Benutzer können CLI-Profile erstellen, die mit IAM-Identity-Center-Konten und -Rollen verknüpft sind. Die CLI ruft automatisch AWS-Anmeldeinformationen aus IAM Identity Center ab und aktualisiert sie in Ihrem Namen. Dadurch müssen temporäre AWS-Anmeldeinformationen nicht mehr aus der IAM Identity Center-Konsole kopiert und eingefügt werden. Für SDK sollten sich Benutzer zur Übernahme von Rollen auf AWS Security Token Service (AWS STS) verlassen, um temporäre Anmeldeinformationen zu erhalten. In bestimmten Fällen sind temporäre Anmeldeinformationen möglicherweise nicht praktisch. Sie sollten sich der Risiken bewusst sein, die durch das Speichern von Zugriffsschlüsseln entstehen, diese häufig ändern und wenn möglich eine Multi-Faktor-Authentifizierung (MFA) verlangen. Bestimmen Sie anhand der Informationen zum letzten Zugriff, wann Zugriffsschlüssel rotiert oder entfernt werden sollten.

Wenn Sie Konsumenten Zugriff auf Ihre AWS-Ressourcen gewähren müssen, verwenden Sie Amazon Cognito -Identitäten-Pools und weisen Sie ihnen temporäre Anmeldeinformationen mit eingeschränkten Berechtigungen für den Zugriff auf Ihre AWS-Ressourcen zu. Die Berechtigungen für jeden Benutzer werden über IAM-Rollen gesteuert, die Sie erstellen. Sie können Regeln definieren, um die Rolle für jeden Benutzer basierend auf Ansprüchen im ID-Token des Benutzers auszuwählen. Sie können eine Standardrolle für authentifizierte Benutzer definieren. Sie können auch eine separate IAM-Rolle mit eingeschränkten Berechtigungen für Gastbenutzer definieren, die nicht authentifiziert sind.

Für Maschinenidentitäten sollten Sie sich auf IAM-Rollen verlassen, um Zugriff auf AWS zu gewähren. Für Amazon Elastic Compute Cloud-Instances (Amazon EC2) können Sie Rollen für Amazon EC2verwenden. Sie können Ihrer Amazon EC2-Instance eine IAM-Rolle zuweisen, damit Ihre in Amazon EC2 ausgeführten Anwendungen temporäre Sicherheitsanmeldeinformationen verwenden können, die AWS über den Instance Metadata Service (IMDS) automatisch erstellt, verteilt und regelmäßig ändert. Die aktuelle Version von IMDS schützt vor Schwachstellen, die die temporären Anmeldeinformation offenlegen, und sollten implementiert werden. Für den Zugriff auf Amazon EC2-Instances mithilfe von Schlüsseln oder Passwörtern ist AWS Systems Manager eine sicherere Möglichkeit, auf Ihre Instances zuzugreifen und diese mit einem vorinstallierten Agent ohne das gespeicherte Secret zu verwalten. Darüber hinaus ermöglichen Ihnen andere AWS-Services wie AWS Lambda die Konfiguration einer IAM-Servicerolle, um dem Service Berechtigungen zum Ausführen von AWS-Aktionen unter Verwendung temporärer Anmeldeinformationen zu erteilen. In Situationen, in denen Sie keine temporären Anmeldeinformationen verwenden können, arbeiten Sie

mit programmgesteuerten Tools wie AWS Secrets Manager, um die Rotation und Verwaltung von Anmeldeinformation zu automatisieren.

Regelmäßiges Überprüfen und Ändern von Anmeldeinformationen: Eine regelmäßige Validierung, vorzugsweise durch ein automatisiertes Tool, ist erforderlich, um zu überprüfen, ob die richtigen Kontrollen angewendet werden. Für Personenidentitäten sollten Sie festlegen, dass Benutzer ihre Passwörter regelmäßig ändern und anstelle von Zugriffsschlüsseln temporäre Anmeldeinformationen verwenden. Bei der Umstellung von IAM-Benutzern zu zentralisierten Identitäten können Sie einen Bericht zu Anmeldeinformationen generieren , um Ihre IAM-Benutzer zu überprüfen. Wir empfehlen außerdem, dass Sie die MFA-Einstellungen in Ihrem Identitätsanbieter erzwingen. Sie können AWS-Config-Regeln einrichten, um diese Einstellungen zu überwachen. Für Maschinenidentitäten sollten Sie sich auf temporäre Anmeldeinformationen mit IAM-Rollen verlassen. In Situationen, in denen dies nicht möglich ist, ist eine häufige Prüfung und Änderung von Zugriffsschlüsseln erforderlich.

Sicheres Speichern und Verwenden von geheimen Schlüsseln: Verwenden Sie für Anmeldeinformationen, die nicht IAM-bezogen sind und für die keine temporären Anmeldeinformationen genutzt werden können, z. B. Datenbankanmeldungen, einen Service, der für die Verwaltung von Secrets entwickelt wurde, z. B. Secrets Manager. Secrets Manager vereinfacht die Verwaltung, Änderung und sichere Speicherung verschlüsselter Secrets mit unterstützten Diensten. Aufrufe für den Zugriff auf die Secrets werden zu Prüfungszwecken in AWS CloudTrail protokolliert und IAM-Berechtigungen können Zugriff auf sie mit den geringsten Rechten gewähren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren Sie Richtlinien zur geringsten Berechtigung: Weisen Sie IAM-Gruppen und -Rollen Zugriffsrichtlinien zu, die in ihrem Umfang möglichst gering und an den Tätigkeitsbereich der Benutzer angepasst sind.
  - Gewähren von geringsten Rechten
- Entfernen Sie unnötige Berechtigungen: Implementieren Sie das Prinzip der geringsten Berechtigung, indem Sie unnötige Berechtigungen zurücknehmen.
  - Reduzieren des Richtlinienbereichs durch Anzeigen der Benutzeraktivität
  - Anzeigen des Rollenzugriffs

- Eine Berechtigungsgrenze ist eine erweiterte Funktion für eine verwaltete Richtlinie. Sie legt die maximalen Berechtigungen fest, die mit einer identitätsbasierten Richtlinie einer IAM-Entität erteilt werden kann. Eine Berechtigungsgrenze erlaubt einer Entität nur die Ausführung jener Aktionen,

die sowohl nach ihren identitätsbasierten Richtlinien als auch nach ihren Berechtigungsgrenzen zulässig sind.
    - Übung: IAM-Berechtigungsgrenzen – Übertragung der Rollenerstellung
- Erwägen Sie die Verwendung von Ressourcen-Tags für Berechtigungen: Mit Tags können Sie den Zugriff auf die AWS-Ressourcen steuern, die das Tagging unterstützen. Sie können IAM-Benutzer und -Rollen auch taggen, um zu steuern, worauf sie zugreifen können.
    - Übung: IAM Tag-basierte Zugriffskontrolle für EC2
    - Attributbasierte Zugriffskontrolle (ABAC)

Ressourcen

Zugehörige Dokumente:

- Erste Schritte mit AWS Secrets Manager
- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Identitätsanbieter und Verbund
- Partnerlösungen im Bereich Sicherheit: Zugriff und Zugriffssteuerung
- Temporäre Sicherheits-Anmeldeinformationen
- Stammbenutzer des AWS-Kontos

Relevante Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)
- Managing user permissions at scale with AWS IAM Identity Center (Verwalten von Benutzerberechtigungen in großem Umfang mit AWS IAM Identity Center)
- Mastering identity at every layer of the cake

SEC02-BP03 Sicheres Speichern und Verwenden von Secrets

Speichern Sie Arbeitskräfte- und Maschinenidentitäten, die Secrets wie Passwörter für Drittanbieter-Anwendungen erfordern, mit automatischer Rotation unter Verwendung der neuesten Branchenstandards in einem spezialisierten Service. Für Anmeldeinformationen, die nicht IAM-bezogen sind und für die keine temporären Anmeldeinformation verwendet werden können, z. B. Datenbankanmeldungen, nutzen Sie einen Service, der für die Verwaltung von Secrets entwickelt

wurde, z. B. AWS Secrets Manager. Mit Secrets Manager können Sie bequem verschlüsselte Secrets mit unterstützten Services verwalten, rotieren und sicher speichern. Aufrufe für den Zugriff auf die Secrets werden zu Prüfungszwecken in AWS CloudTrail protokolliert und IAM-Berechtigungen können Zugriff auf sie mit den geringsten Rechten gewähren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden Sie AWS Secrets Manager: AWS Secrets Manager ist ein AWS-Service für die einfache Verwaltung von Secrets. Geheime Schlüssel können Datenbank-Anmeldeinformationen, Passwörter, API-Schlüssel von Dritten und sogar beliebiger Text sein.

Ressourcen

Ähnliche Dokumente:

- Erste Schritte mit AWS Secrets Manager
- Identitätsanbieter und Verbund

Ähnliche Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)

SEC02-BP04 Verlassen auf einen zentralen Identitätsanbieter

Verlassen Sie sich bei Identitäten von Arbeitskräften auf einen Identitätsanbieter, mit dem Sie Identitäten zentral verwalten können. Dadurch ist es einfacher, den Zugriff über mehrere Anwendungen und Services hinweg zu verwalten, da Sie den Zugriff von einem einzigen Standort aus erstellen, verwalten und widerrufen. Wenn beispielsweise jemand Ihr Unternehmen verlässt, können Sie den Zugriff für alle Anwendungen und Services (einschließlich AWS) an einem Standort widerrufen. Dies reduziert die Notwendigkeit mehrerer Anmeldeinformationen und bietet die Möglichkeit der Integration in bereits vorhandene HR-Prozesse.

Für den Verbund mit einzelnen AWS-Konten können Sie zentrale Identitäten für AWS mit einem SAML 2.0-basierten Anbieter mit AWS Identity and Access Management verwenden. Sie können jeden Anbieter verwenden, der von Ihnen in AWS oder außerhalb von AWS gehostet oder vom AWS Partner bereitgestellt wird und mit dem SAML 2.0 -Protokoll kompatibel ist. Sie können einen Verbund

zwischen Ihrem AWS-Konto und dem von Ihnen gewählten Anbieter verwenden, um einem Benutzer oder einer Anwendung Zugriff zum Aufrufen von AWS-API-Vorgängen zu gewähren, indem Sie über eine SAML-Zusicherung temporäre Sicherheitsanmeldeinformationen abrufen. Webbasiertes SSO wird ebenfalls unterstützt, sodass sich Benutzer über Ihre Website bei der AWS Management Console anmelden können.

Für den Verbund mit mehreren Konten in Ihrer AWS Organizations können Sie Ihre Identitätsquelle in AWS IAM Identity Center (IAM Identity Center)konfigurieren und angeben, wo Ihre Benutzer und Gruppen gespeichert werden. Nach der Konfiguration ist Ihr Identitätsanbieter Ihre Quelle der Wahrheit. Informationen können mithilfe des SCIM-Protokolls (System for Cross-Domain Identity Management) v2.0 synchronisiert werden. Anschließend können Sie Benutzer oder Gruppen abrufen und ihnen IAM Identity Center-Zugriff auf AWS-Konten, Cloud-Anwendungen oder beides gewähren.

IAM Identity Center ist in AWS Organizations integriert, sodass Sie Ihren Identitätsanbieter einmal konfigurieren und dann Zugriff auf vorhandene und neue Konten gewähren können, die in Ihrem Unternehmen verwaltet werden. IAM Identity Center bietet Ihnen einen Standardspeicher, den Sie verwenden können, um Ihre Benutzer und Gruppen zu verwalten. Wenn Sie sich für die Verwendung des IAM Identity Center-Speichers entscheiden, erstellen Sie Ihre Benutzer und Gruppen und weisen deren Zugriffsebene Ihren AWS-Konten und -Anwendungen zu. Beachten Sie dabei die bewährte Methode der geringsten Berechtigung. Alternativ können Sie eine Verbindung zu Ihrem externen Identitätsanbieter über SAML 2.0 oder eine Verbindung zu Ihrem Microsoft AD-Verzeichnis über AWS Directory Service herstellen. Nach der Konfiguration können Sie sich bei der AWS Management Console oder der mobilen AWS-App anmelden, indem Sie sich über Ihren zentralen Identitätsanbieter authentifizieren.

Für die Verwaltung von Endbenutzern oder Verbrauchern Ihrer Workloads, z. B. einer mobilen App, können Sie Amazon Cognito. Es bietet Authentifizierung, Autorisierung und Benutzerverwaltung für Ihre Web- und mobilen Anwendungen. Ihre Benutzer können sich direkt mit einem Benutzernamen und Passwort oder über einen Drittanbieter wie Amazon, Apple, Facebook oder Google anmelden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Zentralisieren Sie den administrativen Zugriff: Erstellen Sie im IAM-Identitätsanbieter (Identity and Access Management) eine Entität, um eine Vertrauensstellung zwischen Ihrem AWS-Konto und dem Identitätsanbieter (IDP) herzustellen. IAM unterstützt Identitätsanbieter, die mit OpenID Connect (OIDC) oder SAML 2.0 (Security Assertion Markup Language 2.0) kompatibel sind.
  - Identitätsanbieter und Verbund

- Zentralisieren Sie den Anwendungszugriff: Erwägen Sie, Amazon Cognito zum Zentralisieren des Anwendungszugriffs zu verwenden. Damit können Sie Ihren Webanwendungen und mobilen Apps auf schnelle und einfache Weise die Benutzerregistrierung und -anmeldung sowie die Zugriffskontrolle hinzufügen. Amazon Cognito lässt sich auf Millionen von Benutzern hochskalieren. Es unterstützt die Anmeldung mit Social-Identity-Anbietern wie Facebook, Google und Amazon sowie mit Unternehmens-Identitätsanbietern über SAML 2.0.

- Entfernen Sie alte IAM-Benutzer und -Gruppen: Sobald Sie einen Identitätsanbieter (IDP) verwenden, sollten Sie nicht mehr benötigte IAM-Benutzer und -Gruppen entfernen.
  - Suchen nach ungenutzten Anmeldeinformationen
  - Löschen einer IAM-Benutzergruppe

Ressourcen

Ähnliche Dokumente:

- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Partnerlösungen im Bereich Sicherheit: Zugriff und Zugriffssteuerung
- Temporäre Sicherheits-Anmeldeinformationen
- Stammbenutzer des AWS-Kontos

Ähnliche Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)
- Managing user permissions at scale with AWS IAM Identity Center (Verwalten von Benutzerberechtigungen in großem Umfang mit AWS IAM Identity Center)
- Mastering identity at every layer of the cake

SEC02-BP05 Regelmäßiges Überprüfen und Rotieren von Anmeldeinformationen

Wenn Sie sich nicht auf temporäre Anmeldeinformationen verlassen können und langfristige Anmeldeinformationen benötigen, prüfen Sie die Anmeldeinformationen, um sicherzustellen, dass die definierten Kontrollen (z. B. Multi-Faktor-Authentifizierung, MFA) erzwungen und regelmäßig rotiert werden sowie über die entsprechende Zugriffsebene verfügen. Eine regelmäßige Validierung,

vorzugsweise durch ein automatisiertes Tool, ist erforderlich, um zu überprüfen, ob die richtigen Kontrollen angewendet werden. Für Personenidentitäten sollten Sie festlegen, dass Benutzer ihre Passwörter regelmäßig ändern und anstelle von Zugriffsschlüsseln temporäre Anmeldeinformationen verwenden. Bei der Umstellung von AWS Identity and Access Management-Benutzern (IAM) zu zentralisierten Identitäten können Sie einen Bericht zu Anmeldeinformationen generieren , um Ihre IAM-Benutzer zu überprüfen. Wir empfehlen außerdem, dass Sie die MFA-Einstellungen in Ihrem Identitätsanbieter erzwingen. Sie können AWS-Config-Regeln einrichten, um diese Einstellungen zu überwachen. Für Maschinenidentitäten sollten Sie sich auf temporäre Anmeldeinformationen mit IAM-Rollen verlassen. In Situationen, in denen dies nicht möglich ist, ist eine häufige Prüfung und Änderung von Zugriffsschlüsseln erforderlich.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Prüfen Sie Anmeldeinformationen regelmäßig: Verwenden Sie Berichte zu Anmeldeinformationen und Identify and Access Management (IAM) Access Analyzer, um IAM-Anmeldeinformationen und -Berechtigungen zu prüfen.
  - IAM Access Analyzer
  - Abrufen von Berichten zu Anmeldeinformationen für Ihr AWS-Konto
  - Übung: Automated IAM user cleanup (Automatisierte Bereinigung von IAM-Benutzern)
- Verwenden Sie Zugriffsebenen, um IAM-Berechtigungen zu prüfen: Um die Sicherheit Ihres AWS-Konto-Kontos zu erhöhen, sollten Sie Ihre IAM-Richtlinien regelmäßig überprüfen und überwachen. Sorgen Sie dafür, dass mit den Richtlinien Zugriffsrechte nur in dem Umfang erteilt werden, wie sie für die jeweiligen Aktionen erforderlich sind.
  - Überprüfen von IAM-Berechtigungen mithilfe von Zugriffsebenen
- Erwägen Sie, die Erstellung und Aktualisierung von IAM-Ressourcen zu automatisieren: Mit AWS CloudFormation kann die Bereitstellung von IAM-Ressourcen einschließlich Rollen und Richtlinien automatisiert werden. So lässt sich die Zahl menschlicher Fehler verringern, da die Vorlagen verifiziert und ihre Versionen kontrolliert werden können.
  - Übung: Automated deployment of IAM groups and roles (Automatisierte Bereitstellung von IAM-Gruppen und -Rollen)

Ressourcen

Zugehörige Dokumente:

- Erste Schritte mit AWS Secrets Manager
- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Identitätsanbieter und Verbund
- Partnerlösungen im Bereich Sicherheit: Zugriff und Zugriffssteuerung
- Temporäre Sicherheits-Anmeldeinformationen

Relevante Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)
- Managing user permissions at scale with AWS IAM Identity Center (Verwalten von Benutzerberechtigungen in großem Umfang mit AWS IAM Identity Center)
- Mastering identity at every layer of the cake

SEC02-BP06 Nutzen von Benutzergruppen und Attributen

Wenn die Anzahl der von Ihnen verwalteten Benutzer zunimmt, müssen Sie diese so organisieren, dass Sie sie im erforderlichen Umfang verwalten können. Platzieren Sie Benutzer mit allgemeinen Sicherheitsanforderungen in Gruppen, die von Ihrem Identitätsanbieter definiert wurden, und implementieren Sie Mechanismen, um sicherzustellen, dass Benutzerattribute, die für die Zugriffskontrolle verwendet werden können (zum Beispiel Abteilung oder Standort), korrekt und auf dem neuesten Stand sind. Verwenden Sie diese Gruppen und Attribute anstelle einzelner Benutzer, um den Zugriff zu steuern. Auf diese Weise können Sie den Zugriff zentral verwalten, indem Sie die Gruppenmitgliedschaft oder Attribute eines Benutzers einmal mit einem Berechtigungssatzändern, anstatt viele einzelne Richtlinien zu aktualisieren, wenn sich die Zugriffsanforderungen eines Benutzers ändern. Sie können AWS IAM Identity Center (IAM Identity Center) verwenden, um Benutzergruppen und Attribute zu verwalten. IAM Identity Center unterstützt die am häufigsten verwendeten Attribute unabhängig davon, ob sie manuell während der Benutzererstellung eingegeben oder automatisch mithilfe einer Synchronisierungs-Engine bereitgestellt werden, wie in der Spezifikation „System for Cross-Domain Identity Management (SCIM)“ definiert.

Platzieren Sie Benutzer mit allgemeinen Sicherheitsanforderungen in Gruppen, die von Ihrem Identitätsanbieter definiert wurden, und implementieren Sie Mechanismen, um sicherzustellen, dass

Benutzerattribute, die für die Zugriffskontrolle verwendet werden können (zum Beispiel Abteilung oder Standort), korrekt und auf dem neuesten Stand sind. Verwenden Sie diese Gruppen und Attribute anstelle einzelner Benutzer, um den Zugriff zu steuern. Auf diese Weise können Sie den Zugriff zentral verwalten, indem Sie die Gruppenmitgliedschaft oder Attribute eines Benutzers einmal ändern, anstatt viele einzelne Richtlinien zu aktualisieren, wenn sich die Zugriffsanforderungen eines Benutzers ändern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Konfigurieren Sie Gruppen, wenn Sie AWS IAM Identity Center (IAM Identity Center) verwenden: IAM Identity Center bietet Ihnen die Möglichkeit, Benutzergruppen zu konfigurieren und Gruppen die gewünschte Berechtigungsstufe zuzuweisen.
  - AWS Single Sign-On – Verwalten von Identitäten
- Informieren Sie sich über die attributbasierte Zugriffskontrolle (ABAC): ABAC (Attribute-based Access Control) ist eine Autorisierungsstrategie, die Berechtigungen basierend auf Attributen definiert.
  - Was ist ABAC für AWS?
  - Übung: IAM Tag-basierte Zugriffskontrolle für EC2

Ressourcen

Ähnliche Dokumente:

- Erste Schritte mit AWS Secrets Manager
- Security best practices in IAM (Bewährte Methoden für die Sicherheit in IAM)
- Identitätsanbieter und Verbund
- Stammbenutzer des AWS-Kontos

Ähnliche Videos:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale (Bewährte Methoden zum Verwalten, Abrufen und Rotieren von Secrets in großem Umfang)
- Managing user permissions at scale with AWS IAM Identity Center (Verwalten von Benutzerberechtigungen in großem Umfang mit AWS IAM Identity Center)

- Mastering identity at every layer of the cake

Ähnliche Beispiele:

- Übung: IAM Tag-basierte Zugriffskontrolle für EC2

SICH 3 Wie verwalten Sie Berechtigungen für Personen und Maschinen?

Verwalten Sie Berechtigungen zum Steuern des Zugriffs auf Personen- und Maschinenidentitäten, die Zugriff auf AWS und Ihren Workload benötigen. Berechtigungen steuern, wer worauf und unter welchen Bedingungen zugreifen kann.

Bewährte Methoden

- SEC03-BP01 Definieren von Zugriffsanforderungen
- SEC03-BP02 Gewähren des Zugriffs mit den geringsten Berechtigungen
- SEC03-BP03 Einrichtung eines Notfallzugriffprozesses
- SEC03-BP04 Kontinuierliche Reduzierung der Berechtigungen
- SEC03-BP05 Definieren eines Integritätsschutzes für Berechtigungen in Ihrer Organisation
- SEC03-BP06 Zugriffsverwaltung basierend auf dem Lebenszyklus
- SEC03-BP07 Analysieren des öffentlichen und kontoübergreifenden Zugriffs
- SEC03-BP08 Sicheres gemeinsames Nutzen von Ressourcen

SEC03-BP01 Definieren von Zugriffsanforderungen

Administratoren, Endbenutzer oder andere Komponenten müssen auf jede Komponente oder Ressource Ihres Workloads zugreifen. Sie müssen eine klare Definition davon haben, wer oder was Zugriff auf die einzelnen Komponenten haben soll. Anschließend wählen Sie den entsprechenden Identitätstyp und die entsprechende Authentifizierungs- und Autorisierungsmethode aus.

Typische Anti-Muster:

- Hartkodierung oder Speicherung von geheimen Daten in Ihrer Anwendung
- Gewähren individueller Berechtigungen für alle Nutzer
- Verwendung langlebiger Anmeldeinformationen

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Administratoren, Endbenutzer oder andere Komponenten müssen auf jede Komponente oder Ressource Ihres Workloads zugreifen. Sie müssen eine klare Definition davon haben, wer oder was Zugriff auf die einzelnen Komponenten haben soll. Anschließend wählen Sie den entsprechenden Identitätstyp und die entsprechende Authentifizierungs- und Autorisierungsmethode aus.

Regulärer Zugriff auf AWS-Konten in der Organisation sollte per Verbundzugriff oder einen zentralen Identitätsanbieter bereitgestellt werden. Sie sollten auch Ihr Identitätsmanagement zentralisieren und sicherstellen, dass es ein etabliertes Verfahren zur Integration des AWS-Zugriffs in den Zugriffslebenszyklus der Mitarbeiter gibt Wenn beispielsweise ein Mitarbeiter in eine Rolle mit einer anderen Zugriffsstufe wechselt, sollte sich auch dessen Gruppenmitgliedschaft so ändern, dass die neuen Zugriffsanforderungen berücksichtigt werden.

Legen Sie bei der Definition der Zugriffsanforderungen für nicht menschliche Identitäten fest, welche Anwendungen und Komponenten Zugriff benötigen und wie die Berechtigungen gewährt werden. Eine empfohlene Vorgehensweise ist die Verwendung von nach dem Modell der geringsten Berechtigung entwickelten IAM-Rollen. AWS-verwaltete Richtlinien bieten vordefinierte IAM-Richtlinien für die meisten typischen Anwendungsfälle.

AWS-Services wie beispielsweise AWS Secrets Manager und AWS Systems Manager Parameter Storekönnen dabei helfen, Secrets in sicherer Weise von Anwendungen oder Workloads zu trennen, wenn es nicht möglich ist, IAM-Rollen zu verwenden. In Secrets Manager können Sie die automatische Rotation Ihrer Anmeldeinformationen einrichten. Mit Systems Manager können Sie auf Parameter in Ihren Skripts, Befehlen, SSM-Dokumenten, Konfigurations- und Automatisierungsworkflows verweisen, indem Sie den bei der Erstellung des Parameters angegebenen eindeutigen Namen verwenden.

Sie können AWS Identity and Access Management Roles Anywhere verwenden, um temporäre Sicherheitsanmeldeinformationen in IAM für Workloads zu erhalten, die außerhalb von AWS ausgeführt werden. Ihre Workloads können dieselben IAM-Richtlinien und IAM-Rollen verwenden, die Sie für AWS-Anwendungen zum Zugriff auf AWS-Ressourcen nutzen.

Verwenden Sie nach Möglichkeit kurzfristige temporäre anstelle langfristiger statischer Anmeldeinformationen. Verwenden Sie für Szenarien, in denen Sie IAM-Nutzer mit programmatischem Zugriff und langfristigen Anmeldeinformationen benötigen, Informationen über die letzte Nutzung von Zugriffsschlüsseln, um Zugriffsschlüssel zu entfernen und zu rotieren.

Ressourcen

Zugehörige Dokumente:

- Attributbasierte Zugriffskontrolle (ABAC)
- AWS IAM Identity Center
- IAM Roles Anywhere
- AWS-verwaltete Richtlinien für IAM Identity Center
- AWS-IAM-Richtlinienbedingungen
- IAM-Anwendungsfälle
- Entfernen von nicht benötigten Anmeldeinformationen
- Arbeiten mit Richtlinien
- Steuerung des Zugriffs auf AWS-Ressourcen auf der Grundlage von AWS-Konto, OU oder Organisation
- Identifizieren, Arrangieren und Verwalten von geheimen Daten mithilfe der erweiterten Suche in AWS Secrets Manager

Zugehörige Videos:

- Become an IAM Policy Master in 60 Minutes or Less (Experte für IAM-Richtlinien in unter 60 Minuten)
- Separation of Duties, Least Privilege, Delegation, and CI/CD (Trennung von Pflichten, geringste Berechtigung, Delegierung und CI/CD)
- Streamlining identity and access management for innovation (Optimieren des Identitäts- und Zugriffsmanagements für Innovation)

SEC03-BP02 Gewähren des Zugriffs mit den geringsten Berechtigungen

Gewähren Sie nur den Zugriff, den Identitäten wirklich benötigen, indem Sie den Zugriff auf bestimmte Aktionen auf bestimmten AWS-Ressourcen unter bestimmten Bedingungen erlauben. Nutzen Sie Gruppen und Identitätsattribute, um Berechtigungen dynamisch in großem Umfang festzulegen, anstatt Berechtigungen für einzelne Benutzer zu definieren. Sie können beispielsweise einer Gruppe von Entwicklern den Zugriff erlauben, nur die Ressourcen für ihr Projekt zu verwalten. Wenn ein Entwickler aus der Gruppe entfernt wird, wird der Zugriff für den Entwickler überall

widerrufen, wo die Gruppe für die Zugriffskontrolle verwendet wurde, ohne dass Änderungen an den Zugriffsrichtlinien erforderlich sind.

Typische Anti-Muster:

- Standardmäßige Gewährung von Administratorberechtigungen für Benutzer
- Verwendung des Root-Kontos für alltägliche Aktivitäten

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Durch die Einführung des Prinzips der Minimal erforderlichen Berechtigungen stellen Sie sicher, dass Identitäten zum Erledigen einer Aufgabe nur die minimal erforderlichen Funktionen ausführen dürfen. Dabei wird ein ausgewogenes Verhältnis zwischen Nutzbarkeit und Effizienz geschaffen. Wenn Sie nach diesem Prinzip arbeiten, schränken Sie den unbeabsichtigten Zugriff ein und stellen sicher, dass Sie überprüfen können, wer Zugriff auf welche Ressourcen hat. In AWS haben Identitäten standardmäßig keine Berechtigungen, mit Ausnahme des Root-Benutzers. Die Anmeldeinformationen für den Root-Benutzer müssen eng kontrolliert und dürfen nur für einige wenige bestimmte Aufgaben verwendet werden.

Sie verwenden Richtlinien, um explizit Berechtigungen zu erteilen, die IAM- oder Ressourcen-Entitäten angefügt sind, z. B. eine IAM-Rolle, die von Verbundidentitäten oder -maschinen verwendet wird, oder Ressourcen (z. B. S3-Buckets). Wenn Sie eine Richtlinie erstellen und anfügen, können Sie die Serviceaktionen, Ressourcen und Bedingungen angeben, die erfüllt sein müssen, damit AWS den Zugriff erlauben kann. AWS unterstützt eine Vielzahl von Bedingungen, mit denen Sie den Zugriff einschränken können. Wenn Sie beispielsweise den `Bedingungsschlüssel` PrincipalOrgIDverwenden, wird die Kennung von AWS Organizations überprüft, sodass der Zugriff innerhalb Ihrer AWS-Organisation gewährt werden kann.

Sie können auch Anforderungen kontrollieren, die AWS-Services in Ihrem Namen stellen, wie das Erstellen einer AWS CloudFormation-Funktion durch AWS Lambda. Dazu verwenden Sie den `CalledVia` -Bedingungsschlüssel. Sie sollten unterschiedliche Richtlinientypen in Ebenen organisieren, um die Berechtigungen in einem Konto insgesamt effektiv zu begrenzen. So können Sie beispielsweise Ihren Anwendungsteams gestatten, ihre eigenen IAM-Richtlinien zu erstellen. Verwenden Sie aber eine Berechtigungsgrenze zur Begrenzung der maximalen Berechtigungen, die das Team gewähren kann.

Es gibt verschiedene AWS-Funktionen, die Ihnen bei der Skalierung des Berechtigungsmanagements und der Einhaltung des Prinzips der geringsten Berechtigungen helfen. Auf Attributen basierende Zugriffssteuerung ermöglicht die Begrenzung der Berechtigungen auf der Grundlage des Tags einer Ressource, um Autorisierungsentscheidungen je nach den der Ressource zugewiesenen Tags und dem aufrufenden IAM-Prinzipal zu treffen. Dadurch können Sie Ihre Tagging- und Ihre Berechtigungsrichtlinie kombinieren, um detailliert gesteuerte Ressourcenzugriffe zu ermöglichen, ohne dass dazu viele spezielle Richtlinien erforderlich sind.

Eine andere Möglichkeit zur Erstellung einer Richtlinie mit geringsten Berechtigungen besteht darin, sie nach der Ausführung einer Aktivität auf CloudTrail-Berechtigungen zu basieren. IAM Access Analyzer kann automatisch IAM-Richtlinien auf der Grundlage einer Aktivität generieren. Sie können auch IAM Access Advisor auf der Ebene der Organisation oder einzelner Konten verwenden, um die für eine bestimmte Richtlinie zuletzt verwendeten Informationen nachzuverfolgen.

Richten Sie regelmäßige Prüfungen dieser Details und einen Plan zum Entfernen nicht benötigter Berechtigungen ein. Sie sollten Integritätsschutz für Berechtigungen in Ihrer AWS-Organisation einrichten, um die Höchstzahl der Berechtigungen innerhalb eines Mitgliedskontos zu begrenzen. Services wie beispielsweise AWS Control Tower bieten präskriptive und verwaltete präventive Steuerungen und ermöglichen Ihnen die Definition Ihrer eigenen Steuerungen.

Ressourcen

Zugehörige Dokumente:

- Berechtigungsgrenzen für IAM-Entitäten
- Techniken zum Erstellen von IAM-Richtlinien mit geringsten Berechtigungen
- IAM Access Analyzer erleichtert die Implementierung geringster Berechtigungen durch die Generierung von IAM-Richtlinien auf der Grundlage der Zugriffsaktivitäten
- Verfeinern der Berechtigungen mithilfe der zuletzt genutzten Informationen
- IAM-Richtlinienarten und wann sie verwendet werden sollten
- Testen von IAM-Richtlinien mit dem IAM-Richtliniensimulator
- Integritätsschutz in AWS Control Tower
- Zero-Trust-Architekturen: Eine AWS-Perspektive
- Implementieren des Prinzips der geringsten Berechtigung mit CloudFormation StackSets

Zugehörige Videos:

- Next-generation permissions management (Berechtigungsmanagement der nächsten Generation)
- Zero Trust: An AWS perspective (Zero Trust: Eine AWS-Perspektive)
- How can I use permissions boundaries to limit IAM users and roles to prevent privilege escalation? (Wie kann ich mit Berechtigungsgrenzen IAM-Benutzer und -Rollen einschränken, um die Eskalation von Berechtigungen zu vermeiden?)

Zugehörige Beispiele:

- Übung: IAM-Berechtigungsgrenzen – Übertragung der Rollenerstellung

SEC03-BP03 Einrichtung eines Notfallzugriffprozesses

Ein Prozess, der den Notfallzugriff auf Ihren Workload im unwahrscheinlichen Fall eines automatisierten Prozesses oder eines Pipeline-Problems ermöglicht. Auf diese Weise können Sie den Zugriff mit der geringsten Berechtigung nutzen, aber sicherstellen, dass Benutzer bei Bedarf die richtige Zugriffsebene erhalten. Richten Sie beispielsweise einen Prozess ein, mit dem Administratoren die Anfrage prüfen und genehmigen, z. B. eine kontoübergreifende AWS-Rolle für den Zugriff im Notfall. Alternativ können Sie ein spezifisches Verfahren festlegen, das Administratoren zur Validierung und Genehmigung einer Notfallanfrage befolgen müssen.

Typische Anti-Muster:

- Fehlen eines Notfallprozesses für die Wiederherstellung nach einem Ausfall mit Ihrer vorhandenen Identitätskonfiguration
- Gewähren langfristiger erhöhter Berechtigungen für Fehlerbehebungs- oder Wiederherstellungszwecke

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Mittel

Implementierungsleitfaden

Die Einrichtung eines Notfallzugriffs kann verschiedene Formen haben, auf die Sie vorbereitet sein sollten. Die erste davon ist der Ausfall Ihres primären Identitätsanbieters. Für diesen Fall benötigen Sie ein zweites Zugriffsverfahren mit den für die Wiederherstellung erforderlichen Berechtigungen. Dieses Verfahren kann ein weiterer Identitätsanbieter oder ein IAM-Benutzer sein. Dieses zweite Verfahren muss eng kontrolliert und überwacht werden und bei Verwendung eine Benachrichtigung ausgeben. Die Identität für den Notfallzugriff sollte von einem Konto stammen, das speziell diesem

Zweck dient, und nur über die Berechtigungen verfügen, die erforderlich sind, um eine Rolle für die Wiederherstellung anzunehmen.

Weiterhin sollten Sie auf den Notfallzugriff vorbereitet sein, wo erhöhte administrative Zugriffsberechtigungen erforderlich sind. Ein typisches Szenario besteht darin, Änderungsberechtigungen auf einen automatisierten Prozess für die Bereitstellung von Änderungen zu beschränken. Wenn bei diesem Prozess ein Problem auftritt, müssen Nutzer möglicherweise erhöhte Berechtigungen anfragen, um die Funktionalität wiederherstellen zu können. Richten Sie dafür einen Prozess ein, bei dem Nutzer erhöhte Zugriffsberechtigungen anfragen und Administratoren diese prüfen und genehmigen können. Die Implementierungspläne, die die bewährten Methoden für die Vorab-Bereitstellung von Zugriff und die Einrichtung von Notfall-, „Break Glass“-Rollen enthalten, werden bereitgestellt im Rahmen von SEC10-BP05 Vorab bereitgestellter Zugriff.

Ressourcen

Zugehörige Dokumente:

- Überwachen und Benachrichtigen auf AWS
- Verwalten vorübergehend erhöhter Zugriffsberechtigungen

Zugehöriges Video:

- Become an IAM Policy Master in 60 Minutes or Less (Experte für IAM-Richtlinien in unter 60 Minuten)

SEC03-BP04 Kontinuierliche Reduzierung der Berechtigungen

Wenn Teams und Workloads bestimmen, welchen Zugriff sie benötigen, entfernen Sie Berechtigungen, die sie nicht mehr verwenden, und erstellen Sie Überprüfungsprozesse, um Berechtigungen mit den geringsten Berechtigungen zu erzielen. Überwachen und reduzieren Sie kontinuierlich ungenutzte Identitäten und Berechtigungen.

Wenn Teams erst mit der Zusammenarbeit begonnen haben und Projekte gerade erst starten, können Sie sich entscheiden, einen umfassenden Zugang (in einer Entwicklungs- oder Testumgebung) zu gewähren, um Innovation und Agilität zu fördern. Dabei sollten Sie den Zugriff ständig überprüfen und, insbesondere bei Produktionsumgebungen, den Zugriff auf die erforderlichen Berechtigungen einschränken und das Prinzip der geringsten Berechtigungen einhalten. AWS

bietet Zugriffsanalysefunktionen, mit denen Sie ungenutzte Zugriffe identifizieren können. Um Sie dabei zu unterstützen, ungenutzte Benutzer, Rollen, Berechtigungen und Anmeldeinformationen zu identifizieren, analysiert AWS die Zugriffsaktivitäten und stellt Informationen zu Zugriffsschlüsseln und zuletzt verwendeten Rollen bereit. Sie können den Zeitstempel des letzten Zugriffs verwenden, um ungenutzte Benutzer und Rollen zu identifizierenund sie zu entfernen. Darüber hinaus können Sie die Informationen zum letzten Service- und Aktionszugriff überprüfen, um Berechtigungen für bestimmte Benutzer und Rollen zu identifizieren und enger zu fassen. Sie können beispielsweise Informationen zum letzten Zugriff verwenden, um die spezifischen Amazon Simple Storage Service-Aktionen (Amazon S3) zu identifizieren, die Ihre Anwendungsrolle erfordert, und den Zugriff auf diese beschränken. Diese Funktion ist in der AWS Management Console und programmgesteuert verfügbar, damit Sie sie in Ihre Infrastruktur-Workflows und automatisierten Tools integrieren können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Konfigurieren Sie AWS Identify and Access Management (IAM) Access Analyzer: AWS IAM Access Analyzer hilft Ihnen, die Ressourcen in Ihrer Organisation und Ihren Konten zu identifizieren, z. B. Amazon Simple Storage Service-Buckets (Amazon S3) oder IAM-Rollen, die mit einer externen Entität geteilt werden.
  - AWS IAM Access Analyzer

Ressourcen

Zugehörige Dokumente:

- Attributbasierte Zugriffskontrolle (ABAC)
- Gewähren von geringsten Rechten
- Entfernen von nicht benötigten Anmeldeinformationen
- Arbeiten mit Richtlinien

Relevante Videos:

- Become an IAM Policy Master in 60 Minutes or Less (Experte für IAM-Richtlinien in unter 60 Minuten)
- Separation of Duties, Least Privilege, Delegation, and CI/CD (Trennung von Pflichten, geringste Berechtigung, Delegierung und CI/CD)

SEC03-BP05 Definieren eines Integritätsschutzes für Berechtigungen in Ihrer Organisation

Richten Sie allgemeine Kontrollen ein, die den Zugriff auf alle Identitäten in Ihrer Organisation einschränken. Sie können beispielsweise den Zugriff auf bestimmte AWS-Regionen einschränken oder verhindern, dass Ihre Bediener gemeinsame Ressourcen löschen, z. B. eine IAM-Rolle, die für Ihr zentrales Sicherheitsteam verwendet wird.

Typische Anti-Muster:

- Ausführen von Workloads in Ihrem Organisationsadministrator-Konto
- Ausführen von Produktions- und Nicht-Produktionsworkloads im selben Konto

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Mittel

Implementierungsleitfaden

Wenn Sie im Zuge Ihres Wachstums zusätzliche Workloads in AWS verwalten, sollten Sie diese Workloads mithilfe von Konten trennen und die Konten mit AWS Organizations verwalten. Wir empfehlen, allgemeinen Integritätsschutz für Berechtigungen einzurichten, der den Zugriff auf alle Identitäten in Ihrer Organisation einschränkt. Sie können beispielsweise den Zugriff auf bestimmte AWS-Regionen einschränken oder verhindern, dass Mitglieder Ihres Teams gemeinsame Ressourcen löschen, z. B. eine IAM-Rolle, die vom zentralen Sicherheitsteam verwendet wird.

Sie können beginnen, indem Sie Beispiel-Servicekontrollrichtlinien implementieren, die beispielsweise verhindern, dass Benutzer wichtige Services deaktivieren. SCPs verwenden die IAM-Richtliniensprache und ermöglichen Ihnen, Kontrollen einzurichten, die alle IAM-Prinzipale (Benutzer und Rollen) einhalten müssen. Sie können den Zugriff auf bestimmte Serviceaktionen und Ressourcen oder basierend auf bestimmten Bedingungen einschränken, um die Zugriffskontrollanforderungen Ihrer Organisation zu erfüllen. Falls erforderlich, können Sie Ausnahmen zum Integritätsschutz definieren. Sie können beispielsweise Serviceaktionen für alle IAM-Entitäten im Konto mit Ausnahme einer bestimmten Administratorrolle einschränken.

Wir empfehlen, die Ausführung von Workloads in Ihrem Veraltungskonto zu vermeiden. Das Verwaltungskonto sollte für den Einsatz und die Bereitstellung von Integritätsschutz für die Sicherheit verwendet werden, der sich auf Mitgliedskonten auswirkt. Manche AWS-Services unterstützen die Verwendung eines delegierten Administratorkontos. Wenn ein solches delegiertes Konto verfügbar ist, sollten Sie es anstelle des Verwaltungskontos verwenden. Sie sollten den Zugriff auf das Organisationsadministratorkonto strengstens einschränken.

Die Verwendung einer Mehrkonten-Strategie ermöglicht größere Flexibilität bei der Anwendung von Integritätsschutz auf Ihre Workloads. Die AWS Security Reference Architecture bietet präskriptive Anleitungen zur Gestaltung Ihrer Kontenstruktur. AWS-Services wie AWS Control Tower bieten Funktionen für die zentrale Verwaltung präventiver und erkennender Kontrollen in ihrer Organisation. Definieren Sie für jedes Konto bzw. jede OU in Ihrer Organisation einen klaren Zweck und schränken Sie die Steuerungen entsprechend diesem Zweck ein.

Ressourcen

Zugehörige Dokumente:

- AWS Organizations
- Service-Kontrollrichtlinien (SCPs),
- Bessere Nutzung von Servicekontrollrichtlinien in einer Mehrkontenumgebung
- AWS Security Reference Architecture (AWS SRA)

Zugehörige Videos:

- Enforce Preventive Guardrails using Service Control Policies (Durchsetzung von präventivem Integritätsschutz mit Servicekontrollrichtlinien)
- Building governance at scale with AWS Control Tower (Governance in großem Umfang mit AWS Control Tower)
- AWS Identity and Access Management deep dive (Tiefer Einblick in AWS Identity and Access Management)

SEC03-BP06 Zugriffsverwaltung basierend auf dem Lebenszyklus

Integrieren Sie Zugriffskontrollen in den Operator- und Anwendungslebenszyklus sowie Ihren zentralen Verbundanbieter. Entfernen Sie beispielsweise den Zugriff eines Benutzers, wenn er die Organisation verlässt oder eine andere Rolle übernimmt.

Wenn Sie Workloads mit separaten Konten verwalten, müssen Sie Ressourcen für diese Konten freigeben. Wir empfehlen, dass Sie Ressourcen mit AWS Resource Access Manager (AWS RAM). Mit diesem Service können Sie AWS-Ressourcen einfach und sicher innerhalb Ihrer AWS Organizations-Organisation und -Organisationseinheiten freigeben. Mithilfe von AWS RAM wird der Zugriff auf gemeinsam genutzte Ressourcen automatisch gewährt oder widerrufen, wenn Konten in die Organisation oder Organisationseinheit verschoben werden, für die sie freigegeben sind. Auf

diese Weise können Sie sicherstellen, dass Ressourcen nur für die Konten freigegeben werden, die Sie beabsichtigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Verwenden eines Lebenszyklus für den Benutzerzugriff: Implementieren Sie eine Lebenszyklusrichtlinie für den Benutzerzugriff für neue Benutzer, Änderungen von Zuständigkeiten und das Ausscheiden von Benutzern, um sicherzustellen, dass nur aktuelle Benutzer Zugriff haben.

Ressourcen

Zugehörige Dokumente:

- Attributbasierte Zugriffskontrolle (ABAC)
- Gewähren von geringsten Rechten
- IAM Access Analyzer
- Entfernen von nicht benötigten Anmeldeinformationen
- Arbeiten mit Richtlinien

Relevante Videos:

- Become an IAM Policy Master in 60 Minutes or Less (Experte für IAM-Richtlinien in unter 60 Minuten)
- Separation of Duties, Least Privilege, Delegation, and CI/CD (Trennung von Pflichten, geringste Berechtigung, Delegierung und CI/CD)

SEC03-BP07 Analysieren des öffentlichen und kontoübergreifenden Zugriffs

Überwachen Sie kontinuierlich Ergebnisse, die den öffentlichen und kontoübergreifenden Zugriff betreffen. Beschränken Sie den öffentlichen und kontoübergreifenden Zugriff ausschließlich auf Ressourcen, die diese Art von Zugriff benötigen.

Typische Anti-Muster:

- Fehlen eines Prozesses für die Kontrolle des kontoübergreifenden und öffentlichen Zugriffs auf Ressourcen

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Niedrig

Implementierungsleitfaden

In AWS können Sie Zugriff auf Ressourcen in einem anderen Konto gewähren. Sie gewähren direkten kontoübergreifenden Zugriff über an Ressourcen angefügte Richtlinien (zum Beispiel Amazon Simple Storage Service (Amazon S3)-Bucket-Richtlinien) oder indem Sie zulassen, dass eine Identität eine IAM-Rolle in einem anderen Konto annimmt. Prüfen Sie bei der Verwendung von Ressourcenrichtlinien, dass der Zugriff Identitäten in Ihrer Organisation gewährt wird und dass Sie die Ressourcen wirklich öffentlich machen wollen. Definieren Sie einen Prozess für die Genehmigung aller Ressourcen, die öffentlich verfügbar sein müssen.

IAM Access Analyzer verwendet nachweisbare Sicherheit , um alle Zugriffspfade zu einer Ressource von außerhalb ihres Kontos zu identifizieren. Dieser Service überprüft Ressourcenrichtlinien kontinuierlich und meldet Ergebnisse des öffentlichen und kontoübergreifenden Zugriffs, um Ihnen die Analyse potenziell umfassender Zugriffe zu erleichtern. Ziehen Sie die Konfiguration von IAM Access Analyzer mit AWS Organizations in Betracht, um Transparenz für alle Ihre Konten zu gewährleisten. IAM Access Analyzer ermöglicht Ihnen auch die Voranzeige von Access-Analyzer-Ergebnissen, bevor Sie Ressourcenberechtigungen bereitstellen. So können Sie sicherstellen, dass mit den Richtlinienänderungen nur der beabsichtigte öffentliche und kontoübergreifende Zugriff auf Ihre Ressourcen gewährt wird. Bei der Arbeit für den Mehrkonten-Zugriff können Sie Vertrauensrichtlinien verwenden, um zu steuern, in welchen Fällen eine Rolle angenommen werden kann. So könnten Sie beispielsweise die Annahme von Rollen auf einen bestimmten Quell-IP-Bereich einschränken.

Sie können auch AWS Config verwenden, um Ressourcen für versehentliche Konfigurationen mit öffentlichem Zugriff durch AWS Config-Richtlinienprüfungen zu melden und zu korrigieren. Services wie AWS Control Tower und AWS Security Hub vereinfachen die Bereitstellung von Prüfungen und Integritätsschutz über eine AWS Organizations hinweg, um öffentlich zugängliche Ressourcen zu identifizieren und zu korrigieren. Beispielsweise verfügt AWS Control Tower über verwalteten Integritätsschutz, der erkennen kann, ob Amazon EBS-Snapshots von allen AWS-Konten wiederhergestellt werden können.

Ressourcen

Zugehörige Dokumente:

- Verwendung von AWS Identity and Access Management Access Analyzer
- Integritätsschutz in AWS Control Tower
- AWS Foundational Security Best Practices Standard

- AWS Config Managed Rules
- AWS Trusted Advisor-Prüfungsreferenz

Zugehörige Videos:

- Best Practices for securing your multi-account environment (Bewährte Methoden für den Schutz Ihrer Mehrkonten-Umgebung)
- Dive Deep into IAM Access Analyzer (Tiefer Einblick in IAM Access Analyzer)

SEC03-BP08 Sicheres gemeinsames Nutzen von Ressourcen

Steuern Sie die Nutzung gemeinsam genutzter Ressourcen über Konten oder innerhalb Ihrer AWS Organizations. Überwachen Sie gemeinsam genutzte Ressourcen und überprüfen Sie den Zugriff auf gemeinsame Ressourcen.

Typische Anti-Muster:

- Verwendung der standardmäßigen IAM-Vertrauensrichtlinie bei der Gewährung kontoübergreifenden Zugriffs für Drittparteien

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Niedrig

Implementierungsleitfaden

Wenn Sie Ihre Workloads mit mehreren AWS-Konten verwalten, müssen Sie möglicherweise Ressourcen von mehreren Konten verwenden lassen. Dies beinhaltet oft die kontoübergreifende Kontofreigabe innerhalb eines AWS Organizations. Verschiedene AWS-Services wie etwa AWS Security Hub, Amazon GuardDutyund AWS Backup verfügen über in Organizations integrierte kontoübergreifende Funktionen. Sie können AWS Resource Access Manager verwenden, um übliche Ressourcen gemeinsam zu nutzen, wie beispielsweise VPC-Subnetze oder Transit-Gateway-Anhänge, AWS Network Firewalloder Amazon SageMaker Runtime-Pipelines. Wenn Sie sicherstellen möchten, dass Ihr Konto nur innerhalb Ihrer Organizations Ressourcen teilt, empfehlen wir die Verwendung von Service-Kontrollrichtlinien (SCPs), um den Zugriff auf externe Prinzipale zu verhindern.

Wenn Sie Ihre Ressourcen teilen, sollten Sie Maßnahmen treffen, um sie gegen unbeabsichtigte Zugriffe zu schützen. Wir empfehlen die Kombination von identitätsbasierten Kontrollen und

Netzwerkkontrollen zur Erstellung eines Datenperimeters für Ihre Organisation.. Diese Kontrollen müssen streng begrenzen, welche Ressourcen gemeinsam genutzt werden, und die gemeinsame Nutzung oder Offenlegung aller anderen Ressourcen verhindern. Sie könnten beispielsweise als Teil Ihres Datenperimeters VPC-Endpunktrichtlinien und die Bedingung `aws:PrincipalOrgId` verwenden, um sicherzustellen, dass die Identitäten, die auf Ihre Amazon S3-Buckets zugreifen, zu Ihrer Organisation gehören.

In manchen Fällen kann es vorkommen, dass Sie Ressourcen außerhalb Ihrer Organizations freigeben oder Drittparteien den Zugriff auf Ihr Konto gewähren möchten. So könnte etwa ein Partner eine Überwachungslösung bereitstellen, die auf Ressourcen in Ihrem Konto zugreifen muss. In solchen Fällen können Sie eine kontoübergreifende IAM-Rolle erstellen, die nur über die von der Drittpartei benötigten Berechtigungen verfügt. Sie sollten auch mithilfe der externen ID-Bedingung eine Vertrauensrichtlinie erstellen. Wenn Sie eine externe ID verwenden, sollten Sie für jede Drittpartei eine eindeutige ID erstellen. Die eindeutige ID sollte nicht von der Drittpartei bereitgestellt oder kontrolliert werden. Wenn die Drittpartei den Zugriff auf Ihre Umgebung nicht mehr benötigt, sollten Sie die Rolle entfernen. Sie sollten darüber hinaus unter allen Umständen die Bereitstellung langfristiger IAM-Anmeldeinformationen für Drittparteien vermeiden. Achten Sie auf andere AWS-Services, die die gemeinsame Nutzung in nativer Weise unterstützen. Beispielsweise ermöglicht das AWS Well-Architected Tool die gemeinsame Nutzung von Workloads mit anderen AWS-Konten.

Bei Verwendung von Services wie Amazon S3 wird empfohlen, ACLs für Ihren Amazon S3-Bucket zu deaktivieren und IAM-Richtlinien zur Festlegung der Zugriffskontrolle zu verwenden. Zur Einschränkung des Zugriffs auf einen Amazon S3-Ursprung von Amazon CloudFrontaus migrieren Sie von der Ursprungszugriffsidentität (OAI) zur Ursprungszugriffssteuerung (OAC), die zusätzliche Funktionen wie beispielsweise die serverseitige Verschlüsselung mit AWS KMS.

Ressourcen

Zugehörige Dokumente:

- Bucket-Besitzer gewährt kontoübergreifende Berechtigung für Objekte, die er nicht besitzt
- Verwendung von Vertrauensrichtlinien mit IAM
- Erstellen von Datenperimetern auf AWS
- Verwenden einer externen ID, um Dritten Zugriff auf Ihre AWS-Ressourcen zu gewähren

Zugehörige Videos:

- Granular Access with AWS Resource Access Manager (Granulärer Zugriff mit AWS Resource Access Manager)
- Securing your data perimeter with VPC endpoints (Schutz Ihres Datenperimeters mit VPC-Endpunkten)
- Establishing a data perimeter on AWS (Einrichten eines Datenperimeters auf AWS)

Erkennung

Frage

- SICH 4 Wie erkennen und untersuchen Sie Sicherheitsereignisse?

SICH 4 Wie erkennen und untersuchen Sie Sicherheitsereignisse?

Erfassen und analysieren Sie Ereignisse mithilfe von Protokollen und Kennzahlen, um Einblick zu erhalten. Ergreifen Sie Maßnahmen bei Sicherheitsereignissen und potenziellen Bedrohungen, um Ihren Workload zu schützen.

Bewährte Methoden

- SEC04-BP01 Konfigurieren der Service- und Anwendungsprotokollierung
- SEC04-BP02 Zentrale Analyse von Protokollen, Ergebnissen und Metriken
- SEC04-BP03 Automatisierte Reaktion auf Ereignisse
- SEC04-BP04 Implementieren von umsetzbaren Sicherheitsereignissen:

SEC04-BP01 Konfigurieren der Service- und Anwendungsprotokollierung

Konfigurieren Sie die Protokollierung in der gesamten Workload, einschließlich Anwendungsprotokolle, Ressourcenprotokolle und AWS-Serviceprotokolle. Stellen Sie beispielsweise sicher, dass AWS CloudTrail, Amazon CloudWatch Logs, Amazon GuardDuty und AWS Security Hub für alle Konten in Ihrer Organisation aktiviert sind.

Eine grundlegende Vorgehensweise besteht darin, eine Reihe von Erkennungsmechanismen auf Kontoebene einzurichten. Diese grundlegenden Mechanismen dienen der Aufzeichnung und Erkennung einer Vielzahl von Aktionen für alle Ressourcen in Ihrem Konto. Sie ermöglichen es Ihnen, eine umfassende Aufklärungsfunktion mit Optionen wie automatisierten Korrekturen und Partnerintegrationen zu erstellen, um Funktionen hinzuzufügen.

In AWS sind folgende Services in dieser Basisgruppe enthalten:

- AWS CloudTrail stellt den Ereignisverlauf Ihrer AWS-Kontoaktivität bereit, einschließlich Aktionen über die AWS Management Console, AWS SDKs, Befehlszeilentools und andere AWS-Services.
- AWS Config überwacht und zeichnet Ihre AWS-Ressourcenkonfigurationen auf. Darüber hinaus ermöglicht es Ihnen, die Auswertung und Korrektur der gewünschten Konfigurationen zu automatisieren.
- Amazon GuardDuty ist ein Service zur Bedrohungserkennung, der Ihre AWS-Konten und -Workloads zu deren Schutz fortlaufend auf böswillige oder unbefugte Verhaltensweisen überwacht.
- AWS Security Hub bietet einen zentralen Ort, an dem Ihre Sicherheitswarnungen oder Ergebnisse von mehreren AWS-Services und optionalen Produkten von Drittanbietern aggregiert, organisiert und priorisiert werden. So erhalten Sie einen umfassenden Überblick über Sicherheitswarnungen und den Compliance-Status.

Aufbauend auf der Grundlage der Kontoebene, bieten viele wichtige AWS-Services, z. B. Amazon Virtual Private Cloud Console (Amazon VPC)Protokollierungsfunktionen auf Service-Ebene. Amazon VPC Flow Logs ermöglichen es Ihnen, Informationen über den IP-Datenverkehr zu und von Netzwerkschnittstellen zu erfassen, die wertvolle Einblicke in den Konnektivitätsverlauf bieten und automatisierte Aktionen basierend auf ungewöhnlichem Verhalten auslösen können.

Für Amazon Elastic Compute Cloud (Amazon EC2)-Instances und anwendungsbasierte Protokollierung, die nicht von AWS-Services stammen, besteht die Möglichkeit zur Speicherung und Analyse von Protokollen mit Amazon CloudWatch Logs. Eine Agent sammelt die Protokolle vom Betriebssystem und den ausgeführten Anwendungen und speichert sie automatisch. Sobald die Protokolle in CloudWatch Logs verfügbar sind, können Sie sie in Echtzeit verarbeitenoder mithilfe von CloudWatch Logs Insights.

Neben dem Erfassen und Aggregieren von Protokollen ist auch das Extrahieren aussagekräftiger Informationen aus dem enormen Umfang an Protokollen und Ereignisdaten wichtig, die von modernen, komplexen Architekturen generiert werden. Weitere Informationen finden Sie auf der Registerkarte Überwachung im Whitepaper zur Säule der Zuverlässigkeit . Protokolle können selbst sensible Daten enthalten, entweder wenn Anwendungsdaten fälschlicherweise den Weg in Protokolldateien gefunden haben, die der CloudWatch Logs-Agent erfasst, oder wenn die regionsübergreifende Protokollierung für die Protokollaggregation konfiguriert ist und es gesetzliche Vorgaben zum grenzüberschreitenden Versand bestimmter Informationen gibt.

Eine Möglichkeit besteht darin, AWS Lambda-Funktionen zu nutzen, welche von Protokollen angestoßen werden. So können Protokolldaten gefiltert und verkleinert werden, bevor sie an einen zentralen Protokollierungsstandort weitergeleitet werden, z. B. einen Amazon Simple Storage Service (Amazon S3)-Bucket. Die nicht bearbeiteten Protokolle können in einem lokalen Bucket aufbewahrt werden, bis eine „angemessene Zeit“ vergangen ist (wie von der Gesetzgebung und Ihrem Rechtsteam festgelegt). Ab diesem Zeitpunkt kann eine Amazon S3-Lebenszyklusregel sie automatisch löschen. Protokolle können in Amazon S3 weiter geschützt werden, indem Sie Amazon S3 Object Lockwo Sie Objekte mit einem WORM-Modell (Write-Once-Read-Many) speichern können.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Aktivierung der Protokollierung von AWS-Services: Aktivieren Sie die Protokollierung von AWS-Services entsprechend Ihren Anforderungen. Die Protokollierungsfunktionen umfassen Folgendes: Amazon VPC Flow Logs, Elastic Load Balancing (ELB)-Protokolle, Amazon S3-Bucket-Protokolle, CloudFront-Zugriffsprotokolle, Amazon Route 53-Abfrageprotokolle und Amazon Relational Database Service (Amazon RDS)-Protokolle.
  - AWS Answers: Native AWS-Sicherheitsprotokollierungsfunktionen
- Bewerten und aktivieren Sie die Protokollierung von betriebssystem- und anwendungsspezifischen Protokollen, um verdächtiges Verhalten zu erkennen.
  - Erste Schritte mit CloudWatch Logs
  - Entwicklertools und Protokollanalyse
- Angemessene Kontrollen für Protokolle anwenden: Protokolle können vertrauliche Informationen enthalten und nur autorisierte Benutzer sollten Zugriff darauf haben. Erwägen Sie, die Berechtigungen auf Amazon S3-Buckets und CloudWatch Logs-Protokollgruppen einzuschränken.
  - Authentifizierung und Zugriffskontrolle für Amazon CloudWatch
  - Identitäts- und Zugriffsverwaltung in Amazon S3
- Konfigurieren Amazon GuardDuty: GuardDuty ist ein Service zur Bedrohungserkennung, der Ihre AWS-Konten und AWS-Workloads zu deren Schutz fortlaufend auf böswillige oder unbefugte Verhaltensweisen überwacht. Aktivieren Sie GuardDuty und konfigurieren Sie automatisierte Warnungen für E-Mails mithilfe der Übung.
- Konfigurieren eines benutzerdefinierten Prüfprotokolls in CloudTrail: Durch das Konfigurieren eines Prüfprotokolls können Sie Protokolle über den Standardzeitraum hinaus speichern und analysieren.
- Aktivieren AWS Config: AWS Config bietet Ihnen einen detaillierten Überblick über die Konfiguration der AWS-Ressourcen in Ihrem AWS-Konto. Hierzu zählt auch, wie die Ressourcen

zueinander in Verbindung stehen und wie sie in der Vergangenheit konfiguriert wurden. So können Sie erkennen, wie sich die Konfigurationen und Beziehungen mit der Zeit ändern.

- Aktivieren AWS Security Hub: Security Hub bietet Ihnen einen umfassenden Überblick über Ihren Sicherheitsstatus in AWS und hilft Ihnen, Ihre Compliance mit den Standards und Best Practices der Sicherheitsbranche zu überprüfen. Security Hub erfasst Sicherheitsdaten von allen AWS-Konten, AWS-Services und unterstützten Produkten von Drittanbieterpartnern und hilft Ihnen, Ihre Sicherheitstrends zu analysieren und die Sicherheitsprobleme mit der höchsten Priorität zu identifizieren.

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch
- Amazon EventBridge
- Erste Schritte: Amazon CloudWatch Logs
- Partnerlösungen im Bereich Sicherheit: Protokollierung und Überwachung

Zugehörige Videos:

- Best Practices for Centrally Monitoring Resource Configuration and Compliance (Best Practices für die zentrale Überwachung der Ressourcenkonfiguration und Compliance)
- Remediating Amazon GuardDuty and AWS Security Hub Findings (Korrektur von Amazon GuardDuty- und AWS Security Hub-Feststellungen)
- Threat management in the cloud: Amazon GuardDuty and AWS Security Hub (Bedrohungsmanagement in der Cloud: Amazon GuardDuty und AWS Security Hub)

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung von aufdeckenden Kontrollen

SEC04-BP02 Zentrale Analyse von Protokollen, Ergebnissen und Metriken

Sicherheitsteams benötigen Protokolle und Suchtools, um potenziell interessante Ereignisse zu erkennen, die auf unbefugte Aktivitäten oder unbeabsichtigte Änderungen hinweisen können. Um mit den enormen Informationsmengen komplexer Architekturen Schritt zu halten, reicht es jedoch nicht

aus, erfasste Daten einfach zu analysieren und die Informationen manuell zu verarbeiten. Nur mittels Analyse und Berichterstellung lassen sich nicht die richtigen Ressourcen zuweisen, um ein Ereignis zeitnah zu bearbeiten.

Zur Erstellung eines kompetenten Sicherheitsteams hat es sich bewährt, den Fluss von Sicherheitsereignissen und -ergebnissen tief in ein Benachrichtigungs- und Workflow-System zu integrieren. Dies kann beispielsweise ein Ticketsystem, ein Bug- oder Fehlersystem oder ein anderes Security Information and Event Management-System (SIEM) sein. Der Workflow wird dadurch aus E-Mail-Berichten und statischen Berichten genommen, sodass Sie Ereignisse oder Ergebnisse weiterleiten, eskalieren und verwalten können. Viele Organisationen integrieren mittlerweile Sicherheitsbenachrichtigungen in ihre Chat- oder Zusammenarbeitsplattformen und in ihre Plattformen für Entwicklerproduktivität. Für Organisationen, die die Automatisierung einführen, bietet ein API-gesteuertes Ticketing-System mit geringer Latenz erhebliche Flexibilität bei der Planung, vor allem in Bezug darauf, was zuerst automatisiert werden soll.

Diese bewährte Methode gilt nicht nur für Sicherheitsereignisse, die anhand von Protokollnachrichten bezüglich Benutzeraktivitäten oder Netzwerkereignissen generiert wurden, sondern auch für solche, die aufgrund von Änderungen in der Infrastruktur ausgelöst wurden. Die Fähigkeit, Änderungen zu erkennen, zu bestimmen, ob eine Änderung angemessen war, und diese Informationen dann an den richtigen Korrekturworkflow weiterzuleiten, ist für die Aufrechterhaltung und Validierung einer sicheren Architektur unerlässlich. Dies gilt im Kontext unerwünschter Änderungen, die nicht besonders auffällig sind, sodass ihre Ausführung derzeit nicht mit einer Kombination aus AWS Identity and Access Management (IAM) und AWS Organizations-Konfiguration verhindert werden kann.

Amazon GuardDuty und AWS Security Hub bieten Aggregations-, Deduplizierungs- und Analysemechanismen für Protokolldatensätze, die Ihnen auch über andere AWS-Services zur Verfügung gestellt werden. GuardDuty speist Informationen aus Quellen wie AWS CloudTrail-Management- und -Datenereignissen, VPC-DNS-Protokollen und VPC Flow Logs ein und aggregiert und analysiert diese Informationen. Security Hub kann Ausgaben von GuardDuty, AWS Config, Amazon Inspector, Amazon Macie, AWS Firewall Manager und zahlreichen Sicherheitsprodukten von Drittanbietern im AWS Marketplace einspeisen, aggregieren und analysieren. Das gilt auch für Ihren eigenen Code, wenn er entsprechend erstellt wurde. Sowohl GuardDuty als auch Security Hub verfügen über ein Administrator-Member-Modell, das Ergebnisse und Einblicke über mehrere Konten hinweg aggregieren kann. Security Hub wird häufig von Kunden verwendet, die über ein On-Premise-SIEM als AWS-seitigen Protokoll- und Alarmpräprozessor und Aggregator verfügen. Über diesen können sie Amazon EventBridge über einen AWS Lambda-basierten Prozessor und Weiterleiter einspeisen.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Bewerten der Funktionen zur Protokollverarbeitung: Bewerten Sie die für die Verarbeitung von Protokollen verfügbaren Optionen.
  - Amazon OpenSearch Service zum Protokollieren und Überwachen von (praktisch) allem verwenden
  - Suchen eines Partners mit Spezialisierung auf Protokollierungs- und Überwachungslösungen
- Testen Sie zum Analysieren von CloudTrail-Protokollen zunächst Amazon Athena.
  - Konfigurieren von Athena zum Analysieren von CloudTrail-Protokollen
- Implementieren der zentralisierten Protokollierung in AWS: Sehen Sie sich die folgende AWS-Beispiellösung zum Zentralisieren der Protokollierung für mehrere Quellen an.
  - Centralize logging solution
- Implementieren der zentralisierten Protokollierung mit einem Partner: APN-Partner verfügen über Lösungen, die Ihnen beim zentralen Analysieren von Protokollen helfen.
  - Protokollierung und Überwachung

Ressourcen

Zugehörige Dokumente:

- Zentralisierte Protokollierung in AWS
- AWS Security Hub
- Amazon CloudWatch
- Amazon EventBridge
- Erste Schritte mit Amazon CloudWatch Logs
- Partnerlösungen im Bereich Sicherheit: Protokollierung und Überwachung

Zugehörige Videos:

- Best Practices for Centrally Monitoring Resource Configuration and Compliance (Best Practices für die zentrale Überwachung der Ressourcenkonfiguration und Compliance)
- Remediating Amazon GuardDuty and AWS Security Hub Findings (Korrektur von Amazon GuardDuty- und AWS Security Hub-Feststellungen)

- Threat management in the cloud: Amazon GuardDuty and AWS Security Hub (Bedrohungsmanagement in der Cloud: Amazon GuardDuty und AWS Security Hub)

SEC04-BP03 Automatisierte Reaktion auf Ereignisse

Die Nutzung der Automatisierung zum Ermitteln und Beheben von Ereignissen reduziert den menschlichen Aufwand und menschliche Fehler und ermöglicht Ihnen die Skalierung der Prüffunktionen. Regelmäßige Prüfungen helfen Ihnen dabei, Automatisierungstools zu optimieren und immer wieder auszuführen.

In AWS können interessante Ereignisse und Informationen zu potenziell unerwarteten Änderungen an einem automatisierten Workflow mithilfe von Amazon EventBridge untersucht werden. Dieser Service bietet eine skalierbare Rules Engine, die sowohl native AWS-Ereignisformate (z. B. AWS CloudTrail-Ereignisse) als auch von Ihnen generierbare benutzerdefinierte Ereignisse behandelt. Mit Amazon GuardDuty können Sie Ereignisse auch an ein Workflow-System für jene weiterleiten, die Vorfallreaktionssysteme (AWS Step Functions) erstellen, oder an ein zentrales Sicherheitskonto oder an einen Bucket zur weiteren Analyse.

Um Änderungen zu erkennen und diese Informationen an den richtigen Workflow weiterzuleiten, können Sie AWS-Config-Regeln und Conformance Packsverwenden. AWS Config erkennt Änderungen an ordnungsgemäß ausgeführten Services (wenn auch mit einer höheren Latenz als dies bei EventBridge der Fall ist) und generiert Ereignisse, die mithilfe von AWS-Config-Regeln-Regeln analysiert werden können. Dies ermöglicht es, einen Rollback durchzuführen, Compliance-Richtlinien zu erzwingen und Informationen an Systeme wie Änderungsverwaltungsplattformen und operative Ticketsysteme weiterzuleiten. Sie können nicht nur eigene Lambda-Funktionen schreiben, um auf AWS Config-Ereignisse zu reagieren, sondern auch das AWS-Config-Regeln Development Kitbenutzen und auf eine Bibliothek mit Open Source- AWS-Config-Regeln zugreifen. Conformance Packs sind eine Sammlung von AWS-Config-Regeln- und Korrekturaktionen, die Sie als einzelne Einheit in Form einer YAML-Vorlage bereitstellen. A beispielhafte Conformance-Pack-Vorlage ist für die Well-Architected-Säule „Sicherheit“ verfügbar.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Implementieren automatisierter Warnungen mit GuardDuty: GuardDuty ist ein Service zur Bedrohungserkennung, der Ihre AWS-Konten und AWS-Workloads fortlaufend auf böswillige oder unbefugte Verhaltensweisen überwacht und so schützt. Aktivieren Sie GuardDuty und konfigurieren Sie automatisierte Warnungen.

- Automatisieren von Untersuchungsprozessen: Entwickeln Sie automatische Prozesse, die ein Ereignis untersuchen und Berichte an einen Administrator senden, um Zeit zu sparen.
  - Übung: Amazon GuardDuty in der Praxis

Ressourcen

Zugehörige Dokumente:

- Zentralisierte Protokollierung in AWS
- AWS Security Hub
- Amazon CloudWatch
- Amazon EventBridge
- Erste Schritte mit Amazon CloudWatch Logs
- Partnerlösungen im Bereich Sicherheit: Protokollierung und Überwachung
- Erste Schritte mit Amazon GuardDuty

Zugehörige Videos:

- Best Practices for Centrally Monitoring Resource Configuration and Compliance (Best Practices für die zentrale Überwachung der Ressourcenkonfiguration und Compliance)
- Remediating Amazon GuardDuty and AWS Security Hub Findings (Korrektur von Amazon GuardDuty- und AWS Security Hub-Feststellungen)
- Threat management in the cloud: Amazon GuardDuty and AWS Security Hub (Bedrohungsmanagement in der Cloud: Amazon GuardDuty und AWS Security Hub)

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung von aufdeckenden Kontrollen

SEC04-BP04 Implementieren von umsetzbaren Sicherheitsereignissen:

Erstellen Sie Warnungen, die an Ihr Team gesendet werden und von diesem bearbeitet werden können. Stellen Sie sicher, dass Warnungen relevante Informationen enthalten, damit das Team Maßnahmen ergreifen kann. Für jeden Aufklärungsmechanismus, den Sie besitzen, sollten Sie auch einen Prozess zur Untersuchung in Form eines Runbooks oder eines Playbookshaben. Wenn Sie

beispielsweise Amazon GuardDutyaktivieren, werden verschiedene Ergebnisse. Sie sollten einen Runbook-Eintrag für jeden Ergebnistyp haben. Wenn beispielsweise ein Trojaner erkannt wird, enthält Ihr Runbook einfache Anweisungen, die jemanden anweisen, den Vorfall zu untersuchen und zu beheben.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Ermitteln verfügbarer Metriken für AWS-Services: Ermitteln Sie die Metriken, die über Amazon CloudWatch für die Services verfügbar sind, die Sie verwenden.
  - AWS-Servicedokumentation
  - Verwenden von Amazon CloudWatch-Metriken
- Konfigurieren Sie Amazon CloudWatch-Alarme.
  - Verwenden von Amazon CloudWatch-Alarmen

Ressourcen

Zugehörige Dokumente:

- Amazon CloudWatch
- Amazon EventBridge
- Partnerlösungen im Bereich Sicherheit: Protokollierung und Überwachung

Zugehörige Videos:

- Best Practices for Centrally Monitoring Resource Configuration and Compliance (Best Practices für die zentrale Überwachung der Ressourcenkonfiguration und Compliance)
- Remediating Amazon GuardDuty and AWS Security Hub Findings (Korrektur von Amazon GuardDuty- und AWS Security Hub-Feststellungen)
- Threat management in the cloud: Amazon GuardDuty and AWS Security Hub (Bedrohungsmanagement in der Cloud: Amazon GuardDuty und AWS Security Hub)

Schutz der Infrastruktur

Fragen

- SICH 5 Wie schützen Sie Ihre Netzwerkressourcen?
- SICH 6 Wie schützen Sie Ihre Datenverarbeitungsressourcen?

SICH 5 Wie schützen Sie Ihre Netzwerkressourcen?

Alle Workloads, die über eine Art Netzwerkverbindung verfügen, unabhängig davon, ob es sich um das Internet oder ein privates Netzwerk handelt, erfordern mehrere Abwehrebenen, um Schutz vor externen und internen Netzwerkbedrohungen sicherzustellen.

Bewährte Methoden

- SEC05-BP01 Erstellen von Netzwerk-Layern
- SEC05-BP02 Kontrollieren des Datenverkehrs auf allen Ebenen
- SEC05-BP03 Automatisieren des Netzwerkschutzes
- SEC05-BP04 Implementieren von Prüfung und Schutz

SEC05-BP01 Erstellen von Netzwerk-Layern

Gruppieren Sie Komponenten mit den gleichen Erreichbarkeitsanforderungen in Ebenen. Beispielsweise sollte ein Datenbank-Cluster in einer Virtual Private Cloud (VPC) ohne erforderlichen Internetzugriff in Subnetzen ohne Route zum oder aus dem Internet platziert werden. In einer serverlosen Arbeitslast, die ohne VPC ausgeführt wird, können ähnliche Ebenen und die Segmentierung mit Microservices dasselbe Ziel erreichen.

Komponenten wie Amazon Elastic Compute Cloud (Amazon EC2)-Instances, Amazon Relational Database Service (Amazon RDS)-Datenbank-Cluster und AWS Lambda-Funktionen, die gemeinsame Verfügbarkeitsanforderungen haben, können in Ebenen unterteilt werden, welche von Subnetzen gebildet werden. Beispielsweise sollte ein Amazon RDS-Datenbank-Cluster in einer VPC ohne erforderlichen Internetzugriff in Subnetzen ohne Route zum oder aus dem Internet platziert werden. Dieser Ansatz auf mehreren Kontrollebenen mildert die Auswirkungen einer fehlerhaften Konfiguration einer einzelnen Ebene, wodurch möglicherweise unbeabsichtigter Zugriff möglich wäre. Für Lambda können Sie Ihre Funktionen in Ihrer VPC ausführen, um die VPC-basierten Kontrollen zu nutzen.

Für Netzwerkkonnektivität, die Tausende von VPCs, AWS-Konten und On-Premise-Netzwerke umfassen kann, empfiehlt sich die Verwendung von AWS Transit Gateway. Es fungiert als Hub, welcher den Datenfluss für alle als Speicher agierenden Netzwerke steuert. Der Datenverkehr

zwischen einer Amazon Virtual Private Cloud und AWS Transit Gateway verbleibt im privaten AWS-Netzwerk, wodurch externe Bedrohungsvektoren wie DDoS-Angriffe (Distributed Denial of Service) und häufige Exploits wie SQL-Injection, Cross-Site-Scripting, Cross-Site-Anforderungsfälschung oder Missbrauch eines fehlerhaften Authentifizierungscodes reduziert werden. Das regionsübergreifende Peering von AWS Transit Gateway verschlüsselt auch den regionsübergreifenden Datenverkehr ohne Single Point of Failure oder Bandbreitenengpässe.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Subnetze in VPC erstellen: Erstellen Sie Subnetze für jede Ebene (in Gruppen mit mehreren Availability Zones) und ordnen Sie Routing-Tabellen zu, um das Routing zu steuern.
  - VPCs und Subnetze
  - Routing-Tabellen

Ressourcen

Zugehörige Dokumente:

- AWS Firewall Manager
- Amazon Inspector
- Amazon VPC-Sicherheit
- Erste Schritte mit AWS WAF

Zugehörige Videos:

- AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs
- Application Acceleration and Protection with Amazon CloudFront, AWS WAF, and AWS Shield (Anwendungsbeschleunigung und -schutz mit Amazon CloudFront, AWS WAF und AWS Shield)

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung von VPC

SEC05-BP02 Kontrollieren des Datenverkehrs auf allen Ebenen

Bei der Architektur Ihrer Netzwerktopologie sollten Sie die Konnektivitätsanforderungen der einzelnen Komponenten überprüfen. Beispielsweise bei Komponenten, welche Internetzugang (ein- und ausgehend), Konnektivität zu VPCs, Edge-Services und oder externe Rechenzentren erfordern.

Mit einer VPC können Sie Ihre Netzwerktopologie definieren, die eine AWS-Region mit einem von Ihnen festgelegten privaten IPv4-Adressbereich oder einem von AWS ausgewählten IPv6-Adressbereich umfasst. Sie sollten mehrere Kontrollmechanismen mit einem umfassenden Verteidigungsansatz für den ein- und ausgehenden Datenverkehr anwenden, einschließlich der Verwendung von Sicherheitsgruppen (Stateful Inspection Firewall), Netzwerk-ACLs, Subnetzen und Routing-Tabellen. Innerhalb einer VPC können Sie Subnetze in einer Availability Zone erstellen. Jedes Subnetz ist mit einer Routing-Tabelle mit Routing-Regeln verknüpft, mit denen Sie die Pfade des Datenverkehrs innerhalb des Subnetzes steuern können. Sie können ein routingfähiges Internet-Subnetz über eine Route definieren, die zu einem Internet- oder NAT-Gateway geleitet wird, das dieser oder einer anderen VPC zugehörig ist.

Wenn eine Instance, eine Amazon Relational Database Service (Amazon RDS)-Datenbank oder ein anderer Service innerhalb einer VPC gestartet wird, verfügt sie über eine eigene Sicherheitsgruppe pro Netzwerkschnittstelle. Diese Firewall befindet sich außerhalb der Betriebssystemebene. Sie können damit Regeln für zulässigen ein- und ausgehenden Datenverkehr festlegen. Des Weiteren haben Sie die Möglichkeit, Beziehungen zwischen Sicherheitsgruppen zu definieren. Beispielsweise akzeptieren Instances innerhalb einer Sicherheitsgruppe der Datenbankebene nur Datenverkehr von Instancces innerhalb der Anwendungsebene unter Bezugnahme auf die Sicherheitsgruppen, die auf die beteiligten Instances angewendet werden. Sofern Sie keine Nicht-TCP-Protokolle verwenden, sollte es nicht notwendig sein, eine Amazon Elastic Compute Cloud (Amazon EC2)-Instance ohne Load Balancer oder CloudFront. Dies schützt sie vor unbeabsichtigtem Zugriff aufgrund eines Betriebssystem- oder Anwendungsfehlers. Einem Subnetz kann auch eine Netzwerk-ACL zugeordnet sein, die als zustandslose Firewall fungiert. Sie sollten die Netzwerk-ACL so konfigurieren, dass der zulässige Datenverkehr zwischen den Ebenen beschränkt wird. Beachten Sie, dass Sie Regeln für den ein- und ausgehenden Datenverkehr definieren müssen.

Manche AWS-Services erfordern Komponenten für den Zugriff auf das Internet, um API-Aufrufe dort zu tätigen, wo sich AWS-API-Endpunkte befinden. Andere AWS-Services verwenden VPC-Endpunkte innerhalb Ihrer Amazon VPCs. Viele AWS-Services wie Amazon S3 und Amazon DynamoDB unterstützen VPC-Endpunkte. Diese Technologie wurde in AWS PrivateLink.
Wir empfehlen Ihnen die Verwendung dieses Ansatzes für den Zugriff auf AWS-Services, Drittanbieterservices und Ihre eigenen Services, die sicher in anderen VPCs gehostet sind.

Sämtlicher Netzwerkverkehr in AWS PrivateLink bleibt im globalen AWS-Backbone und durchquert nie das Internet. Die Konnektivität kann nur von Benutzern des Service eingeleitet werden, nicht vom Anbieter des Service. Die Verwendung von AWS PrivateLink für den Zugriff auf externe Services ermöglicht die Erstellung isolierter VPCs ohne Internetzugriff und hilft beim Schutz Ihrer VPCs vor externen Bedrohungsvektoren. Drittanbieterservices können AWS PrivateLink verwenden, um ihren Kunden die Verbindung mit Services über private IP-Adressen von ihren VPCs aus zu ermöglichen. Für VPC-Komponenten, die eine Verbindung mit dem Internet herstellen müssen, können diese nur ausgehend (einseitig) über ein AWS-verwaltetes NAT-Gateway, ein ausgehendes Internet-Gateway oder einen von Ihnen erstellten und verwalteten Web-Proxy erfolgen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Kontrollieren des Netzwerkdatenverkehrs in einer VPC: Implementieren Sie VPC-Best-Practices zum Kontrollieren des Datenverkehrs.
  - Amazon VPC-Sicherheit
  - VPC-Endpunkte
  - Amazon VPC-Sicherheitsgruppe
  - Netzwerk-ACLs
- Kontrollieren des Datenverkehrs am Edge: Implementieren Sie Edge-Services wie Amazon CloudFront, um eine zusätzliche Schutzebene und andere Funktionen bereitzustellen.
  - Amazon CloudFront-Anwendungsfälle
  - AWS Global Accelerator
  - AWS Web Application Firewall (AWS WAF)
  - Amazon Route 53
  - Amazon VPC-Eingangs-Routing
- Kontrollieren des privaten Netzwerkverkehrs: Implementieren Sie Services, die Ihren privaten Datenverkehr für Ihre Workload schützen.
  - Amazon VPC-Peering
  - Amazon VPC-Endpunkt-Services (AWS PrivateLink)
  - Amazon VPC Transit Gateway
  - AWS Direct Connect
  - AWS-Site-to-Site-VPN

- AWS-Client-VPN
- Amazon S3-Zugriffspunkte

Ressourcen

Ähnliche Dokumente:

- AWS Firewall Manager
- Amazon Inspector
- Erste Schritte mit AWS WAF

Ähnliche Videos:

- AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs
- Application Acceleration and Protection with Amazon CloudFront, AWS WAF, and AWS Shield (Anwendungsbeschleunigung und -schutz mit Amazon CloudFront, AWS WAF und AWS Shield)

Ähnliche Beispiele:

- Übung: Automatisierte Bereitstellung von VPC

SEC05-BP03 Automatisieren des Netzwerkschutzes

Automatisieren Sie Schutzmechanismen, um ein selbstverteidigendes Netzwerk bereitzustellen, das auf Threat Intelligence und Erkennung von Anomalien beruht. Zum Beispiel können Tools zur Erkennung und Verhinderung von Eindringversuchen sich an aktuelle Bedrohungen anpassen und deren Auswirkungen reduzieren. Eine Webanwendungs-Firewall ist ein Beispiel dafür, wie Sie den Netzwerkschutz automatisieren können, indem Sie beispielsweise die AWS WAF Security Automations-Lösung (https://github.com/awslabs/aws-waf-security-automationsverwenden, um Netzwerkverkehr zu blockieren, welcher von schadhaften IP-Adressen stammt.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren des Schutzes für webbasierten Datenverkehr: AWS bietet eine Lösung, die AWS CloudFormation verwendet, um automatisch eine Reihe von AWS WAF-Regeln zum

Filtern gängiger webbasierter Angriffe bereitzustellen. Benutzer können aus vorkonfigurierten Schutzfunktionen wählen, die die in einer AWS WAF Web Access Control List (Web ACL) enthaltenen Regeln definieren.

  - Sicherheitsautomatisierung mit AWS WAF
- Erwägen von AWS Partner-Lösungen: AWS-Partner bieten Hunderte branchenführende Produkte, die mit vorhandenen Kontrollen in Ihren On-Premises-Umgebungen gleichwertig oder identisch sind oder sich in diese integrieren lassen. Diese Produkte ergänzen die vorhandenen AWS-Services, sodass Sie eine umfassende Sicherheitsarchitektur bereitstellen und eine nahtlosere Erfahrung in Ihren Cloud- und On-Premises-Umgebungen ermöglichen können.
  - Sicherheit der Infrastruktur

Ressourcen

Zugehörige Dokumente:

- AWS Firewall Manager
- Amazon Inspector
- Amazon VPC-Sicherheit
- Erste Schritte mit AWS WAF

Zugehörige Videos:

- AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs
- Application Acceleration and Protection with Amazon CloudFront, AWS WAF, and AWS Shield (Anwendungsbeschleunigung und -schutz mit Amazon CloudFront, AWS WAF und AWS Shield)

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung von VPC

SEC05-BP04 Implementieren von Prüfung und Schutz

Untersuchen und filtern Sie Ihren Datenverkehr auf jeder Ebene. Mit dem VPC Network Access Analyzer können Sie Ihre VPC-Konfigurationen auf potenziell unbeabsichtigten Zugriff überprüfen. Sie können Ihre Netzwerkzugriffsanforderungen festlegen und potenzielle Netzwerkpfade identifizieren, die diese nicht erfüllen. Für Komponenten, die über HTTP-basierte Protokolle

abgefertigt werden, kann eine Webanwendungs-Firewall zum Schutz vor gängigen Angriffen beitragen. AWS WAF ist eine Firewall für Webanwendungen, mit der Sie HTTP(s)-Anforderungen überwachen und blockieren können, die Ihren konfigurierbaren Regeln entsprechen und an eine Amazon API Gateway-API, Amazon CloudFront oder Application Load Balancer weitergeleitet werden. Für den Einstieg in AWS WAF können Sie Von AWS verwaltete Regeln in Kombination mit Ihren eigenen vorhandenen Partnerintegrationen.

Für die Verwaltung von AWS WAF, AWS Shield Advanced-Schutzmaßnahmen und Amazon VPC-Sicherheitsgruppen in AWS Organizations können Sie AWS Firewall Manager verwenden. Dies ermöglicht Ihnen die zentrale Konfiguration und Verwaltung von Firewall-Regeln für Ihre Konten und Anwendungen, was eine Skalierung einfacher macht. Außerdem können Sie schnell auf Angriffe reagieren, indem Sie AWS Shield Advancedoder Lösungen verwenden, die unerwünschte Anfragen an Ihre Webanwendungen automatisch blockieren. Firewall Manager lässt sich auch mit AWS Network Firewall kombinieren. AWS Network Firewall ist ein verwalteter Service, der eine Regel-Engine nutzt, um Ihnen die detaillierte Kontrolle über zustandsbehafteten und zustandlosen Netzwerkdatenverkehr zu ermöglichen. Er unterstützt Suricata-kompatible Open-Source-IPS-Spezifikationen (Intrusion Prevention System) für Regeln zum Schutz Ihrer Workload.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Konfigurieren von Amazon GuardDuty: GuardDuty ist ein Service zur Bedrohungserkennung, der Ihre AWS-Konten und AWS-Workloads fortlaufend auf schädliche oder unbefugte Verhaltensweisen überwacht und dadurch schützt. Aktivieren Sie GuardDuty und konfigurieren Sie automatisierte Warnungen.
  - Amazon GuardDuty
  - Übung: Automatisierte Bereitstellung von aufdeckenden Kontrollen
- Konfigurieren von Virtual Private Cloud (VPC) Flow Logs: VPC Flow Logs ist eine Funktion, mit deren Hilfe Sie Informationen zum ein- und ausgehenden IP-Datenverkehr an den Netzwerkschnittstellen Ihrer VPC erfassen können. Flussprotokolldaten können in Amazon CloudWatch Logs und Amazon Simple Storage Service (Amazon S3) veröffentlicht werden. Sobald das Flussprotokoll fertig ist, können Sie seine Daten auf den ausgewählten Zielort abrufen und dort einsehen.
- Erwägen von VPC-Datenverkehrabbildung: Die Datenverkehrabbildung ist eine Amazon Amazon VPC-Funktion, mit der Sie Netzwerkdatenverkehr von einer Elastic-Network-Schnittstelle von Amazon Elastic Compute Cloud (Amazon EC2)-Instances kopieren und diesen dann zur

Inhaltsprüfung, Bedrohungsüberwachung und Fehlerbehebung an Out-of-Band-Sicherheits- und -Überwachungs-Appliances senden können.

- VPC-Datenverkehrabbildung

Ressourcen

Ähnliche Dokumente:

- AWS Firewall Manager
- Amazon Inspector
- Amazon VPC-Sicherheit
- Erste Schritte mit AWS WAF

Ähnliche Videos:

- AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs
- Application Acceleration and Protection with Amazon CloudFront, AWS WAF, and AWS Shield (Anwendungsbeschleunigung und -schutz mit Amazon CloudFront, AWS WAF und AWS Shield)

Ähnliche Beispiele:

- Übung: Automatisierte Bereitstellung von VPC

SICH 6 Wie schützen Sie Ihre Datenverarbeitungsressourcen?

Datenverarbeitungsressourcen in Ihrem Workload erfordern mehrere Ebenen der Abwehr zum Schutz vor externen und internen Bedrohungen. Zu den Datenverarbeitungsressourcen zählen EC2-Instances, Container, AWS Lambda-Funktionen, Datenbankservices, IoT-Geräte und mehr.

Bewährte Methoden

- SEC06-BP01 Schwachstellenmanagement
- SEC06-BP02 Verringern der Angriffsfläche
- SEC06-BP03 Implementieren von verwalteten Services
- SEC06-BP04 Automatisieren des Datenverarbeitungsschutzes
- SEC06-BP05 Personen das Ausführen von Aktionen aus der Ferne ermöglichen

- SEC06-BP06 Validieren der Softwareintegrität

SEC06-BP01 Schwachstellenmanagement

Überprüfen Sie häufig Schwachstellen in Ihrem Code, Ihren Abhängigkeiten und in Ihrer Infrastruktur, um Schutz vor neuen Bedrohungen zu bieten.

Bei der Konfiguration Ihrer Datenverarbeitungsinfrastruktur können Sie die Erstellung und Aktualisierung von Ressourcen mit AWS CloudFormation automatisieren. CloudFormation ermöglicht die Erstellung von Vorlagen, die in YAML oder JSON geschrieben sind. Dafür können Sie entweder AWS-Beispiele verwenden oder Ihre eigenen Vorlagen schreiben. So können Sie standardmäßig sichere Infrastrukturvorlagen erstellen, die Sie mit CloudFormation Guardverifizieren können. Das spart Ihnen Zeit und reduziert das Risiko von Konfigurationsfehlern. Sie können für den Aufbau Ihrer Infrastruktur und die Bereitstellung Ihrer Anwendungen auf Continuous Delivery zurückgreifen, z. B. mit AWS CodePipeline, um das Erstellen, Testen und Freigeben zu automatisieren.

Sie sind für das Patch-Management für Ihre AWS-Ressourcen verantwortlich, einschließlich Amazon Elastic Compute Cloud(Amazon EC2)-Instances, Amazon Machine Images (AMIs) und viele andere Datenverarbeitungsressourcen. Für Amazon EC2-Instances automatisiert AWS Systems Manager das Patchen verwalteter Instances mit sicherheitsrelevanten und anderen Arten von Updates.
Sie können Patch Manager verwenden, um Patches für Betriebssysteme und Anwendungen anzuwenden. (Für Windows-Server ist der Anwendungs-Support auf Updates von Microsoft-Anwendungen beschränkt.) Sie können Patch Manager verwenden, um Service Packs auf Windows-Instances zu installieren und kleinere Versions-Upgrades auf Linux-Instances vorzunehmen. Sie können Flotten von Amazon EC2-Instances oder Ihre On-Premises-Server und virtuelle Maschinen (VMs) nach Betriebssystemtyp patchen. Das beinhaltet unterstützte Versionen von Windows Server, Amazon Linux, Amazon Linux 2, CentOS, Debian Server, Oracle Linux, Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES) und Ubuntu Server. Sie können Instances scannen, um nur fehlende Patches angezeigt zu bekommen oder Sie können scannen und automatisch alle fehlenden Patches installieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Konfigurieren von Amazon Inspector: Amazon Inspector testen die Netzwerkzugänglichkeit Ihrer Amazon Elastic Compute Cloud (Amazon EC2)-Instances und den Sicherheitsstatus der Anwendungen, die auf diesen Instances ausgeführt werden. Amazon Inspector bewertet Anwendungen hinsichtlich Exposition, Schwachstellen und Abweichungen von Best Practices.

  - Was ist Amazon Inspector?
- Scannen von Quellcode: Durchsuchen Sie Bibliotheken und Abhängigkeiten nach Schwachstellen.
  - Amazon CodeGuru
  - OWASP: Tools zur Quellcodeanalyse

Ressourcen

Zugehörige Dokumente:

- AWS Systems Manager
- Replacing a Bastion Host with Amazon EC2 Systems Manager (Ersetzen eines Bastion-Host mit Amazon EC2 Systems Manager)
- Übersicht zur Sicherheit von AWS Lambda

Zugehörige Videos:

- Running high-security workloads on Amazon EKS (Ausführen von Workloads mit hoher Sicherheit auf Amazon EKS)
- Securing Serverless and Container Services
- Bewährte Sicherheitsmethoden für den Amazon EC2-Instance-Metadatenservice

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung der Web Application Firewall

SEC06-BP02 Verringern der Angriffsfläche

Reduzieren Sie Ihre Gefährdung mit Blick auf unbefugte Zugriffe, indem Sie Betriebssysteme härten und Komponenten, Bibliotheken und extern nutzbare Services minimieren. Reduzieren Sie zunächst ungenutzte Komponenten für alle Workloads, unabhängig davon, ob es sich um Betriebssystempakete, Anwendungen für Amazon Elastic Compute Cloud (Amazon EC2)-basierte Workloads oder externe Softwaremodule in Ihrem Code handelt. Viele Leitfäden für Härtung und Sicherheit sind für gängige Betriebssysteme und Serversoftware verfügbar. Sie können zum Beispiel mit dem Center for Internet Security (CIS) beginnen und dann iterieren.

In Amazon EC2 können Sie zur Erfüllung der spezifischen Sicherheitsanforderungen Ihrer Organisation Ihre eigenen Amazon Machine Images (AMIs) erstellen, die Sie gepatcht und gehärtet haben. Die Patches und anderen Sicherheitskontrollen, die Sie auf das AMI anwenden, sind zum Zeitpunkt ihrer Erstellung wirksam. Sie sind nicht dynamisch, es sei denn, Sie nehmen nach dem Starten Änderungen vor (z. B. mit AWS Systems Manager).

Sie können den Prozess zur Erstellung sicherer AMIs mit EC2 Image Builder vereinfachen. EC2 Image Builder senkt den Aufwand für die Erstellung und Pflege goldener Images deutlich, ohne dass die Automatisierung implementiert und gewartet werden muss. Wenn Software-Updates verfügbar sind, erzeugt Image Builder automatisch ein neues Image, ohne dass Benutzer Image-Builds manuell anstoßen müssen. EC2 Image Builder ermöglicht Ihnen das einfache Validieren der Funktionalität und Sicherheit Ihrer Images mit von AWS bereitgestellten und Ihren eigenen Tests, bevor Sie die Images in der Produktion nutzen. Sie können auch von AWS bereitgestellte Sicherheitseinstellungen anwenden, um Ihre Images weiter abzusichern und interne Sicherheitskriterien zu erfüllen. Unter Verwendung von AWS bereitgestellter Vorlagen können Sie beispielsweise Security Technical Implementation Guide (STIG)-konforme Images erstellen.

Mit Drittanbieter-Tools zur statischen Code-Analyse können Sie häufige Sicherheitsprobleme wie nicht geprüfte Funktionseingangsgrenzen sowie zutreffende CVEs identifizieren. Sie können Amazon CodeGuru für unterstützte Sprachen verwenden. Sie können auch Drittanbieter-Tools zur Überprüfung von Abhängigkeiten verwenden, um zu ermitteln, ob Bibliotheken, welche von Ihnen genutzt werden, auf dem neuesten Stand sind, frei von CVEs sind und die passende Lizenzierung enthalten, die den Anforderungen Ihrer Softwarepolitik entsprechen.

Amazon Inspector bietet Ihnen die Möglichkeit, Konfigurationsbewertungen Ihrer Instances bezüglich bekannter CVEs durchzuführen. Darüber hinaus können Sie eine Bewertung im Hinblick auf Sicherheits-Benchmarks vornehmen und Benachrichtigungen bei Fehlern automatisieren. Amazon Inspector kann auf Produktions-Instances und in Build-Pipelines ausgeführt werden, um Entwickler und Techniker bezüglich vorhandener Fehler zu benachrichtigen. Sie können programmgesteuert auf ermittelte Fehler zugreifen oder Ihr Team auf Backlogs und Bug-Verfolgungssysteme verweisen. EC2 Image Builder kann verwendet werden, um Server-Images (AMIs) mit automatischem Patching, von AWS bereitgestellter Durchsetzung von Sicherheitsrichtlinien und anderen Anpassungen zu verwalten. Implementieren Sie bei der Verwendung von Containern ECR Image Scanning in Ihrer Build-Pipeline und scannen Sie regelmäßig Ihr Image-Repository, um nach CVEs in Ihren Containern zu suchen.

Amazon Inspector und andere Tools sind zwar effektiv bei der Identifizierung von Konfigurationen und vorhandenen CVEs, doch andere Methoden sind erforderlich, um Ihren Workload auf

Anwendungsebene zu testen. Fuzzing ist eine bekannte Methode zur Suche von Fehlern mithilfe von Automatisierung, um falsch formatierte Daten in Eingabefeldern und anderen Bereichen Ihrer Anwendung zu finden.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Härten des Betriebssystems: Konfigurieren Sie Betriebssysteme so, dass sie den Best Practices entsprechen.
  - Sichern von Amazon Linux
  - Sichern von Microsoft Windows Server
- Härten von containerisierten Ressourcen: Konfigurieren Sie containerisierte Ressourcen so, dass sie den Best Practices für Sicherheit entsprechen.
- Implementieren Sie Best Practices für AWS Lambda.
  - Best Practices für AWS Lambda

Ressourcen

Zugehörige Dokumente:

- AWS Systems Manager
- Replacing a Bastion Host with Amazon EC2 Systems Manager (Ersetzen eines Bastion-Host mit Amazon EC2 Systems Manager)
- Übersicht zur Sicherheit von AWS Lambda

Zugehörige Videos:

- Running high-security workloads on Amazon EKS (Ausführen von Workloads mit hoher Sicherheit auf Amazon EKS)
- Securing Serverless and Container Services
- Bewährte Sicherheitsmethoden für den Amazon EC2-Instance-Metadatenservice

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung der Web Application Firewall

SEC06-BP03 Implementieren von verwalteten Services

Implementieren Sie Services zur Verwaltung von Ressourcen wie Amazon Relational Database Service (Amazon RDS), AWS Lambda und Amazon Elastic Container Service (Amazon ECS), um Ihre Aufgaben zur Wahrung der Sicherheit im Rahmen des Modells der gemeinsamen Verantwortung zu reduzieren. Amazon RDS unterstützt Sie beispielsweise beim Einrichten, Betreiben und Skalieren einer relationalen Datenbank und automatisiert Verwaltungsaufgaben wie Hardwarebereitstellung, Datenbankeinrichtung, Patching und Sicherungen. Das bedeutet, dass Sie mehr Zeit haben, sich auf alternative Möglichkeiten zum Absichern Ihrer Anwendung zu konzentrieren, die im AWS Well-Architected Framework beschrieben werden. Mit Lambda können Sie Code ausführen, ohne Server bereitstellen oder verwalten zu müssen. So müssen Sie sich nur auf die Konnektivität, den Aufruf und die Sicherheit auf Codeebene konzentrieren – nicht auf Infrastruktur oder Betriebssystem.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Ermitteln verfügbarer Services: Ermitteln, testen und implementieren Sie Services zur Verwaltung von Ressourcen wie Amazon RDS, AWS Lambda und Amazon ECS.

Ressourcen

Zugehörige Dokumente:

- AWS-Website:
- AWS Systems Manager
- Replacing a Bastion Host with Amazon EC2 Systems Manager (Ersetzen eines Bastion-Host mit Amazon EC2 Systems Manager)
- Übersicht zur Sicherheit von AWS Lambda

Zugehörige Videos:

- Running high-security workloads on Amazon EKS (Ausführen von Workloads mit hoher Sicherheit auf Amazon EKS)
- Securing Serverless and Container Services
- Bewährte Sicherheitsmethoden für den Amazon EC2-Instance-Metadatenservice

Zugehörige Beispiele:

- Übung: AWS Certificate Manager – Anfordern eines öffentlichen Zertifikats

SEC06-BP04 Automatisieren des Datenverarbeitungsschutzes

Automatisieren Sie Ihre Schutz-Rechenmechanismen, einschließlich Schwachstellenmanagement, Reduzierung der Angriffsfläche und Verwaltung von Ressourcen. Die Automatisierung hilft Ihnen, Zeit in die Sicherung anderer Aspekte Ihres Workloads zu investieren und das Risiko menschlicher Fehler zu reduzieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren der Konfigurationsverwaltung: Erzwingen und validieren Sie sichere Konfigurationen automatisch mithilfe eines Service oder Tools zur Konfigurationsverwaltung.
  - AWS Systems Manager
  - AWS CloudFormation
  - Übung: Automatisierte Bereitstellung von VPC
  - Übung: Automatisierte Bereitstellung der EC2-Webanwendung

- Automatisieren des Patchings von Amazon Elastic Compute Cloud (Amazon EC2)-Instances: AWS Systems Manager Patch Manager automatisiert das Patching verwalteter Instances mit sicherheitsrelevanten und anderen Arten von Updates. Sie können Patch Manager verwenden, um Patches für Betriebssysteme und Anwendungen anzuwenden.
  - AWS Systems Manager Patch Manager
  - Centralized multi-account and multi-region patching with AWS Systems Manager Automation (Zentralisiertes Patching über mehrere Konten und Regionen mit AWS Systems Manager-Automatisierung)

- Implementieren von Maßnahmen zur Erkennung und Verhinderung von Eindringversuchen: Implementieren Sie ein Tool zur Erkennung und Verhinderung von Eindringversuchen, um böswillige Aktivitäten auf Instances zu überwachen und zu stoppen.
- Erwägen von AWS Partner-Lösungen: AWS-Partner bieten Hunderte branchenführende Produkte, die mit vorhandenen Kontrollen in Ihren On-Premises-Umgebungen gleichwertig oder identisch

sind oder sich in diese integrieren lassen. Diese Produkte ergänzen die vorhandenen AWS-Services, sodass Sie eine umfassende Sicherheitsarchitektur bereitstellen und eine nahtlosere Erfahrung in Ihren Cloud- und On-Premises-Umgebungen ermöglichen können.

- Sicherheit der Infrastruktur

Ressourcen

Zugehörige Dokumente:

- AWS CloudFormation
- AWS Systems Manager
- AWS Systems Manager Patch Manager
- Centralized multi-account and multi-region patching with AWS Systems Manager Automation (Zentralisiertes Patching über mehrere Konten und Regionen mit AWS Systems Manager-Automatisierung)
- Sicherheit der Infrastruktur
- Replacing a Bastion Host with Amazon EC2 Systems Manager (Ersetzen eines Bastion-Host mit Amazon EC2 Systems Manager)
- Übersicht zur Sicherheit von AWS Lambda

Zugehörige Videos:

- Running high-security workloads on Amazon EKS (Ausführen von Workloads mit hoher Sicherheit auf Amazon EKS)
- Securing Serverless and Container Services
- Bewährte Sicherheitsmethoden für den Amazon EC2-Instance-Metadatenservice

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung der Web Application Firewall
- Übung: Automatisierte Bereitstellung der EC2-Webanwendung

SEC06-BP05 Personen das Ausführen von Aktionen aus der Ferne ermöglichen

Durch das Entfernen der Möglichkeit für interaktiven Zugriff wird das Risiko menschlicher Fehler und das Potenzial einer manuellen Konfiguration oder Verwaltung reduziert. Verwenden Sie beispielsweise einen Änderungsmanagement-Workflow, um Amazon Elastic Compute Cloud (Amazon EC2)-Instances unter Verwendung von Infrastruktur als Code bereitzustellen und Amazon EC2-Instances dann mit Tools wie AWS Systems Manager zu verwalten, statt direkten Zugriff oder Zugriff über einen Bastion-Host zuzulassen. AWS Systems Manager automatisiert eine Vielzahl von Wartungs- und Bereitstellungsaufgaben mithilfe von Funktionen wie Automatisierung -Workflows, Dokumenten (Playbooks) und dem Run Command. AWS CloudFormation-Stacks werden anhand von Pipelines erstellt und können Ihre Infrastrukturbereitstellungs- und Verwaltungsaufgaben ohne direkte Verwendung der AWS Management Console oder APIs automatisieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Ersetzen des Konsolenzugriffs: Ersetzen Sie den Konsolenzugriff (SSH oder RDP) auf Instances mit AWS Systems Manager Run Command, um Verwaltungsaufgaben zu automatisieren.

- AWS Systems Manager Run Command

Ressourcen

Zugehörige Dokumente:

- AWS Systems Manager
- AWS Systems Manager Run Command
- Replacing a Bastion Host with Amazon EC2 Systems Manager (Ersetzen eines Bastion-Host mit Amazon EC2 Systems Manager)
- Übersicht zur Sicherheit von AWS Lambda

Zugehörige Videos:

- Running high-security workloads on Amazon EKS (Ausführen von Workloads mit hoher Sicherheit auf Amazon EKS)
- Securing Serverless and Container Services

- Bewährte Sicherheitsmethoden für den Amazon EC2-Instance-Metadatenservice

Zugehörige Beispiele:

- Übung: Automatisierte Bereitstellung der Web Application Firewall

SEC06-BP06 Validieren der Softwareintegrität

Implementieren Sie Mechanismen (z. B. Codesignierung), um zu überprüfen, ob die Software, der Code und die Bibliotheken, die in der Workload verwendet werden, aus vertrauenswürdigen Quellen stammen und nicht manipuliert wurden. Sie sollten beispielsweise das Codesignierungszertifikat der Binärdateien und Skripte überprüfen, um den Autor zu bestätigen, und sicherzustellen, dass es seit der Erstellung durch den Autor nicht manipuliert wurde. AWS Signer kann Sie beim Sicherstellen der Vertrauenswürdigkeit und Integrität Ihres Codes unterstützen, indem der Codesignierungslebenszyklus zentral verwaltet wird, einschließlich Signierungszertifizierung und öffentliche und private Schlüssel. Informieren Sie sich über die Verwendung erweiterter Muster und Best Practices für die Codesignierung mit AWS Lambda. Darüber hinaus kann eine Prüfsumme der Software, die Sie herunterladen, im Vergleich zu der Prüfsumme vom Anbieter helfen, sicherzustellen, dass sie nicht manipuliert wurde.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Untersuchen von Mechanismen: Die Codesignierung ist ein Mechanismus, der zur Validierung der Softwareintegrität verwendet werden kann.
  - NIST: Sicherheitsüberlegungen für die Codesignierung

Ressourcen

Zugehörige Dokumente:

- AWS Signer
- New – Code Signing, a Trust and Integrity Control for (Neu: Codesignierung, eine Vertrauens- und Integritätskontrolle für AWS Lambda)

Datenschutz

Fragen

- SICH 7 Wie klassifizieren Sie Ihre Daten?
- SICH 8 Wie schützen Sie Ihre Daten im Ruhezustand?
- SICH 9 Wie schützen Sie Ihre Daten bei der Übertragung?

SICH 7 Wie klassifizieren Sie Ihre Daten?

Die Datenklassifizierung bietet eine Möglichkeit, Daten basierend auf Wichtigkeit und Sensibilität zu kategorisieren, um Ihnen dabei zu helfen, angemessene Schutz- und Aufbewahrungskontrollen zu bestimmen.

Bewährte Methoden

- SEC07-BP01 Identifizieren der Daten innerhalb Ihrer Workload
- SEC07-BP02 Definieren von Datenschutzkontrollen:
- SEC07-BP03 Automatisieren der Identifizierung und Klassifizierung
- SEC07-BP04 Definieren des Datenlebenszyklusmanagements:

SEC07-BP01 Identifizieren der Daten innerhalb Ihrer Workload

Sie müssen den Typ und die Klassifizierung der von Ihrem Workload verarbeiteten Daten, die zugehörigen Geschäftsprozesse, den Dateneigentümer, die geltenden gesetzlichen und Compliance-Anforderungen, wo sie gespeichert werden, sowie die resultierenden Kontrollen, die durchgesetzt werden müssen, verstehen. Dies kann Klassifizierungen umfassen, um anzugeben, ob die Daten öffentlich verfügbar sein sollen, ob die Daten nur zur internen Verwendung dienen, wie z. B. personenbezogene Daten des Kunden (PII), oder ob die Daten für einen eingeschränkten Zugriff vorgesehen sind, wie z. B. geistiges Eigentum, gesetzlich privilegierte oder als sensibel gekennzeichnete Daten. Indem Sie entsprechend den Sicherheitsanforderungen jeder Workload ein passendes Datenklassifizierungssystem verwalten, können Sie die für die Daten geeigneten Kontrollen und Zugriffsebenen/Schutzmaßnahmen zuweisen. Öffentliche Inhalte sind beispielsweise für jedermann zugänglich. Wichtige Inhalte hingegen werden verschlüsselt und sicher gespeichert. Hierfür ist der autorisierte Zugriff auf einen Schlüssel für die Entschlüsselung erforderlich.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Erwägen der Datenermittlung mit Amazon Macie: Macie erkennt vertrauliche Daten wie persönlich identifizierbare Informationen (PII) oder geistiges Eigentum.
  - Amazon Macie

Ressourcen

Zugehörige Dokumente:

- Amazon Macie
- Data Classification Whitepaper
- Erste Schritte mit Amazon Macie

Zugehörige Videos:

- Einführung des neuen Amazon Macie

SEC07-BP02 Definieren von Datenschutzkontrollen:

Schützen Sie Daten entsprechend ihrer Klassifizierungsstufe. Schützen Sie beispielsweise Daten, die als öffentlich klassifiziert werden, indem Sie relevante Empfehlungen anwenden und gleichzeitig sensible Daten durch zusätzliche Kontrollen schützen.

Durch die Verwendung von Ressourcen-Tags, separater AWS-Konten je nach Sensibilität (und möglicherweise auch nach Vorbehalt, Enklave oder Interessensgemeinschaft), IAM-Richtlinien, AWS Organizations-SCPs, AWS Key Management Service (AWS KMS) und AWS CloudHSM können Sie Ihre Richtlinien für die Datenklassifizierung und den Datenschutz mit Verschlüsselung definieren und implementieren. Wenn Sie beispielsweise S3-Buckets mit hoch kritischen Daten oder Amazon Elastic Compute Cloud (Amazon EC2)-Instances haben, die vertrauliche Daten verarbeiten, können Sie diese mit dem Tag `Project=ABC` kennzeichnen. Nur Ihr direktes Team weiß, was der Projektcode bedeutet, und es bietet eine Möglichkeit, die attributbasierte Zugriffskontrolle zu verwenden. Sie können für die AWS KMS-Kodierungsschlüssel mithilfe von Schlüsselrichtlinien Zugriffsebenen definieren. Auf diese Weise stellen Sie sicher, dass nur die entsprechenden Services sicher auf die sensiblen Inhalte zugreifen können. Wenn Sie Autorisierungsentscheidungen basierend auf Tags treffen, sollten Sie sicherstellen, dass die Berechtigungen für die Tags mithilfe von Tag-Richtlinien in AWS Organizations entsprechend definiert sind.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Definieren eines Datenidentifikations- und -klassifizierungsschemas: Eine Identifikation und Klassifizierung Ihrer Daten wird durchgeführt, um potenzielle Auswirkungen und den Typ der gespeicherten Daten zu bewerten und festzulegen, welche Personen Zugriff auf die Daten haben sollen.
  - AWS-Dokumentation
- Ermitteln verfügbarer AWS-Kontrollen: Ermitteln Sie die Sicherheitskontrollen für die AWS-Services, die Sie verwenden oder verwenden möchten. Die Dokumentation vieler Services umfasst einen Sicherheitsabschnitt.
  - AWS-Dokumentation
- Identifizieren von AWS-Compliance-Ressourcen: Ermitteln Sie die Ressourcen, die AWS zur Verfügung stellt, um Sie zu unterstützen.
  - https://aws.amazon.com/compliance/

Ressourcen

Zugehörige Dokumente:

- AWS-Dokumentation
- Data Classification Whitepaper
- Erste Schritte mit Amazon Macie
- Fehlender Text

Zugehörige Videos:

- Einführung des neuen Amazon Macie

SEC07-BP03 Automatisieren der Identifizierung und Klassifizierung

Durch die Automatisierung der Identifizierung und Klassifizierung von Daten können Sie die richtigen Kontrollen implementieren. Die Verwendung von Automatisierung für diesen Zweck anstelle des direkten Zugriffs durch eine Person reduziert das Risiko menschlichen Versagens und unbeabsichtigter Offenlegung. Sie sollten die Nutzung eines Tools wie Amazon Maciein

Betracht ziehen. Das Tool verwendet Machine Learning, um sensible Daten in AWS automatisch zu erkennen, zu klassifizieren und zu schützen. Amazon Macie erkennt vertrauliche Daten wie persönlich identifizierbare Informationen (PII) oder geistiges Eigentum und stellt Ihnen Dashboards und Warnungen zur Verfügung, die sichtbar machen, wie auf diese Daten zugegriffen wird bzw. wie diese bewegt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Verwenden von Amazon Simple Storage Service (Amazon S3) Inventory: Amazon S3 Inventory ist eines der Tools, mit denen Sie den Replikations- und Verschlüsselungsstatus Ihrer Objekte prüfen und melden können.
  - Amazon S3 Inventory
- Verwenden von Amazon Macie: Amazon Macie nutzt Machine Learning, um in Amazon S3 gespeicherte Daten automatisch zu erkennen und zu klassifizieren.
  - Amazon Macie

Ressourcen

Ähnliche Dokumente:

- Amazon Macie
- Amazon S3 Inventory
- Data Classification Whitepaper
- Erste Schritte mit Amazon Macie

Ähnliche Videos:

- Einführung des neuen Amazon Macie

SEC07-BP04 Definieren des Datenlebenszyklusmanagements:

Ihre definierte Lebenszyklusstrategie sollte auf Vertraulichkeitsstufen sowie auf gesetzlichen und organisatorischen Anforderungen basieren. Aspekte, einschließlich des Zeitraums für die Aufbewahrung von Daten, Datenvernichtungsprozesse, Datenzugriffsverwaltung, Datentransformation und Datenfreigabe sollten berücksichtigt werden. Wenn Sie eine

Datenklassifizierungsmethode erwägen, achten Sie auf ein ausgewogenes Verhältnis zwischen Nutzbarkeit und Zugriff. Berücksichtigen Sie auch die unterschiedlichen Zugriffsebenen und Nuancen bei der Implementierung eines sicheren und dennoch anwendbaren Ansatzes für jede Ebene. Verwenden Sie immer einen umfassenden Ansatz zur Verteidigung und reduzieren Sie den menschlichen Zugriff auf Daten und Mechanismen zum Umwandeln, Löschen oder Kopieren von Daten. Legen Sie beispielsweise fest, dass Benutzer sich bei einer Anwendung stark authentifizieren müssen, und geben Sie der Anwendung anstelle der Benutzer die erforderliche Zugriffsberechtigung, um Aktionen aus der Ferne auszuführen. Stellen Sie außerdem sicher, dass Benutzer einen vertrauenswürdigen Netzwerkpfad verwenden und Zugriff auf die Verschlüsselungsschlüssel benötigen. Nutzen Sie Tools wie Dashboards oder die automatisierte Berichterstellung, um Benutzern Informationen zu diesen Daten bereitzustellen, statt ihnen direkten Zugriff auf die Daten zu gewähren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Identifizieren von Datentypen: Identifizieren Sie die Datentypen, die Sie in Ihrer Workload speichern oder verarbeiten. Diese Daten können Text, Bilder, Binärdatenbanken usw. sein.

Ressourcen

Zugehörige Dokumente:

- Data Classification Whitepaper
- Erste Schritte mit Amazon Macie

Zugehörige Videos:

- Einführung des neuen Amazon Macie

SICH 8 Wie schützen Sie Ihre Daten im Ruhezustand?

Schützen Sie Ihre Daten im Ruhezustand, indem Sie mehrere Kontrollen implementieren, um das Risiko eines unbefugten Zugriffs oder eines Missbrauchs zu reduzieren.

Bewährte Methoden

- SEC08-BP01: Implementieren einer sicheren Schlüsselverwaltung
- SEC08-BP02 Erzwingen der Verschlüsselung im Ruhezustand

- SEC08-BP03 Automatisieren des Schutzes von Daten im Ruhezustand:
- SEC08-BP04 Durchsetzen der Zugriffskontrolle
- SEC08-BP05 Verwenden von Mechanismen, die den direkten Zugriff auf Daten verhindern

SEC08-BP01: Implementieren einer sicheren Schlüsselverwaltung

Durch die Definition eines Verschlüsselungsansatzes, der die Speicherung, regelmäßige Änderung und Zugriffskontrolle von Schlüsseln umfasst, können Sie Ihren Inhalt vor nicht autorisierten Benutzern und vor unnötiger Offenlegung gegenüber autorisierten Benutzern schützen. AWS Key Management Service (AWS KMS) erleichtert die Verwaltung der Verschlüsselungsschlüssel und lässt sich in zahlreiche AWS-Services integrieren. Der Service bietet eine langlebige, sichere und redundante Speicherung Ihrer AWS KMS-Schlüssel. Sie können sowohl Schlüsselaliase als auch schlüsselspezifische Richtlinien festlegen. Die Richtlinien erleichtern das Festlegen von Schlüsseladministratoren und Schlüsselbenutzern. Mit dem Cloud-basierten Hardwaresicherheitsmodul (HSM) AWS CloudHSM können Sie zudem auf einfache Weise eigene Verschlüsselungsschlüssel erstellen und in der AWS Cloud verwenden. Es hilft Ihnen, unternehmensspezifische, vertragliche und gesetzliche Compliance-Anforderungen hinsichtlich der Datensicherheit zu erfüllen. Dazu werden nach FIPS 140-2 Level 3 validierte HSMs verwendet.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren von AWS KMS: Der AWS KMS erleichtert Ihnen das Erstellen und Verwalten von Schlüsseln sowie die Kontrolle der Verschlüsselung in einer Vielzahl von AWS-Services und in Ihren Anwendungen.AWS KMS ist ein sicherer und widerstandsfähiger Service, der FIPS 140-2-validierte Hardwaresicherheitsmodule zum Schutz Ihrer Schlüssel nutzt.
  - Erste Schritte: AWS Key Management Service (AWS KMS)
- Erwägen des AWS-Verschlüsselungs-SDK: Verwenden Sie das AWS-Verschlüsselungs-SDK mit AWS KMS-Integration, wenn Ihre Anwendung Daten clientseitig verschlüsseln muss.
  - AWS-Verschlüsselungs-SDK

Ressourcen

Ähnliche Dokumente:

- AWS Key Management Service

- Kryptografische AWS-Services und -Tools
- Erste Schritte: AWS Key Management Service (AWS KMS)
- Protecting Amazon S3 Data Using Encryption (Amazon S3-Daten durch Verschlüsselung schützen)

Ähnliche Videos:

- How Encryption Works in AWS (So funktioniert die Verschlüsselung in AWS)
- Securing Your Block Storage on AWS (Sichern Ihres Blockspeichers in AWS)

SEC08-BP02 Erzwingen der Verschlüsselung im Ruhezustand

Sie sollten sicherstellen, dass die Verschlüsselung die einzige Möglichkeit zum Speichern von Daten bietet. AWS Key Management Service (AWS KMS) lässt sich nahtlos in viele AWS-Services integrieren, um Ihnen die Verschlüsselung aller Daten im Ruhezustand zu erleichtern. In Amazon Simple Storage Service (Amazon S3) können Sie beispielsweise die Standardverschlüsselung für einen Bucket festlegen, sodass alle neuen Objekte automatisch verschlüsselt werden. Darüber hinaus bietet Virtuelle Server-Instances in der Amazon Elastic Compute Cloud (Amazon EC2) und Amazon S3 Unterstützung für das Erzwingen der Verschlüsselung durch Festlegen der Standardverschlüsselung. Sie können AWS-Config-Regeln verwenden, um automatisch zu überprüfen, ob Sie die Verschlüsselung nutzen, z. B. für Amazon Elastic Block Store (Amazon EBS)-Volumes, Amazon Relational Database Service (Amazon RDS)-Instancesund Amazon S3-Buckets.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Erzwingen der Verschlüsselung von Daten im Ruhezustand für Amazon Simple Storage Service (Amazon S3): Implementieren Sie die Standardverschlüsselung für Amazon S3-Buckets.
  - Wie kann ich die Standardverschlüsselung für einen S3-Bucket aktivieren?
- Verwenden von AWS Secrets Manager: Secrets Manager ist ein AWS-Service für die einfache Verwaltung geheimer Schlüssel. Geheime Schlüssel können Datenbank-Anmeldeinformationen, Passwörter, API-Schlüssel von Dritten und sogar beliebiger Text sein.
  - AWS Secrets Manager
- Konfigurieren der Standardverschlüsselung für neue EBS-Volumes: Legen Sie fest, dass alle neu erstellten EBS-Volumes verschlüsselt erstellt werden sollen. Dabei können Sie den von AWS bereitgestellten Standardschlüssel oder einen von Ihnen erstellten Schlüssel verwenden.

    - Standardverschlüsselung für EBS-Volumes
- Konfigurieren verschlüsselter Amazon Machine Images (AMIs): Beim Kopieren eines vorhandenen AMI mit aktivierter Verschlüsselung werden Stammvolumes und Snapshots automatisch verschlüsselt.
    - AMIs mit verschlüsselten Snapshots
- Konfigurieren der Amazon Relational Database Service (Amazon RDS)-Verschlüsselung: Konfigurieren Sie die Verschlüsselung für Ihre Amazon RDS-Datenbank-Cluster und Snapshots im Ruhezustand durch Aktivieren der Verschlüsselungsoption.
    - Verschlüsseln von Amazon RDS-Ressourcen
- Konfigurieren der Verschlüsselung in weiteren AWS-Services: Bestimmen Sie die Verschlüsselungsfunktionen für die AWS-Services, die Sie nutzen.
    - AWS-Dokumentation

Ressourcen

Ähnliche Dokumente:

- AMIs mit verschlüsselten Snapshots
- AWS Crypto Tools
- AWS-Dokumentation
- AWS-Verschlüsselungs-SDK
- AWS KMS Cryptographic Details Whitepaper (Whitepaper mit kryptografischen Details zu AWS KMS)
- AWS Key Management Service
- AWS Secrets Manager
- Kryptografische AWS-Services und -Tools
- Amazon EBS-Verschlüsselung
- Standardverschlüsselung für EBS-Volumes
- Verschlüsseln von Amazon RDS-Ressourcen
- Wie kann ich die Standardverschlüsselung für einen S3-Bucket aktivieren?
- Protecting Amazon S3 Data Using Encryption (Amazon S3-Daten durch Verschlüsselung schützen)

Ähnliche Videos:

- How Encryption Works in AWS (So funktioniert die Verschlüsselung in AWS)
- Securing Your Block Storage on AWS (Sichern Ihres Blockspeichers in AWS)

SEC08-BP03 Automatisieren des Schutzes von Daten im Ruhezustand:

Verwenden Sie automatisierte Tools zur kontinuierlichen Validierung und Durchsetzung von Kontrollen, z. B. um sicherzustellen, dass nur verschlüsselte Speicherressourcen verwendet werden. Sie können die Validierung automatisieren, damit alle EBS-Volumes mit AWS-Config-Regelnspeichern AWS Security Hub kann auch verschiedene Kontrollen durch automatisierte Prüfungen auf Sicherheitsstandards überprüfen. Darüber hinaus können Ihre AWS-Config-Regeln nicht konforme Ressourcen automatisch korrigieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

Daten im Ruhezustand stellen alle Daten dar, die Sie für einen beliebigen Zeitraum in Ihrem Workload im nichtflüchtigen Speicher speichern. Die Daten können sich in Blockspeichern, Objektspeichern, Datenbanken, Archiven, IoT-Geräten und sonstigen Speichermedien befinden. Durch den Schutz Ihrer ruhenden Daten verringert sich das Risiko eines nicht autorisierten Zugriffs, wenn die Verschlüsselung und entsprechende Zugriffskontrollen implementiert werden.

Erzwingen der Verschlüsselung von Daten im Ruhezustand: Sie sollten sicherstellen, dass die Verschlüsselung die einzige Möglichkeit zum Speichern von Daten bietet. AWS KMS lässt sich nahtlos in viele AWS-Services integrieren, um Ihnen die Verschlüsselung aller Daten im Ruhezustand zu erleichtern. In Amazon Simple Storage Service (Amazon S3) können Sie beispielsweise die Standardverschlüsselung für einen Bucket festlegen, sodass alle neuen Objekte automatisch verschlüsselt werden. Darüber hinaus bietet Amazon EC2 und Amazon S3 Unterstützung für das Erzwingen der Verschlüsselung durch Festlegen der Standardverschlüsselung. Sie können AWS Managed Config Rules verwenden, um automatisch zu überprüfen, ob Sie die Verschlüsselung nutzen, z. B. für EBS-Volumes, Amazon Relational Database Service (Amazon RDS)-Instancesund Amazon S3-Buckets.

Ressourcen

Zugehörige Dokumente:

- AWS Crypto Tools
- AWS-Verschlüsselungs-SDK

Zugehörige Videos:

- How Encryption Works in AWS (So funktioniert die Verschlüsselung in AWS)
- Securing Your Block Storage on AWS (Sichern Ihres Blockspeichers in AWS)

SEC08-BP04 Durchsetzen der Zugriffskontrolle

Erzwingen Sie eine Zugriffskontrolle mit minimal erforderlichen Berechtigungen und Mechanismen, einschließlich Datensicherungen, Isolierung und Versionsverwaltung, zum Schutz Ihrer ruhenden Daten. Verhindern Sie, dass Operatoren öffentlichen Zugriff auf Ihre Daten gewähren.

Verschiedene Kontrollen z. B. für den Zugriff (mit dem Prinzip der geringsten Berechtigung), Backups (siehe Whitepaper zur Zuverlässigkeit), Isolierung und Versionsverwaltung können helfen, Ihre Daten im Ruhezustand zu schützen. Der Zugriff auf Ihre Daten sollte mit den zuvor in diesem Whitepaper behandelten Erkennungsmechanismen überprüft werden, einschließlich CloudTrail und Service Level-Protokoll, z. B. Amazon Simple Storage Service (Amazon S3)-Zugriffsprotokolle. Sie sollten inventarisieren, welche Daten öffentlich zugänglich sind, und planen, wie Sie die verfügbare Datenmenge im Laufe der Zeit reduzieren können. Amazon S3 Glacier Vault Lock und Amazon S3 Object Lock sind Funktionen, die eine obligatorische Zugriffskontrolle ermöglichen. Sobald eine Tresorrichtlinie mit der Compliance-Option gesperrt ist, kann sie nicht einmal der Root-Benutzer ändern, bis die Sperre abläuft. Der Mechanismus erfüllt die Anforderungen an die Aufzeichnungs- und Datenverwaltung der SEC, CFTC und FINRA. Weitere Informationen finden Sie in diesem Whitepaper.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Erzwingen der Zugriffskontrolle: Erzwingen Sie die Zugriffskontrolle nach dem Prinzip der geringsten Berechtigung, einschließlich des Zugriffs auf Verschlüsselungsschlüssel.
  - Einführung in die Verwaltung von Zugriffsberechtigungen für Ihre Amazon S3-Ressourcen
- Trennen von Daten anhand unterschiedlicher Klassifizierungsstufen: Verwenden Sie unterschiedliche AWS-Konten für die von AWS Organizations verwalteten Datenklassifizierungsstufen.
  - AWS Organizations
- Überprüfen von AWS KMS-Richtlinien: Überprüfen Sie die gewährte Zugriffsebene in den AWS KMS-Richtlinien.

  - Übersicht über die Verwaltung des Zugriffs auf Ihre AWS KMS-Ressourcen
- Überprüfen der Berechtigungen für Amazon S3-Buckets und -Objekte: Überprüfen Sie regelmäßig den in Amazon S3-Bucket-Richtlinien gewährten Zugriff. Als Best Practice gilt, keine öffentlich lesbaren oder schreibbaren Buckets zu haben. Erwägen Sie, AWS Config zur Erkennung von öffentlich verfügbaren Buckets und Amazon CloudFront für die Bereitstellung von Inhalten aus Amazon S3 zu verwenden.
  - AWS-Config-Regeln
  - Amazon S3 + Amazon CloudFront: Die perfekte Kombination in der Cloud
- Aktivieren Sie die Amazon S3-Versionsverwaltung und Objektsperre.
  - Verwenden von Versioning
  - Sperren von Objekten mit der Amazon S3-Objektsperre
- Verwenden von Amazon S3 Inventory: Amazon S3 Inventory ist eines der Tools, mit denen Sie den Replikations- und Verschlüsselungsstatus Ihrer Objekte prüfen und melden können.
  - Amazon S3 Inventory
- Überprüfen von Amazon EBS- und AMI-Freigabeberechtigungen: Mit Freigabeberechtigungen können Images und Volumes für AWS-Konten außerhalb Ihrer Workload freigegeben werden.
  - Teilen eines Amazon EBS-Snapshots
  - Gemeinsame AMIs

Ressourcen

Ähnliche Dokumente:

- AWS KMS Cryptographic Details Whitepaper (Whitepaper mit kryptografischen Details zu AWS KMS)

Ähnliche Videos:

- Securing Your Block Storage on AWS (Sichern Ihres Blockspeichers in AWS)

SEC08-BP05 Verwenden von Mechanismen, die den direkten Zugriff auf Daten verhindern

Halten Sie alle Benutzer davon ab, unter normalen Betriebsbedingungen direkt auf sensible Daten und Systeme zuzugreifen. Verwenden Sie beispielsweise einen Änderungsmanagement-Workflow, um Amazon Elastic Compute Cloud (Amazon EC2)-Instances mithilfe von Tools zu verwalten, statt

direkten Zugriff oder Zugriff über einen Bastion-Host zuzulassen. Dies kann mit AWS Systems Manager Automationerreicht werden. Dabei werden Automatisierungsdokumente verwendet, welche die Anweisungen enthalten, um Automationsaufgaben auszuführen. Diese Dokumente können in der Quellcodeverwaltung gespeichert und von Kollegen vor ihrer Ausführung geprüft und gründlich getestet werden. Das Vorgehen minimiert die Risiken im Vergleich zu direktem Shell-Zugriff. Geschäftliche Benutzer könnten statt direktem Zugriff ein Dashboard erhalten, um Abfragen auszuführen. Bestimmen Sie, wenn keine CI/CD-Pipelines verwendet werden, welche Kontrollen und Prozesse erforderlich sind, um einen normalerweise deaktivierten Mechanismus für den Notfallzugriff bereitzustellen.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Implementieren von Mechanismen, die den direkten Zugriff auf Daten verhindern: Mechanismen umfassen die Verwendung von Dashboards wie Amazon QuickSight, um Benutzern Daten anzuzeigen, anstatt direkt abzufragen.
  - Amazon QuickSight
- Automatisieren der Konfigurationsverwaltung: Führen Sie Aktionen aus der Ferne aus und erzwingen und validieren Sie sichere Konfigurationen automatisch. Verwenden Sie dazu einen Service oder ein Tool zur Konfigurationsverwaltung. Vermeiden Sie die Verwendung von Bastion-Hosts oder den direkten Zugriff auf EC2-Instances.
  - AWS Systems Manager
  - AWS CloudFormation
  - CI/CD-Pipeline für AWS CloudFormation-Vorlagen in AWS

Ressourcen

Zugehörige Dokumente:

- AWS KMS Cryptographic Details Whitepaper (Whitepaper mit kryptografischen Details zu AWS KMS)

Zugehörige Videos:

- How Encryption Works in AWS (So funktioniert die Verschlüsselung in AWS)
- Securing Your Block Storage on AWS (Sichern Ihres Blockspeichers in AWS)

SICH 9 Wie schützen Sie Ihre Daten bei der Übertragung?

Schützen Sie Ihre Daten während der Übertragung, indem Sie mehrere Kontrollen implementieren, um das Risiko eines unbefugten Zugriffs oder Verlusts zu reduzieren.

Bewährte Methoden

- SEC09-BP01 Implementieren einer sicheren Schlüssel- und Zertifikatverwaltung
- SEC09-BP02 Erzwingen einer Verschlüsselung bei der Übertragung
- SEC09-BP03 Automatisieren der Erkennung von unbeabsichtigtem Datenzugriff
- SEC09-BP04 Authentifizieren der Netzwerkkommunikation

SEC09-BP01 Implementieren einer sicheren Schlüssel- und Zertifikatverwaltung

Speichern Sie Verschlüsselungsschlüssel und Zertifikate sicher und ändern Sie sie in angemessenen Zeitintervallen mit strenger Zugriffskontrolle. Um dies zu erreichen, verwenden Sie am besten einen verwalteten Service, wie z. B. AWS Certificate Manager (ACM). Damit können Sie problemlos öffentliche und private TLS-Zertifikate (Transport Layer Security) zur Verwendung mit AWS-Services und Ihren internen verbundenen Ressourcen verwalten und bereitstellen. TLS-Zertifikate werden verwendet, um die Netzwerkkommunikation zu sichern und die Identität von Websites über das Internet sowie Ressourcen in privaten Netzwerken zu bestimmen. ACM lässt sich in AWS-Ressourcen wie Elastic Load Balancers (ELBs), AWS-Verteilungen und APIs auf API Gateway integrieren und verarbeitet auch automatische Zertifikatserneuerungen. Wenn Sie ACM verwenden, um eine private Root-CA bereitzustellen, können von ihr sowohl Zertifikate als auch private Schlüssel zur Verwendung in Amazon Elastic Compute Cloud (Amazon EC2)-Instances, Containern usw. bereitgestellt werden.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren einer sicheren Schlüssel- und Zertifikatverwaltung: Implementieren Sie die definierte Lösung zur sicheren Schlüssel- und Zertifikatverwaltung.
  - AWS Certificate Manager
  - Hosten und Verwalten einer ganzen privaten Zertifikatinfrastruktur in AWS
- Implementieren sicherer Protokolle: Verwenden Sie sichere Protokolle wie Transport Layer Security (TLS) oder IPsec, die Authentifizierung und Vertraulichkeit bieten, um das Risiko der Manipulation

oder des Verlusts von Daten zu reduzieren. Überprüfen Sie die AWS-Dokumentation auf Protokolle und Sicherheitsinformationen, die für die von Ihnen verwendeten Services relevant sind.

Ressourcen

Zugehörige Dokumente:

- AWS-Dokumentation

SEC09-BP02 Erzwingen einer Verschlüsselung bei der Übertragung

Erzwingen Sie Ihre definierten Verschlüsselungsanforderungen basierend auf geeigneten Standards und Empfehlungen, damit Sie Ihre Unternehmens-, Rechts- und Compliance-Anforderungen erfüllen können. AWS-Services bieten HTTPS-Endpunkte, die für die Kommunikation TLS nutzen. Dadurch werden die Daten bei der Kommunikation mit den AWS-APIs während der Übertragung verschlüsselt. Unsichere Protokolle, wie z. B. HTTP, können in einer VPC durch die Verwendung von Sicherheitsgruppen überprüft und blockiert werden. HTTP-Anforderungen können auch automatisch an HTTPS in Amazon CloudFront oder auf einen Application Load Balancer. Sie haben uneingeschränkte Kontrolle über Ihre Datenverarbeitungsressourcen und können die Verschlüsselung während der Übertragung in alle Ihre Services implementieren. Darüber hinaus können Sie die VPN-Konnektivität mit Ihrer VPC von einem externen Netzwerk aus verwenden, um die Verschlüsselung des Datenverkehrs zu erleichtern. Sollten Sie besondere Anforderungen haben, finden Sie Lösungen von Drittanbietern im AWS Marketplace.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Erzwingen der Verschlüsselung bei der Übertragung: Die definierten Verschlüsselungsanforderungen sollten sich nach den neuesten Standards und Best Practices richten und nur sichere Protokolle zulassen. Konfigurieren Sie beispielsweise eine Sicherheitsgruppe, die nur das HTTPS-Protokoll für einen Application Load Balancer oder eine Amazon Elastic Compute Cloud (Amazon EC2)-Instance zulässt.
- Konfigurieren von sicheren Protokollen bei Edge-Services: Konfigurieren Sie HTTPS mit Amazon CloudFront und die erforderlichen Verschlüsselungsverfahren.
  - Verwenden von HTTPS mit CloudFront

- Verwenden eines Virtual Private Network (VPN) für die externe Konnektivität: Ziehen Sie ein IPsec-VPN in Betracht, um Punkt-zu-Punkt- oder Netzwerk-zu-Netzwerk-Verbindungen zu sichern und so den Datenschutz und die Datenintegrität zu gewährleisten.
  - VPN-Verbindungen
- Konfigurieren von sicheren Protokollen bei Load Balancern: Aktivieren Sie HTTPS-Listener, um die Verbindung zu Load Balancern zu sichern.
  - HTTPS-Listener für den Application Load Balancer
- Konfigurieren von sicheren Protokollen für Instances: Ziehen Sie eine HTTPS-Verschlüsselung für Instances in Betracht.
  - Tutorial: SSL/TLS unter Amazon Linux 2 konfigurieren
- Konfigurieren sicherer Protokolle in Amazon Relational Database Service (Amazon RDS): Verwenden Sie Secure Socket Layer (SSL) oder Transport Layer Security (TLS) zum Verschlüsseln der Verbindung zu Datenbank-Instances.
  - Verwenden von SSL zum Verschlüsseln einer Verbindung zu einer Datenbank-Instance
- Konfigurieren sicherer Protokolle in Amazon Redshift: Konfigurieren Sie Ihr Cluster so, dass eine SSL- oder TLS-Verbindung vorgeschrieben ist.
  - Konfigurieren von Sicherheitsoptionen für Verbindungen
- Konfigurieren sicherer Protokolle in weiteren AWS-Services: Bestimmen Sie die Funktionen zur Verschlüsselung während der Übertragung für die AWS-Services, die Sie nutzen.

Ressourcen

Zugehörige Dokumente:

- AWS-Dokumentation

SEC09-BP03 Automatisieren der Erkennung von unbeabsichtigtem Datenzugriff

Verwenden Sie Tools wie Amazon GuardDuty zum automatischen Erkennen von verdächtigen Aktivitäten oder Versuchen, Daten außerhalb definierter Grenzen zu verschieben. GuardDuty kann beispielsweise ungewöhnliche Amazon Simple Storage Service (Amazon S3)-Leseaktivitäten erkennen. Verwendet wird dafür Exfiltration:S3/AnomalousBehavior. Zusätzlich zu GuardDuty können auch Amazon VPC Flow Logs, die die Netzwerkverkehrsinformationen erfassen, zusammen mit Amazon EventBridge verwendet werden, um die Erkennung anormaler Verbindungen – sowohl

erfolgreich als auch abgelehnt – zu berichten. Mit Amazon S3 Access Analyzer können Sie ermitteln, welche Daten für wen in Ihren Amazon S3-Buckets zugänglich sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren der Erkennung von unbefugtem Datenzugriff: Setzen Sie Tools oder Erkennungsmechanismen ein, die automatisch erkennen, wenn versucht wird, Daten außerhalb festgelegter Grenzen zu verschieben. Damit lässt sich beispielsweise ein Datenbanksystem erkennen, das Daten auf einen unbekannten Host kopiert.
  - VPC Flow Logs
- Erwägen von Amazon Macie: Amazon Macie ist ein vollständig verwalteter Service für Datensicherheit und Datenschutz, der mithilfe von Machine Learning und Mustervergleichen Ihre sensiblen Daten in AWS erkennt und schützt.
  - Amazon Macie

Ressourcen

Ähnliche Dokumente:

- VPC Flow Logs
- Amazon Macie

SEC09-BP04 Authentifizieren der Netzwerkkommunikation

Überprüfen Sie die Identität der Kommunikation mithilfe von Protokollen, die die Authentifizierung unterstützen, wie Transport Layer Security (TLS) oder IPsec.

Durch die Verwendung von Netzwerkprotokollen, die die Authentifizierung unterstützen, kann eine Vertrauensstellung zwischen den kommunizierenden Einheiten hergestellt werden. Dadurch wird die im Protokoll verwendete Verschlüsselung hinzugefügt, um das Risiko zu verringern, dass die Kommunikation geändert oder abgefangen wird. Häufig verwendete Protokolle, die die Authentifizierung implementieren, sind Transport Layer Security (TLS), das in vielen AWS-Services verwendet wird, sowie IPsec, welches in AWS Virtual Private Network (AWS VPN) verwendet wird.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Implementieren sicherer Protokolle: Verwenden Sie sichere Protokolle wie TLS oder IPsec, die Authentifizierung und Vertraulichkeit bieten, um das Risiko der Manipulation oder des Verlusts von Daten zu reduzieren. Überprüfen Sie die AWS-Dokumentation auf Protokolle und Sicherheitsinformationen, die für die von Ihnen verwendeten Services relevant sind.

Ressourcen

Ähnliche Dokumente:

- AWS-Dokumentation

Vorfallsreaktion

Frage

- SICH 10 Wie können Sie Vorfälle voraussagen, darauf reagieren und diese beheben?

SICH 10 Wie können Sie Vorfälle voraussagen, darauf reagieren und diese beheben?

Die Vorbereitung ist entscheidend für eine rechtzeitige und effektive Untersuchung, Reaktion auf und Wiederherstellung nach Sicherheitsvorfällen, um Unterbrechungen der Geschäftsabläufe zu minimieren.

Bewährte Methoden

- SEC10-BP01 Identifizieren wichtiger Mitarbeiter und externer Ressourcen
- SEC10-BP02 Entwickeln von Vorfallmanagementplänen
- SEC10-BP03 Vorbereiten forensischer Funktionen
- SEC10-BP04 Automatische Eingrenzung
- SEC10-BP05 Vorab bereitgestellter Zugriff
- SEC10-BP06 Vorabbereitstellen von Tools
- SEC10-BP07 Durchführen von Gamedays

SEC10-BP01 Identifizieren wichtiger Mitarbeiter und externer Ressourcen

Ermitteln Sie interne und externe Mitarbeiter und Ressourcen, die bei Auftreten eines Vorfalls reagieren können.

Wenn Sie Ihren Ansatz zur Vorfallreaktion in der Cloud definieren, müssen Sie in Zusammenarbeit mit anderen Teams (z. B. Rechtsberater, Geschäftsleitung, Business-Stakeholder, AWS-Support-Services usw.) wichtige Mitarbeiter, Interessengruppen und relevante Kontakte identifizieren. Um Abhängigkeiten zu reduzieren und die Reaktionszeit zu verkürzen, müssen Sie sicherstellen, dass Ihr Team, die spezialisierten Sicherheitsteams und die Kundendienstmitarbeiter über die Services informiert sind, die Sie nutzen, und die Gelegenheit erhalten, praktische Erfahrungen zu sammeln.

Wir empfehlen Ihnen, externe AWS-Sicherheitspartner zu identifizieren, die Ihnen externes Fachwissen und eine andere Perspektive bieten können, um Ihre Reaktionsfähigkeit zu verbessern. Ihre vertrauenswürdigen Sicherheitspartner können Ihnen dabei helfen, potenzielle Risiken oder Bedrohungen zu identifizieren, mit denen Sie möglicherweise nicht vertraut sind.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Identifizieren wichtiger Mitarbeiter in Ihrer Organisation: Pflegen Sie eine Kontaktliste mit Mitarbeitern Ihrer Organisation, die bei Eintreten eines Vorfalls hinzugezogen werden müssen, um darauf zu reagieren und die Sicherheit wiederherzustellen.
- Identifizieren externer Partner: Beauftragen Sie gegebenenfalls externe Partner, die bei der Reaktion auf einen Vorfall und bei der Wiederherstellung der Sicherheit behilflich sein können.

Ressourcen

Zugehörige Dokumente:

- AWS Security Incident Response Guide (AWS-Sicherheitsleitfaden für die Vorfallreaktion)

Zugehörige Videos:

- How to prepare for and respond to security incidents in your AWS environment (Vorbereiten und Reagieren auf Sicherheitsvorfälle in Ihrer AWS-Umgebung)

Zugehörige Beispiele:

SEC10-BP02 Entwickeln von Vorfallmanagementplänen

Erstellen Sie Pläne, die Ihnen helfen, auf einen Vorfall zu reagieren, während des Vorfalls zu kommunizieren und im Anschluss den ursprünglichen Zustand wiederherzustellen. Beispielsweise können Sie einen Vorfallreaktionsplan mit den wahrscheinlichsten Szenarien für Ihren Workload und Ihre Organsation starten. Diese Pläne sollten Vorgehensweisen zur internen und externen Kommunikation und Eskalation enthalten.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Ein Vorfallreaktionsplan ist von entscheidender Bedeutung, um auf Sicherheitsvorfälle zu reagieren, sie einzudämmen und ihre potenziellen Folgen zu beheben. Ein Vorfallmanagementplan ist ein strukturierter Prozess für die Identifizierung und Behebung von Sicherheitsvorfällen sowie die zeitgerechte Reaktion darauf.

In der Cloud gibt es viele der betrieblichen Rollen und Anforderungen, die auch für eine On-Premises-Umgebung typisch sind. Bei der Erstellung eines Vorfallmanagementplans ist es wichtig, Reaktions- und Wiederherstellungsstrategien zu berücksichtigen, die optimal zu Ihren Anforderungen an geschäftliche Ergebnisse und Compliance passen. Wenn Sie beispielsweise Workloads in AWS bearbeiten, die mit FedRAMP in den USA kompatibel sind, sollten Sie den NIST SP 800-61 Computer Security Handling Guide berücksichtigen. Ähnlich gilt beim Betrieb von Workloads mit personenbezogenen Informationen in Europa, dass Sie an Szenarien denken sollten, in denen Sie diese schützen und auf Probleme reagieren müssen, die im Zusammenhang mit den Bestimmungen zu Datenspeicherorten der Regulierungen der Datenschutz-Grundverordnung (DSGVO) der EU stehen.

Wenn Sie einen Vorfallmanagementplan für Ihre Workloads in AWS erstellen, beginnen Sie mit dem AWS-Modell der geteilten Verantwortungzum Aufbau eines gründlichen Verteidigungskonzepts im Rahmen Ihrer Vorfallreaktionen. In diesem Modell kümmert sich AWS um die Sicherheit der Cloud und Sie sind für die Sicherheit in der Cloud verantwortlich. Dies bedeutet, dass Sie die Kontrolle behalten und für die Sicherheitskontrollen verantwortlich sind, für deren Implementierung Sie sich entscheiden. Der Leitfaden für AWS Security Incident Response enthält zentrale Konzepte und grundlegende Anleitungen für den Aufbau eines cloudbasierten Vorfallmanagementplans.

Ein effektiver Vorfallmanagementplan muss kontinuierlich iteriert und stets an die Ziele Ihrer Cloud-Operationen angepasst werden. Erwägen Sie die Verwendung der nachfolgend erläuterten Implementierungspläne für die Erstellung und Weiterentwicklung Ihres Vorfallmanagementplans.

- Aufklärung und Training für die Reaktion auf Vorfälle: Wenn eine Abweichung von Ihrer definierten Baseline auftritt (etwa eine irrtümliche Bereitstellung oder eine fehlerhafte Konfiguration), müssen Sie darauf reagieren und den Vorfall untersuchen. Um dies erfolgreich tun zu können, müssen Sie wissen, welche Steuerungen und Funktionen Sie für die Reaktion auf Sicherheitsvorfälle innerhalb Ihrer AWS-Umgebung verwenden können und welche Prozesse Sie berücksichtigen müssen, um Ihre Teams, die an Notfallreaktionen beteiligt sind, darauf vorzubereiten und entsprechend auszubilden und zu schulen.
  - Playbooks und Runbooks sind effektive Mechanismen für die Gewährleistung von Konsistenz beim Training zur Reaktion auf Vorfälle. Beginnen Sie mit der Erstellung einer ersten Liste häufig durchgeführter Verfahren während einer Vorfallreaktion und entwickeln Sie diese ständig weiter, während Sie diese anzuwenden lernen.
  - Machen Sie die Playbooks und Runbooks im Rahmen geplanter Ernstfallübungen bekannt. Simulieren Sie bei solchen Ernstfallübungen die Vorfallreaktion in einer kontrollierten Umgebung, damit Ihr Team weiß, wie es zu reagieren hat, und um sicherzustellen, dass die an Vorfallreaktionen beteiligten Teams die entsprechenden Abläufe gut kennen. Überprüfen Sie die Ergebnisse dieser Simulationen, um Verbesserungsmöglichkeiten zu erkennen und um weiteren Bedarf an Trainings oder Tools feststellen zu können.
  - Die Sicherheit fällt in den Verantwortungsbereich aller. Sorgen Sie für gemeinsames Wissen zum Vorfallreaktionsprozess, indem Sie alle Personen daran beteiligen, die normalerweise an Ihren Workloads arbeiten. Eine Ernstfallübung betrifft alle Bereiche Ihres Unternehmens: Betrieb, Tests, Entwicklung, Sicherheit, Geschäftsbetrieb und Geschäftsleiter.
- Dokumentieren Sie den Vorfallmanagementplan: Dokumentieren Sie die Tools und die Prozesse zur Aufzeichnung, Behandlung, Fortschrittskommunikation und Benachrichtigung im Zusammenhang mit aktiven Vorfällen. Ein Vorfallmanagementplan verfolgt das Ziel, sicherzustellen, dass der Normalbetrieb so schnell wie möglich wiederhergestellt wird, dass die geschäftlichen Auswirkungen minimiert bleiben und dass alle beteiligten Parteien stets darüber informiert sind. Beispiele für Vorfälle sind der Verlust oder die Beeinträchtigung der Netzwerkkonnektivität, nicht reagierende Prozesse oder APIs, das Ausbleiben der Durchführung einer geplanten Aufgabe (beispielsweise ausbleibendes Patching), die Nichtverfügbarkeit von Anwendungsdaten oder Services, ungeplante Serviceunterbrechungen aufgrund von Sicherheitsvorfällen, Offenlegungen von Anmeldeinformationen oder Fehler durch falsche Konfigurationen.
  - Identifizieren Sie den primären Eigentümer, der für die Behebung des Vorfalls verantwortlich ist. Dies kann beispielsweise der Workload-Eigentümer sein. Machen Sie deutlich, wer für den Vorfall verantwortlich sein wird und wie die Kommunikation ablaufen soll. Wenn mehrere

Parteien am Prozess der Vorfallbehebung beteiligt sind, etwa noch ein externer Anbieter, dann sollten Sie eine Verantwortungs-Matrix (RACI-Matrix)erstellen, die die Rollen und Verantwortlichkeiten der einzelnen Teams oder Personen für die Behebung des Vorfalls aufführt.

Eine RACI-Matrix führt Folgendes auf:

- R: Responsible – Zuständige Partei, die die Arbeiten durchführt
- A: Accountable (Verantwortlich) – Verantwortliche(r) Partei oder Beteiligter mit endgültiger Autorität über die Durchführung der konkreten Aufgabe
- C: Consulted (Konsultiert) – Hinzugezogene Partei, deren Meinung eingeholt wird, typischerweise gehören dazu sachkundige Experten
- I: Informed – Informierte Partei, die über den Fortschritt auf dem Laufenden gehalten wird, oft nur bei Abschluss der Aufgabe oder Fertigstellung des Liefergegenstands.

- Kategorisierung von Vorfällen: Das Definieren und Kategorisieren von Vorfällen nach ihrem Schweregrad und ihren Auswirkungen ermöglicht das strukturierte Vorgehen bei der Beurteilung und Behebung von Vorfällen. Die folgenden Empfehlungen illustrieren eine Auswirkung-bis-Lösung-Dringlichkeitsmatrix für die Quantifizierung eines Vorfalls. So gilt etwa ein Vorfall mit geringen Auswirkungen und niedriger Dringlichkeit als Vorfall mit niedrigem Schweregrad.
  - Hoch (H): Ihre Geschäftstätigkeit ist stark betroffen. Kritische Funktionen Ihrer Anwendung im Zusammenhang mit AWS-Ressourcen sind nicht verfügbar. Reserviert für schwerste Vorfälle mit Auswirkungen auf Produktionssysteme. Die Auswirkungen des Vorfalls nehmen schnell zu und die Behebung muss möglichst schnell erfolgen.
  - Mittel (M): Ein Geschäftsservice oder eine Anwendung im Zusammenhang mit AWS-Ressourcen ist in mittelschwerer Weise betroffen und funktioniert mit Einschränkungen. Anwendungen, die zu Service-Level-Zielen (SLOs) beitragen, sind im Rahmen des Service Level Agreement (SLA) betroffen. Systeme können auch ohne allzu große Auswirkungen auf Finanzen oder den Ruf des Unternehmens mit reduzierter Kapazität funktionieren.
  - Niedrig (L): Nichtkritische Funktionen Ihres Geschäftsservice oder Ihrer Anwendung im Zusammenhang mit AWS-Ressourcen sind betroffen. Systeme können ohne allzu große Auswirkungen auf Finanzen oder den Ruf des Unternehmens mit reduzierter Kapazität weiterarbeiten.
- Standardisieren Sie Sicherheitskontrollen: Das Ziel der Standardisierung der Sicherheitskontrollen besteht darin, Konsistenz, Nachverfolgbarkeit und Wiederholbarkeit hinsichtlich der betrieblichen Ergebnisse zu erzielen. Unterstützen Sie die Standardisierung für zentrale Aktivitäten, die für die Vorfallreaktion von zentraler Bedeutung sind, z. B.:

- Identitäts- und Zugriffsmanagement: Richten Sie Mechanismen für die Kontrolle des Zugriffs auf Ihre Daten sowie für die Verwaltung der Berechtigungen für menschliche und maschinelle Identitäten ein. Erweitern Sie Ihr eigenes Identitäts- und Zugriffsmanagement in die Cloud und nutzen Sie Verbundsicherheit mit Single Sign-on und rollenbasierten Berechtigungen zur Optimierung des Zugriffsmanagements. Empfehlungen zu bewährten Methoden und Verbesserungspläne für die Standardisierung des Zugriffsmanagements finden Sie im Abschnitt zum Thema Identitäts- und Zugriffsmanagement im Whitepaper „Security Pillar“.
- Management von Schwachstellen: Richten Sie Mechanismen zur Identifizierung von Schwachstellen in Ihrer AWS-Umgebung ein, die von Angreifern ausgenutzt werden können, um Ihr System zu beschädigen oder zu missbrauchen. Implementieren Sie präventive und erkennende Kontrollen als Sicherheitsmechanismen, um auf Sicherheitsvorfälle reagieren und mögliche Auswirkungen mindern zu können. Standardisieren Sie Prozesse wie die Bedrohungsmodellierung im Rahmen Ihres Infrastrukturbuilds und Ihres Anwendungsbereitstellungslebenszyklus.
- Konfigurationsverwaltung: Definieren Sie Standardkonfigurationen und automatisieren Sie Verfahren für die Bereitstellung von Ressourcen in der AWS Cloud. Die Standardisierung der Bereitstellung von Infrastruktur und Ressourcen hilft bei der Eindämmung der Gefahr von Fehlkonfigurationen durch irrtümliche Bereitstellungen oder versehentliche Fehlkonfigurationen durch menschliche Bediener. Im Abschnitt zu den Designprinzipien des Whitepapers „Operational Excellence Pillar“ finden Sie Anleitungen und Verbesserungspläne zur Implementierung dieser Steuerung.
- Protokollierung und Überwachung für Audit Control: Implementieren Sie Mechanismen zur Überwachung Ihrer Ressourcen auf Ausfälle, Leistungseinbußen und Sicherheitsprobleme. Die Standardisierung dieser Kontrollen sorgt auch für Prüfungsprotokolle zu den in Ihrem System stattfindenden Aktivitäten und hilft so bei der zeitnahen Beurteilung und Behebung von Problemen. Bewährte Methoden unter SEC04 („Wie erkennen und untersuchen Sie Sicherheitsereignisse?“) bieten Anleitungen für die Implementierung dieser Steuerung.

- Verwenden Sie Automatisierung: Eine Automatisierung ermöglicht die zeitnahe Behebung von Vorfällen in großem Umfang. AWS bietet verschiedene Services für die Automatisierung im Kontext der Vorfallreaktionsstrategie. Konzentrieren Sie sich auf das angemessene Gleichgewicht zwischen Automatisierung und manuellen Eingriffen. Beim Aufbau Ihrer Vorfallreaktion in Playbooks und Runbooks sollten Sie wiederholbare Schritte automatisieren. Verwenden Sie AWS-Services wie AWS Systems Manager Incident Manager, um IT-Vorfälle schneller beheben zu können. Verwenden Sie Entwicklertools für die Versionssteuerung und die Automatisierung von Amazon Machine Images (AMI) sowie Infrastructure as Code (IaC)-Bereitstellungen ohne menschliche

Interventionen. Automatisieren Sie wo möglich die Erkennung und die Compliancеprüfung mithilfe verwalteter Services wie Amazon GuardDuty, Amazon Inspector, AWS Security Hub, AWS Config und Amazon Macie. Optimieren Sie die Erkennungsfunktionen mit Machine Learning wie Amazon DevOps Guru, um abnorme Betriebsmuster zu erkennen, bevor sie zu Problemen führen.

- Führen Sie Ursachenanalysen durch und setzen Sie Erkenntnisse um: Implementieren Sie Mechanismen zum Erfassen von Erkenntnissen für abschließende Überprüfungen. Wenn die Ursache für einen Vorfall ein größerer Defekt, ein Konstruktionsfehler oder eine Fehlkonfiguration ist oder wenn die Möglichkeit der Wiederholung besteht, wird dies als Problem klassifiziert. In solchen Fällen sollten Sie das Problem analysieren und lösen, um Unterbrechungen des normalen Betrieb zu minimieren.

Ressourcen

Zugehörige Dokumente:

- Leitfaden für AWS Security Incident Response
- NIST: Computer Security Incident Handling Guide

Zugehörige Videos:

- Automating Incident Response and ForensicsAWS (Automatisieren der Vorfallreaktion und Forensik in AWS)
- DIY guide to runbooks, incident reports, and incident response (DIY-Leitfaden für Runbooks, Vorfallberichte und Vorfallreaktion)
- Prepare for and respond to security incidents in your AWS environment (Vorbereiten und Reagieren auf Sicherheitsvorfälle in Ihrer AWS-Umgebung)

Zugehörige Beispiele:

- Übung: Playbook für Vorfallreaktion mit Jupyter – AWS IAM
- Übung: Vorfallreaktion mit AWS-Konsole und CLI

SEC10-BP03 Vorbereiten forensischer Funktionen

Es ist wichtig, dass Ihre Notfallteams wissen, wann und wie forensische Untersuchungen sich in Ihren Reaktionsplan eingliedern. Ihre Organisation sollte definieren, welche Nachweise erfasst und welche

Tools dafür verwendet werden. Identifizieren und bereiten Sie forensische Untersuchungsfunktionen vor, die geeignet sind, und beziehen Sie externe Spezialisten, Tools und Automatisierung mit ein. Eine wichtige Entscheidung, die Sie vorab treffen sollten, ist, ob Sie Daten von einem Live-System erfassen. Manche Daten wie die Inhalte von flüchtigem Speicher oder aktiver Netzwerkverbindungen gehen verloren, wenn das System abgeschaltet oder neu gestartet wird.

Ihr Notfallteam kann Tools wie AWS Systems Manager, Amazon EventBridge und AWS Lambda kombinieren, um automatisch Forensiktools in einem laufenden System auszuführen und mittels VPC-Datenverkehrsspiegelung ein Netzwerkpaketabbild zu erhalten, sodass nicht persistente Nachweise gesammelt werden können. Führen Sie andere Aktivitäten wie Protokollanalysen oder die Analyse von Datenträgerabbildern in einem dedizierten Sicherheitskonto mit individuellen Forensik-Workstations und für Ihr Notfallteam zugänglichen Tools aus.

Legen Sie relevante Protokolle regelmäßig in einem Datenspeicher mit hoher Widerstandsfähigkeit und Integrität ab. Notfallteams sollten auf diese Protokolle zugreifen können. AWS bietet verschiedene Tools zur Vereinfachung der Protokolluntersuchung, z. B. Amazon Athena, Amazon OpenSearch Service (OpenSearch Service) und Amazon CloudWatch Logs Insights. Zudem sollten Sie Nachweise mit Amazon Simple Storage Service (Amazon S3) Object Lock sicher aufbewahren. Dieser Service arbeitet nach dem WORM-Modell (Write Once, Read Many) und verhindert, dass Objekte über einen gewissen Zeitraum gelöscht oder überschrieben werden. Da forensische Untersuchungstechniken eine spezielle Schulung erfordern, müssen Sie möglicherweise externe Spezialisten engagieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Ermitteln forensischer Funktionen: Recherchieren Sie die forensischen Untersuchungsfunktionen in Ihrer Organisation, verfügbare Tools und externe Spezialisten.
- Automating Incident Response and Forensics

Ressourcen

Zugehörige Dokumente:

- How to automate forensic disk collection in AWS (Automatisieren der forensischen Datenträgererfassung in AWS)

SEC10-BP04 Automatische Eingrenzung

Automatisieren Sie die Eingrenzung eines Vorfalls und die Wiederherstellung, um die Reaktionszeiten und Auswirkungen auf Ihr Unternehmen zu reduzieren.

Sobald Sie die Prozesse und Tools aus Ihren Playbooks erstellt und trainiert haben, können Sie die Logik in eine codebasierte Lösung überführen, die von vielen Notfallteams als Tool verwendet werden kann, um die Antwort zu automatisieren und Abweichungen oder Unsicherheit im Notfallteam zu beseitigen. Dies kann den Lebenszyklus einer Reaktion beschleunigen. Das nächste Ziel besteht darin, diesen Code vollständig zu automatisieren, damit er von den Warnungen oder Ereignissen selbst aufgerufen wird, statt von einem Mitarbeiter des Notfallteams. So wird eine ereignisgesteuerte Antwort erstellt. Diese Prozesse sollten auch relevante Daten automatisch zu Ihren Sicherheitssystemen hinzufügen. Bei einem Vorfall mit Datenverkehr von einer unerwünschten IP-Adresse kann beispielsweise automatisch eine AWS WAF-Sperrliste oder eine Network Firewall-Regelgruppe ergänzt werden, um weitere Aktivitäten zu verhindern.

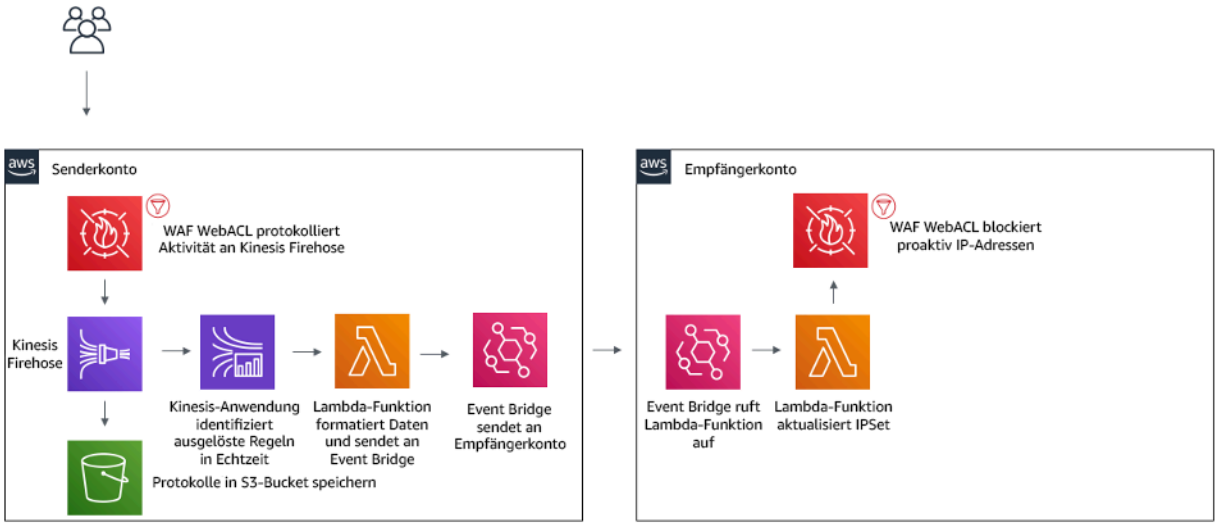


Abbildung 3: Automatisierte Blockierung bekannter böswilliger IP-Adressen mit AWS WAF

Bei einem ereignisgesteuerten Antwortsystem löst ein Mechanismus zur Aufdeckung eine Reaktion aus, um das Ereignis automatisch zu beheben. Sie können ereignisgesteuerte Antwortfunktionen verwenden, um die Wertschöpfung zwischen Aufdeckung und Reaktion zu beschleunigen. Zum Erstellen dieser ereignisgesteuerten Architektur können Sie AWS Lambda verwenden. Dabei handelt es sich um einen serverlosen Datenverarbeitungsservice, der Ihren Code als Reaktion auf Ereignisse ausführt und automatisch die zugrunde liegenden Datenverarbeitungsressourcen für Sie verwaltet. Angenommen, Sie haben ein AWS-Konto mit aktiviertem AWS CloudTrail-Service. Wenn AWS CloudTrail jemals deaktiviert wird (über den API-Aufruf `cloudtrail:StopLogging`), können Sie

Amazon EventBridge verwenden, um das spezifische `cloudtrail:StopLogging` -Ereignis zu überwachen und eine AWS Lambda-Funktion zum Aufrufen von `cloudtrail:StartLogging` nutzen, um die Protokollierung neu zu starten.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

Automatisieren Sie die Eindämmungsfunktionen.

Ressourcen

Zugehörige Dokumente:

- AWS Security Incident Response Guide (AWS-Sicherheitsleitfaden für die Vorfallreaktion)

Zugehörige Videos:

- How to prepare for and respond to security incidents in your AWS environment (Vorbereiten und Reagieren auf Sicherheitsvorfälle in Ihrer AWS-Umgebung)

SEC10-BP05 Vorab bereitgestellter Zugriff

Stellen Sie sicher, dass Notfallteams über den richtigen vorab bereitgestellten Zugriff in AWS verfügen, um die Zeit von der Untersuchung bis zur Wiederherstellung zu verkürzen.

Typische Anti-Muster:

- Verwenden des Root-Kontos für die Reaktion auf Vorfälle
- Verändern bestehender Benutzerkonten
- Direkte Manipulation von IAM-Berechtigungen bei Bereitstellung von Just-in-time-Berechtigungserhöhungen

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Mittel

Implementierungsleitfaden

AWS empfiehlt die Reduzierung oder Ausschaltung der Abhängigkeit von langlebigen Anmeldeinformationen wenn möglich und ihren Ersatz durch Just-in-Time-

Berechtigungseskalationsmechanismen. Langlebige Anmeldeinformationen sind anfällig für Sicherheitsrisiken und erhöhen den Verwaltungsaufwand. Für die meisten Managementaufgaben sowie für Vorfallreaktionsaufgaben empfehlen wir die Implementierung eines Identitätsverbunds neben der temporären Eskalierung für den administrativen Zugriff. In diesem Modell beantragt ein Benutzer seine Erhöhung auf eine höhere Berechtigungsstufe (etwa zu einer Vorfallreaktionsrolle). Anschließend wird, sofern der Benutzer grundsätzlich dafür infrage kommt, eine Anfrage an einen Genehmiger gesendet. Wenn die Anfrage genehmigt wurde, erhält der Benutzer einen Satz temporärer AWS-Anmeldeinformationen für die Durchführung seiner Aufgaben. Wenn diese Anmeldeinformationen ablaufen, muss der Benutzer eine neue Erhöhungsanfrage stellen.

Wir empfehlen für die meisten Vorfallreaktionsszenarien die Verwendung temporärer Berechtigungseskalierungen. Die korrekte Vorgehensweise ist die Verwendung von AWS Security Token Service und von Sitzungsrichtlinien zur Festlegung der Zugriffsbereiche.

Es gibt Szenarien, in denen Verbundidentitäten nicht verfügbar sind, zum Beispiel:

- Ausfall durch Problem mit einem Identitätsanbieter (IdP)
- Fehlerhafte Konfiguration oder menschlicher Fehler, die/der das Managementsystem für den Verbundzugriff beschädigt
- Böswillige Aktivität, z. B. ein DDoS-Angriff (Distributed Denial of Service) oder anderweitig verursachte Nichtverfügbarkeit des Systems

Für diese Fälle sollte Notfall- „Break Glass“- Zugriff konfiguriert werden, um Untersuchungen und die schnelle Behebung des Vorfalls zu ermöglichen. Wir empfehlen die Verwendung eines IAM-Benutzers mit ausreichenden Berechtigungen für die Durchführung von Aufgaben und den Zugriff auf AWS-Ressourcen. Verwenden Sie die Root-Anmeldeinformationen nur für Aufgaben, die Root-Benutzerzugriff erfordern. Zur Prüfung, ob die Vorfallreaktionskräfte über die korrekte Zugriffsstufe auf AWS und andere relevante Systeme verfügen, empfehlen wir die Bereitstellung dedizierter Benutzerkonten. Die Benutzerkonten erfordern privilegierten Zugriff und müssen eng kontrolliert und überwacht werden. Die Konten müssen mit den geringstmöglichen Berechtigungen versehen sein, die für die erforderlichen Aufgaben benötigt werden, und die Zugriffsstufe muss auf den Playbooks basieren, die Teil des Vorfallmanagementplans sind.

Verwenden Sie als bewährte Methode zweckgerichtet erstellte und dedizierte Benutzer und Rollen. Die vorübergehende Eskalierung des Zugriffs eines Benutzers oder einer Rolle über IAM-Richtlinien macht es unklar, welche Zugriffsmöglichkeiten Benutzer während eines Vorfalls hatten, und birgt die Gefahr, dass die eskalierten Berechtigungen später nicht widerrufen werden.

Es ist wichtig, so viele Abhängigkeiten wie möglich zu entfernen, um sicherzustellen, dass Zugriff bei einer möglichst großen Anzahl von Ausfallszenarien möglich ist. Erstellen Sie deshalb ein Playbook, um sicherzustellen, dass Vorfallreaktionsbenutzer als AWS Identity and Access Management-Benutzer in einem dedizierten Sicherheitskonto erstellt und nicht durch einen vorhandenen Verbund oder eine Single Sign-On (SSO)-Lösung verwaltet werden. Alle einzelnen Reaktionskräfte müssen ein eigenes benanntes Konto haben. Die Kontokonfiguration muss eine Richtlinie für sichere Passwörter und Multi-Faktor-Authentifizierung (MFA) durchsetzen. Wenn die Playbooks zur Vorfallreaktion nur Zugriff auf die AWS Management Console benötigen, sollten für den Benutzer keine Zugriffsschlüssel konfiguriert werden und er sollte auch explizit keine Zugriffsschlüssel erstellen dürfen. Dies kann mit IAM-Richtlinien oder Service-Kontrollrichtlinien (SCPs) konfiguriert werden, wie in den bewährten AWS-Sicherheitsmethoden für AWS Organizations SCPs erläutert. Die Benutzer sollten keine Berechtigungen außer der Möglichkeit zur Übernahme von Vorfallreaktionsrollen in anderen Konten haben.

Während eines Vorfalls kann es erforderlich sein, anderen internen oder externen Personen Zugriff zu gewähren, um Untersuchungs-, Korrektur- oder Wiederherstellungsaktivitäten zu unterstützen. Verwenden Sie in diesem Fall den vorher erwähnten Playbook-Mechanismus. Darüber hinaus muss ein Prozess vorhanden sein, um sicherzustellen, dass jeglicher zusätzliche Zugriff sofort nach Abschluss des Vorfalls widerrufen wird.

Zur Sicherstellung, dass die Verwendung von Vorfallreaktionsrollen in korrekter Weise überwacht und geprüft werden kann, ist es entscheidend, dass die für diesen Zweck erstellten IAM-Benutzerkonten nicht zwischen Personen weitergegeben werden und dass der AWS-Konto-Root-Benutzer nicht verwendet wird, sofern dies nicht für eine bestimmte Aufgabe erforderlich ist. Wenn der Root-Benutzer erforderlich ist (zum Beispiel wenn der IAM-Zugriff auf ein bestimmtes Konto nicht verfügbar ist), verwenden Sie einen separaten Prozess mit einem Playbook, um die Verfügbarkeit des Root-Benutzer-Passworts und des MFA-Tokens zu prüfen.

Erwägen Sie zur Konfiguration der IAM-Richtlinien für die Vorfallreaktionsrollen die Verwendung von IAM Access Analyzer zum Erstellen von Richtlinien auf der Grundlage von AWS CloudTrail-Protokollen. Gewähren Sie dazu der Vorfallreaktionsrolle in einem Nicht-Produktionskonto Administratorzugriff und durchlaufen Sie das Playbook. Sobald dies geschehen ist, kann eine Richtlinie erstellt werden, die nur die entsprechenden Aktionen zulässt. Diese Richtlinie kann dann auf alle Vorfallreaktionsrollen über alle Konten hinweg angewendet werden. Möglicherweise möchten Sie eine separate IAM-Richtlinie für jedes Playbook erstellen, um Management und Auditing zu vereinfachen. Beispiel-Playbooks können Reaktionspläne für Ransomware-Angriffe, Datenschutzverletzungen, Verlust von produktionsrelevantem Zugriff oder andere Szenarien enthalten.

Verwenden Sie die Vorfallreaktionsbenutzerkonten zur Annahme dedizierter Vorfallreaktions-IAM-Rollen in anderen AWS-Konten. Diese Rollen müssen so konfiguriert sein, dass sie nur von Benutzern im Sicherheitskonto angenommen werden können, und das Vertrauensverhältnis muss erfordern, dass der aufrufende Prinzipal per MFA authentifiziert wurde. Die Rollen müssen eng gefasste IAM-Richtlinien verwenden, um den Zugriff zu kontrollieren. Stellen Sie sicher, dass alle `AssumeRole-` Anfragen für diese Rollen in CloudTrail protokolliert und gemeldet werden und dass alle mit diesen Rollen durchgeführten Aktivitäten protokolliert werden.

Es wird nachdrücklich empfohlen, die IAM-Benutzerkonten und die IAM-Rollen deutlich zu benennen, damit sie in CloudTrail-Protokollen leicht zu finden sind. Ein Beispiel ist die Benennung der IAM-Konten als *<USER_ID>*`-BREAK-GLASS` und der IAM-Rollen als `BREAK-GLASS-ROLE`.

CloudTrail wird verwendet, um API-Aktivitäten in Ihren AWS-Konten zu protokollieren, und sollte zur Konfiguration von Alarmen zur Nutzung der Vorfallreaktionsrollen eingesetzt werden. Weitere Informationen finden Sie im Blog-Beitrag zur Konfiguration von Alarmen bei Verwendung von Root-Schlüsseln. Die Anweisungen können können geändert werden, um die Metrik Amazon CloudWatch so zu konfigurieren, dass sie nach `AssumeRole-` Ereignissen gefiltert wird, die mit der Vorfallreaktions-IAM-Rolle zusammenhängen.

```
{ $.eventName = "AssumeRole" && $.requestParameters.roleArn =
 "<INCIDENT_RESPONSE_ROLE_ARN>" && $.userIdentity.invokedBy NOT EXISTS && $.eventType !
= "AwsServiceEvent" }
```

Da die Vorfallreaktionsrollen sehr wahrscheinlich eine hohe Zugriffsstufe haben, ist es wichtig, dass diese Alarme an eine breite Gruppe gehen und dass sofort darauf reagiert wird.

Während eines Vorfalls kann es geschehen, dass eine Reaktionskraft Zugriff auf Systeme benötigt, die nicht direkt von IAM gesichert sind. Dazu können Amazon Elastic Compute Cloud-Instances, Amazon Relational Database Service-Datenbanken oder SaaS-Plattformen gehören. Es wird nachdrücklich empfohlen, anstelle nativer Protokolle wie SSH oder RDP AWS Systems Manager Session Manager für alle administrativen Zugriffe auf Amazon EC2-Instances zu verwenden. Dieser Zugriff kann mit IAM (sicher und geprüft) kontrolliert werden. Es kann auch möglich sein, Teile Ihrer Playbooks mit AWS Systems Manager-Run-Command-Dokumentenzu automatisieren, wodurch sich möglicherweise Benutzerfehler reduzieren und Wiederherstellungszeiten verkürzen lassen. Für den Zugriff auf Datenbanken und Tools von Drittanbietern empfehlen wir die Speicherung von Anmeldeinformationen in AWS Secrets Manager und die Gewährung des Zugriffs auf die Vorfallreaktionsrollen.

Schließlich sollte die Verwaltung der Vorfallreaktions-IAM-Benutzerkonten Ihren Joiners-, Movers- und Leavers-Prozessen hinzugefügt sowie regelmäßig geprüft und getestet werden, um sicherzustellen, dass nur die beabsichtigten Zugriffsrechte gewährt werden.

Ressourcen

Zugehörige Dokumente:

- Verwaltung des vorübergehend erhöhten Zugriffs auf Ihre AWS-Umgebung
- Leitfaden für AWS Security Incident Response
- AWS Elastic Disaster Recovery
- AWS Systems Manager Incident Manager
- Einrichten einer Kontopasswortrichtlinie für IAM-Benutzer
- Verwenden der Multi-Faktor-Authentifizierung (MFA) in AWS
- Konfigurieren des kontoübergreifenden Zugriffs mit MFA
- Verwenden von IAM Access Analyzer zum Erstellen von IAM-Richtlinien
- Bewährte Methoden für AWS Organizations-Servicekontrollrichtlinien in einer Mehrkontenumgebung
- Empfang von Benachrichtigungen, wenn die Root-Zugriffsschlüssel Ihres AWS-Kontos verwendet werden
- Erstellen detaillierter Sitzungsberechtigungen mithilfe von IAM-verwalteten Richtlinien

Zugehörige Videos:

- Automating Incident Response and ForensicsAWS (Automatisieren der Vorfallreaktion und Forensik in AWS)
- DIY guide to runbooks, incident reports, and incident response (DIY-Leitfaden für Runbooks, Vorfallberichte und Vorfallreaktion)
- Prepare for and respond to security incidents in your AWS environment (Vorbereiten und Reagieren auf Sicherheitsvorfälle in Ihrer AWS-Umgebung)

Zugehörige Beispiele:

- Übung: AWS-Kontoeinrichtung und Root-Benutzer
- Übung: Vorfallreaktion mit AWS-Konsole und CLI

SEC10-BP06 Vorabbereitstellen von Tools

Stellen Sie sicher, dass Sicherheitspersonal über die richtigen Tools in AWS verfügt, um die Zeit von der Untersuchung bis zur Wiederherstellung zu verkürzen.

Zur Automatisierung von Sicherheitstechnik und Betriebsfunktionen können Sie eine umfassende Palette von APIs und Tools von AWS verwenden. Sie können die Identitätsverwaltung, Netzwerksicherheit, Datenschutz und Überwachungsfunktionen vollständig automatisieren und diese mithilfe gängiger Softwareentwicklungsmethoden bereitstellen, die Sie bereits eingerichtet haben. Wenn Sie die Sicherheitsautomatisierung erstellen, kann Ihr System eine Reaktion überwachen, prüfen und initiieren, statt nur Ihre Sicherheitslage zu überwachen und manuell auf Ereignisse zu reagieren. Eine effektive Möglichkeit zum automatischen Bereitstellen durchsuchbarer und relevanter Protokolldaten in all Ihren AWS-Services für das Notfallteam besteht in der Aktivierung von Amazon Detective.

Wenn Ihre Vorfallreaktionsteams auf Warnungen weiterhin auf die gleiche Weise reagieren, riskieren sie eine Abstumpfung der Warnung. Im Laufe der Zeit kann das Team für Warnungen desensibilisiert werden und entweder Fehler bei der Verarbeitung normaler Situationen machen oder außergewöhnliche Warnungen übersehen. Automatisierung hilft, eine Abstumpfung von Warnungen zu vermeiden, indem Funktionen verwendet werden, die sich wiederholende und gewöhnliche Warnungen verarbeiten, sodass Mitarbeiter die nötigen freien Kapazitäten haben, um sich um sensible und einzigartige Vorfälle zu kümmern. Die Integration von Systemen zur Erkennung von Anomalien wie Amazon GuardDuty, AWS CloudTrail Insights und Amazon CloudWatch Anomaly Detection kann den durch schwellenwertbasierte Warnmeldungen verursachten Aufwand reduzieren.

Sie können manuelle Prozesse verbessern, indem Sie die Schritte im Prozess automatisieren. Nachdem Sie das Korrekturmuster für ein Ereignis definiert haben, können Sie dieses Muster in umsetzbare Logik zerlegen und den Code schreiben, um diese Logik auszuführen. Notfallteams können anschließend diesen Code ausführen, um das Problem zu beheben. Mit der Zeit können Sie immer mehr Schritte automatisieren und schließlich häufige Vorfälle automatisch verarbeiten.

Für Tools, die im Betriebssystem Ihrer Amazon Elastic Compute Cloud (Amazon EC2)-Instance ausgeführt werden, sollten Sie den AWS Systems Manager Run Command verwenden. Mit diesem können Sie einen Agent auf Ihrer Amazon EC2-Instance installieren und das Betriebssystem remote und sicher verwalten. Sie benötigen dafür den Systems Manager Agent (SSM Agent), der bei vielen Amazon Machine Images (AMIs) standardmäßig installiert ist. Beachten sollten Sie jedoch, dass kompromittierte Instances keine vertrauenswürdigen Reaktionen und Antworten von Tools oder den installierten Agents mehr senden und so behandelt werden sollten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Vorabbereitstellen von Tools: Stellen Sie sicher, dass in AWS die richtigen Tools für das Sicherheitspersonal vorab bereitgestellt wurden, damit bei einem Vorfall eine entsprechende Reaktion erfolgen kann.
  - Übung: Vorfallreaktion mit AWS Management Console und CLI
  - Playbook für Vorfallreaktion mit Jupyter – AWS IAM
  - AWS-Sicherheitsautomatisierung
- Implementieren des Ressourcenmarkierung: Markieren Sie Ressourcen mit Informationen, z. B. einem Code für die zu untersuchende Ressource, damit Sie Ressourcen während eines Vorfalls identifizieren können.
  - AWS-Markierungsstrategien

Ressourcen

Ähnliche Dokumente:

- AWS Security Incident Response Guide (AWS-Sicherheitsleitfaden für die Vorfallreaktion)

Ähnliche Videos:

- DIY guide to runbooks, incident reports, and incident response

SEC10-BP07 Durchführen von Gamedays

Testtage werden auch als Simulationen oder Übungen bezeichnet. Dabei handelt es sich um interne Ereignisse, die eine strukturierte Möglichkeit bieten, Ihre Vorfallmanagementpläne und -verfahren in einem realistischen Szenario zu üben. Diese Ereignisse sollten dem Notfallteam als Übung dienen und es sollten die gleichen Tools und Techniken wie in einem echten Szenario zum Einsatz kommen. Im Grunde sollten sogar echte Umgebungen nachgebildet werden. Bei Testtagen geht es im Wesentlichen um die Vorbereitung und die schrittweise Verbesserung Ihrer Reaktionsfähigkeiten. Vorteile von Testtagen:

- Validieren der Bereitschaft
- Fördern von Vertrauen – Lernen durch Simulationen und Schulung von Mitarbeitern

- Einhaltung der Compliance oder vertraglicher Verpflichtungen
- Generieren von Artefakten für die Akkreditierung
- Agilität – inkrementelle Verbesserung
- Schnelleres Arbeiten und Verbessern von Tools
- Verfeinern von Kommunikation und Eskalation
- Gewinn von Vertrautheit mit seltenen und unerwarteten Vorfällen

Diese Vorteile zeigen, weshalb die Teilnahme an einer Simulationsaktivität die Effizienz der Organisation bei kritischen Ereignissen erhöht. Die Entwicklung einer realistischen und nützlichen Simulationsaktivität kann schwierig sein. Obwohl das Testen Ihrer Verfahren oder der Automatisierung für bekannte Ereignisse gewisse Vorteile hat, ist es ebenso wertvoll, an kreativen Aktivitäten zur Simulation von Sicherheitsvorfällen (Security Incident Response Simulations (SIRS)) teilzunehmen, um sich auf unerwartete Ereignisse vorzubereiten und sich kontinuierlich zu verbessern.

Erstellen Sie individuelle Simulationen, die auf Ihre Umgebung, Ihr Team und Ihre Tools zugeschnitten sind. Ermitteln Sie ein Problem und richten Sie Ihre Simulation darauf aus. Das könnten beispielsweise weitergegebene Anmeldeinformationen, ein mit unerwünschten Systemen kommunizierender Server oder eine Fehlkonfiguration sein, die zu unzulässigen Risiken führt. Identifizieren Sie mit Ihrer Organisation vertraute Ingenieure zum Erstellen des Szenarios und eine andere Gruppe, die mitmacht. Das Szenario sollte realistisch und ausreichend anspruchsvoll sein, damit es auch nützlich ist. Es sollte Möglichkeiten für den praktischen Umgang mit Protokollen, Benachrichtigungen, Eskalationen und der Ausführung von Runbooks oder der Automatisierung bieten. Während der Simulation sollte Ihr Notfallteam sein technisches und organisatorisches Können üben und Führungskräfte sollten zur Verbesserung ihrer Vorfallmanagementkompetenzen einbezogen werden. Am Ende der Simulation sollten Sie die Leistungen des Teams würdigen und nach Optionen zum Iterieren, Wiederholen und Erweitern für weitere Simulationen suchen.

AWS hat Vorlagen für Runbooks zur Vorfallreaktion erstellt, die Sie nicht nur zur Vorbereitung Ihrer Reaktionsmaßnahmen, sondern auch als Basis für eine Simulation verwenden können. Bei der Planung kann eine Simulation in fünf Phasen aufgeteilt werden.

Sammeln von Beweisen: In dieser Phase erhält ein Team Warnmeldungen aus unterschiedlichen Quellen, z. B. von einem internen Ticketing-System und von Überwachungstools, aus anonymem Tipps oder sogar aus öffentlichen Nachrichten. Die Teams beginnen dann mit der Überprüfung der Infrastruktur- und Anwendungsprotokolle zum Bestimmen der Kompromittierungsquelle. In diesem

Schritt sollten auch interne Eskalationen und das Führungsteam für Vorfälle einbezogen werden. Nach der Identifizierung gehen die Teams zur Eindämmung des Vorfalls über.

Eindämmen des Vorfalls: An diesem Punkt haben die Teams bereits festgestellt, dass es einen Vorfall gegeben hat, und die Kompromittierungsquelle wurde ermittelt. Jetzt sollten die Teams Maßnahmen ergreifen, indem sie beispielsweise kompromittierte Anmeldeinformationen deaktivieren, eine Datenverarbeitungsressource isolieren oder einer Rolle die Berechtigungen entziehen.

Ausräumen des Vorfalls: Nach der Eindämmung des Vorfalls gehen die Teams jetzt zum Minimieren der Schwachstellen oder Infrastrukturkonfigurationen über, die anfällig für die Kompromittierung waren. Dafür könnten beispielsweise alle Anmeldeinformationen für eine Workload, Zugriffssteuerungslisten (ACLs) oder Netzwerkkonfigurationen geändert werden.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Ausführung Ernstfallübungen: Führen Sie simulierte Ereignisse zur Vorfall- reaktion (Gamedays) für verschiedene Bedrohungen aus, bei denen wichtige Mitarbeiter und das Management einbezogen werden.
- Erfassen von Erkenntnissen: Die aus den Ernstfallübungen gewonnenen Erkenntnisse sollten in das Feedback zur Verbesserung Ihrer Prozesse einfließen.

Ressourcen

Zugehörige Dokumente:

- AWS Security Incident Response Guide (AWS-Sicherheitsleitfaden für die Vorfallreaktion)
- AWS Elastic Disaster Recovery

Zugehörige Videos:

- DIY guide to runbooks, incident reports, and incident response

Zuverlässigkeit

Themen

- Grundlagen

- Workload-Architektur
- Änderungsverwaltung
- Fehlerverwaltung

Grundlagen

Fragen

- ZUV 1 Was ist bei der Verwaltung von Servicekontingenten und Einschränkungen zu beachten?
- ZUV 2 Was ist bei der Planung der Netzwerktopologie zu beachten?

ZUV 1 Was ist bei der Verwaltung von Servicekontingenten und Einschränkungen zu beachten?

Für cloudbasierte Workload-Architekturen gibt es Servicekontingente (die auch als Service Limits bezeichnet werden). Diese Kontingente dienen dazu, nicht versehentlich mehr Ressourcen bereitzustellen als nötig und Anfrageraten für API-Vorgänge zu begrenzen, um Services vor Missbrauch zu schützen. Darüber hinaus gibt es Ressourceneinschränkungen, z. B. die Rate, mit der Bits durch ein Glasfaserkabel geschleust werden können, oder die Speichermenge auf einer physischen Festplatte.

Bewährte Methoden

- REL01-BP01 Kenntnis von Servicekontingenten und Einschränkungen
- REL01-BP02 Verwalten von Servicekontingenten für mehrere Konten und Regionen
- REL01-BP03 Berücksichtigen von festen Servicekontingenten und Einschränkungen durch die Architektur
- REL01-BP04 Überwachen und Verwalten von Kontingenten
- REL01-BP05 Automatisieren der Kontingentverwaltung
- REL01-BP06 Sicherstellen eines ausreichenden Spielraums zwischen den aktuellen Kontingenten und der maximalen Nutzung, damit ein Failover möglich ist

REL01-BP01 Kenntnis von Servicekontingenten und Einschränkungen

Sie wissen über die Standardkontingente und Anfragen zur Kontingenterhöhung für Ihre Workload-Architektur Bescheid. Außerdem wissen Sie, welche Ressourceneinschränkungen, z. B. bezüglich Datenträgern oder Netzwerken, potenziell große Auswirkungen haben.

Service Quotas ist ein AWS-Service, mit dem Sie Ihre Kontingente für über 100 AWS-Services von einem Standort aus verwalten können. Neben der Suche nach den Kontingentwerten können Sie auch Kontingenterhöhungen über die Service Quotas-Konsole oder über das AWS SDK anfordern und nachverfolgen. AWS Trusted Advisor bietet eine Servicekontingent-Prüfung, die Ihre Nutzung und Ihre Kontingente für bestimmte Aspekte einiger Services anzeigt. Die Standardkontingente pro Service finden Sie ebenfalls in der AWS-Dokumentation für den jeweiligen Service. Weitere Informationen finden Sie unter Amazon VPC Quotas. Ratenlimits für gedrosselte APIs werden innerhalb des API Gateway selbst festgelegt. Dazu wird ein Nutzungsplan konfiguriert. Andere Limits, die für ihre jeweiligen Services konfiguriert werden, sind bereitgestellte IOPS, zugewiesener RDS-Speicher und EBS-Volume-Zuweisungen. Amazon Elastic Compute Cloud (Amazon EC2) verfügt über ein eigenes Service Limits-Dashboard, mit dem Sie Ihre Limits für Instances, Amazon Elastic Block Store (Amazon EBS) und Elastic IP-Adressen verwalten können. Wenn Sie einen Anwendungsfall haben, bei dem sich Servicekontingente auf die Leistung Ihrer Anwendung auswirken und eine Anpassung an Ihre Anforderungen nicht möglich ist, wenden Sie sich an den AWS Support, um zu ermitteln, ob es Lösungen gibt.

Gängige Antimuster:

- Bereitstellen einer Workload ohne Berücksichtigung der Servicekontingente für die verwendeten AWS-Services.
- Entwerfen einer Workload ohne Untersuchung und Berücksichtigung der Designeinschränkungen für AWS-Services.
- Bereitstellen einer vielgenutzten Workload, die eine bekannte vorhandene Workload ersetzt, ohne vorher die notwendigen Kontingente zu konfigurieren oder sich mit AWS Support in Verbindung zu setzen.
- Planen eines Ereignisses, das Datenverkehr zur Workload lenken soll, ohne vorher die notwendigen Kontingente zu konfigurieren oder sich mit AWS Support in Verbindung zu setzen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die Servicekontingente, API-Drosselungslimits und Designeinschränkungen kennen, können Sie diese Aspekte bei Entwurf, Implementierung und Betrieb der Workload berücksichtigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Überprüfen Sie die AWS-Servicekontingente in der veröffentlichten Dokumentation und in Service Quotas.
  - AWS Service Quotas (früher als Limits bezeichnet)
- Ermitteln Sie alle für die Workload erforderlichen Services durch Untersuchung des Bereitstellungscodes.
- Verwenden Sie AWS Config, um alle AWS-Ressourcen zu finden, die in Ihren AWS-Konten verwendet werden.
  - AWS Config-unterstützte AWS-Ressourcentypen
- Sie können auch Ihre AWS CloudFormation verwenden, um die genutzten AWS-Ressourcen zu ermitteln. Sehen Sie sich die Ressourcen an, die in der AWS Management Console oder über den Befehl „list-stack-resources“ in der Befehlszeilenschnittstelle erstellt wurden. Sie können zudem Ressourcen anzeigen, die für die Bereitstellung in der Vorlage selbst konfiguriert sind.
  - Anzeigen Ihrer AWS CloudFormation-Stack-Daten und -Ressourcen auf der AWS Management Console
  - AWS CLI for CloudFormation: list-stack-resources (AWS CLI für CloudFormation: list-stack-resources)
- Ermitteln Sie die geltenden Servicekontingente. Nutzen Sie die programmgesteuert über Trusted Advisor und Service Quotas zugänglichen Informationen.

Ressourcen

Ähnliche Dokumente:

- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- Bewährte AWS Trusted Advisor Trusted Advisor-Methoden (Prüfungen) (siehe Abschnitt „Servicelimits“)
- AWS Limit Monitor in AWS Answers
- Amazon EC2 Service Limits
- Was ist Service Quotas?

Ähnliche Videos:

- AWS Live re:Inforce 2019 – Service Quotas

REL01-BP02 Verwalten von Servicekontingenten für mehrere Konten und Regionen

Wenn Sie mehrere AWS-Konten oder AWS-Regionen verwenden, müssen Sie die entsprechenden Kontingente in allen Umgebungen anfordern, in denen die Produktions-Workloads ausgeführt werden.

Servicekontingente werden pro Konto aufgezeichnet. Sofern nicht anders angegeben, gilt jedes Kontingent für eine bestimmte AWS-Region. Zusätzlich zu den Produktionsumgebungen verwalten Sie auch Kontingente in allen anwendbaren Nicht-Produktionsumgebungen, damit Tests und Entwicklung nicht behindert werden.

Gängige Antimuster:

- Es wird zugelassen, dass die Ressourcennutzung in einer Isolationszone zunimmt, ohne dass es einen Mechanismus zur Aufrechterhaltung der Kapazität in den anderen Zonen gibt.
- Alle Kontingente werden manuell und in jeder Isolationszone einzeln festgelegt.
- Es wird nicht sichergestellt, dass regional isolierte Bereitstellungen groß genug sind, um bei Ausfall einer Bereitstellung den zunehmenden Datenverkehr aus einer anderen Region zu bewältigen.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die aktuelle Last bei Nichtverfügbarkeit einer Isolationszone bewältigen können, kann dies dabei helfen, dass beim Failover eine geringere Anzahl von Fehlern auftritt, anstatt dass dieser einen Denial-of-Service für die Kunden verursacht.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Wählen Sie relevante Konten und Regionen anhand von Serviceanforderungen, regulatorischen Anforderungen sowie Anforderungen für die Latenz und die Notfallwiederherstellung aus.
- Ermitteln Sie Servicekontingente für alle relevanten Konten, Regionen und Availability Zones. Die Limits gelten für ein Konto und eine Region.
- Was ist Service Quotas?

Ressourcen

Relevante Dokumente:

- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- Bewährte AWS Trusted Advisor Trusted Advisor-Methoden (Prüfungen) (siehe Abschnitt „Servicelimits“)
- AWS Limit Monitor in AWS Answers
- Amazon EC2 Service Quotas
- Was ist Service Quotas?

Relevante Videos:

- AWS Live re:Inforce 2019 - Service Quotas

REL01-BP03 Berücksichtigen von festen Servicekontingenten und Einschränkungen durch die Architektur

Achten Sie auf unveränderliche Servicekontingente und physische Ressourcen und gestalten Sie sie so, dass sie keine Auswirkungen auf die Zuverlässigkeit haben.

Beispiele sind Netzwerkbandbreite, AWS Lambda-Nutzlastgröße, Drosselungs-/Burst-Rate für API Gateway und gleichzeitige Benutzerverbindungen zu einem Amazon Redshift-Cluster.

Gängige Antimuster:

- Benchmarking erfolgt unter Ausnutzung des Burst-Limits nur für sehr kurze Zeit, anschließend wird aber erwartet, dass der Service über einen längeren Zeitraum mit der betreffenden Kapazität ausgeführt wird.
- Auswahl eines Designs, bei dem pro Benutzer oder Kunde eine Ressource eines Services verwendet wird, ohne die Einschränkungen des Designs zu kennen, die bei dessen Skalierung zum Ausfall führen.

Vorteile der Einführung dieser bewährten Methode: Durch die Nachverfolgung fester Kontingente in AWS-Services und von Einschränkungen in anderen Teilen der Workload (wie z. B. Einschränkungen in Bezug auf Konnektivität, IP-Adressen und Services von Drittanbietern) können Sie Trends hin

zu einem Kontingent erkennen. Außerdem können Sie so das Kontingent erweitern, bevor es überschritten wird.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Berücksichtigen Sie feste Servicekontingente. Achten Sie bei der Gestaltung der Architektur auf feste Servicekontingente und Einschränkungen.
  - AWS Service Quotas

Ressourcen

Ähnliche Dokumente:

- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- Bewährte AWS Trusted Advisor Trusted Advisor-Methoden (Prüfungen) (siehe Abschnitt „Servicelimits“)
- AWS Limit Monitor in AWS Answers
- Amazon EC2 Service Quotas
- Was ist Service Quotas?

Ähnliche Videos:

- AWS Live re:Inforce 2019 – Service Quotas

REL01-BP04 Überwachen und Verwalten von Kontingenten

Überprüfen Sie die potenzielle Nutzung und erhöhen Sie Ihre Kontingente entsprechend, um einen geplanten Nutzungsanstieg zu ermöglichen.

Für unterstützte Dienste können Sie Ihre Kontingente verwalten, indem Sie CloudWatch-Alarme konfigurieren. So wird die Nutzung überwacht und bei einer sich abzeichnenden Erschöpfung von Kontingenten werden Sie benachrichtigt. Diese Alarme können über Service Quotas oder Trusted Advisor ausgelöst werden. Sie können auch Metrikfilter für CloudWatch Logs verwenden, um Muster in Protokollen zu suchen und zu extrahieren und dadurch zu bestimmen, ob die Nutzung Kontingentschwellenwerte erreicht.

Gängige Antimuster:

- Es werden Alarme für den Fall konfiguriert, dass Service Quotas zur Neige gehen, aber es gibt keinen Prozess für die Reaktion auf eine entsprechende Warnung.
- Es werden nur Alarme für Services konfiguriert, die von Service Quotas unterstützt werden, und es erfolgt keine Überwachung anderer Services.

Vorteile der Einführung dieser bewährten Methode: Durch die automatische Verfolgung der AWS-Servicekontingente und die Überwachung ihrer Nutzung können Sie feststellen, wann ein Kontingent zu Neige geht. Mithilfe dieser Überwachungsdaten können Sie zudem abschätzen, wann Kontingente zur Kosteneinsparung verringert werden sollten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Überwachen und verwalten Sie Ihre Kontingente. Überprüfen Sie Ihre potenzielle Nutzung in AWS, erhöhen Sie Ihre regionalen Servicekontingente entsprechend und planen Sie eine steigende Nutzung ein.
  - Erfassen Sie die aktuelle Ressourcennutzung (z. B. Buckets, Instances). Erfassen Sie die aktuelle Ressourcennutzung mithilfe von Service-API-Vorgängen wie der DescribeInstances-API von Amazon EC2.
  - Erfassen Sie Ihre aktuellen Kontingente. Verwenden Sie AWS Service Quotas, AWS Trusted Advisor und AWS-Dokumentation.
    - Verwenden Sie AWS Service Quotas, ein AWS-Service, der Sie bei der Verwaltung von mehr als 100 AWS-Services an einem einzigen Ort unterstützt.
    - Ermitteln Sie Ihre aktuellen Service-Limits anhand von Trusted-Advisor-Service-Limits.
    - Ermitteln Sie aktuelle Servicekontingente mithilfe von Service-API-Vorgängen, sofern unterstützt.
    - Protokollieren von angeforderten Kontingenterhöhungen und deren Status Nachdem eine Kontingenterhöhung genehmigt wurde, sollten Sie sicherstellen, dass Sie Ihre Datensätze aktualisieren, um die Kontingentänderung widerzuspiegeln.

Ressourcen

Ähnliche Dokumente:

- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- AWS Trusted Advisor Überprüfen bewährter Methoden für Service Limits
- AWS Limit Monitor in AWS Answers
- Amazon EC2 Service Quotas
- Was ist Service Quotas?
- Überwachen von Service Quotas unter Verwendung von Amazon CloudWatch-Alarmen

Ähnliche Videos:

- AWS Live re:Inforce 2019 – Service Quotas

REL01-BP05 Automatisieren der Kontingentverwaltung

Implementieren Sie Tools, um vor dem Erreichen von Schwellenwerten benachrichtigt zu werden. Durch die Verwendung von AWS Service Quotas-APIs können Sie Anfragen zur Kontingenterhöhung automatisieren.

Wenn Sie Ihre Konfigurationsmanagementdatenbank (CMDB) oder das Ticketing-System mit Service Quotas integrieren, können Sie die Verfolgung von Kontingenterhöhungsanfragen und von aktuellen Kontingenten automatisieren. Zusätzlich zum AWS SDK bietet Service Quotas Automatisierung unter Verwendung der AWS Command Line Interface (AWS CLI).

Gängige Antimuster:

- Die Kontingente und die Nutzung werden in Tabellen verfolgt.
- Es werden Berichte zur täglichen, wöchentlichen oder monatlichen Nutzung ausgeführt und anschließend wird die Nutzung mit den Kontingenten verglichen.

Vorteile der Einführung dieser bewährten Methode: Durch die automatisierte Nachverfolgung der AWS-Servicekontingente und die Überwachung ihrer Nutzung können Sie feststellen, wann ein Kontingent zu Neige geht. Sie können die Automatisierung einrichten, damit Sie beim Anfordern einer Kontingenterhöhung bei Bedarf unterstützt werden. Wenn sich Ihre Nutzung in die entgegengesetzte Richtung entwickelt, sollten Sie einige Kontingente reduzieren, um von den verringerten Risiken (im Falle von kompromittierten Anmeldeinformationen) und von Kosteneinsparungen zu profitieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Richten Sie eine automatisierte Überwachung ein. Implementieren Sie Tools mithilfe von SDKs, um vor dem Erreichen von Schwellenwerten benachrichtigt zu werden.
  - Nutzen Sie Service Quotas und erweitern Sie den Service mit einer Lösung zur automatisierten Kontingentüberwachung, z. B. mit AWS Limit Monitor oder einem Angebot aus AWS Marketplace.
    - Was ist Service Quotas?
    - Quota Monitor on AWS – AWS-Lösung
  - Richten Sie automatische Reaktionen anhand von Schwellenwerten für Kontingente mit Amazon SNS- und AWS Service Quotas-APIs ein.
  - Testen Sie die Automatisierung.
    - Konfigurieren Sie Limit-Schwellenwerte.
    - Integrieren Sie Änderungsereignisse von AWS Config-Bereitstellungspipelines, Amazon EventBridge oder Ereignisse von Drittanbietern.
    - Legen Sie unnatürlich niedrige Schwellenwerte für Kontingente fest, um die Reaktionen zu testen.
    - Richten Sie Trigger ein, damit bei Benachrichtigungen geeignete Maßnahmen ergriffen werden und bei Bedarf der AWS Support kontaktiert wird.
    - Lösen Sie Änderungsereignisse manuell aus.
    - Führen Sie eine Ernstfallübung aus, um den Prozess für die Kontingenterhöhung zu testen.

Ressourcen

Ähnliche Dokumente:

- APN-Partner: Partner, die Sie bei der Konfigurationsverwaltung unterstützen können
- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- Bewährte AWS Trusted Advisor Trusted Advisor-Methoden (Prüfungen) (siehe Abschnitt „Servicelimits“)
- Quota Monitor on AWS – AWS-Lösung
- Amazon EC2 Service Limits
- Was ist Service Quotas?

Ähnliche Videos:

- AWS Live re:Inforce 2019 – Service Quotas

REL01-BP06 Sicherstellen eines ausreichenden Spielraums zwischen den aktuellen Kontingenten und der maximalen Nutzung, damit ein Failover möglich ist

Eine ausgefallene Ressource kann bis zu ihrer ordnungsgemäßen Beendigung weiterhin zu den Kontingenten zählen. Stellen Sie sicher, dass etwaige Überschneidungen aller ausgefallenen Ressourcen mit ihrem Ersatz bis zur Beendigung der ausgefallenen Ressourcen durch Ihre Kontingente abgedeckt sind. Berücksichtigen Sie bei der Berechnung des Spielraums auch den Ausfall einer Availability Zone.

Gängige Antimuster:

- Es werden Servicekontingente auf Grundlage des aktuellen Bedarfs eingerichtet, ohne dass Failover-Szenarien berücksichtigt werden.

Vorteile der Einführung dieser bewährten Methode: Wenn Ereignisse die Verfügbarkeit potenziell beeinträchtigen könnten, haben Sie die Möglichkeit, in der Cloud Strategien für die Abschwächung solcher Ereignisse oder für die Wiederherstellung im Anschluss daran zu implementieren. Zu solchen Strategien gehört häufig auch das Schaffen zusätzlicher Ressourcen, um solche zu ersetzen, die fehlgeschlagen sind. Ihre Kontingentstrategie muss diese zusätzlichen Ressourcen berücksichtigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Stellen Sie einen ausreichenden Spielraum zwischen Ihrem Servicekontingent und der maximalen Nutzung sicher, damit ein Failover möglich ist.
  - Ermitteln Sie die Servicekontingente unter Berücksichtigung von Bereitstellungsmustern, Verfügbarkeitsanforderungen und Nutzungsanstieg.
  - Fordern Sie bei Bedarf Kontingenterhöhungen an. Planen Sie den erforderlichen Zeitraum bis zur Bewilligung von Kontingenterhöhungen.
    - Ermitteln Sie die Anforderungen bezüglich der Zuverlässigkeit („Anzahl der Neunen“).
    - Legen Sie Fehlerszenarien fest (z. B. Verlust einer Komponente, Availability Zone oder Region).

- Führen Sie eine Bereitstellungsmethode ein (z. B. Canary, Blau/Grün-Bereitstellung, Rot/Schwarz-Bereitstellung oder schrittweise).
- Beziehen Sie einen angemessenen Puffer (z. B. 15 %) in aktuelle Limits ein.
- Planen Sie den Nutzungsanstieg (z. B. Überwachen des Nutzungstrends).

Ressourcen

Relevante Dokumente:

- AWS Marketplace: CMDB-Produkte zur Nachverfolgung von Limits
- AWS Service Quotas (früher als Service Limits bezeichnet)
- Bewährte AWS Trusted Advisor Trusted Advisor-Methoden (Prüfungen) (siehe Abschnitt „Servicelimits“)
- Amazon EC2 Service Limits
- Was ist Service Quotas?

Relevante Videos:

- AWS Live re:Inforce 2019 - Service Quotas

ZUV 2 Was ist bei der Planung der Netzwerktopologie zu beachten?

Workloads existieren häufig in mehreren Umgebungen. Dazu gehören mehrere Cloud-Umgebungen (öffentlich zugängliche und private) und möglicherweise die vorhandene Infrastruktur Ihres Rechenzentrums. Die Pläne müssen Netzwerkaspekte umfassen, wie z. B. die Konnektivität innerhalb und zwischen Systemen, die Verwaltung öffentlicher und privater IP-Adressen und die Auflösung von Domänennamen.

Bewährte Methoden

- REL02-BP01 Bereitstellen einer hochverfügbaren Netzwerkkonnektivität für öffentliche Endpunkte der Workload
- REL02-BP02 Bereitstellen redundanter Konnektivität zwischen privaten Netzwerken in der Cloud und in On-Premises-Umgebungen
- REL02-BP03 Berücksichtigen von Erweiterungen und Verfügbarkeit bei der Zuweisung von IP-Adressen für Subnetze

- REL02-BP04 Vorziehen von Nabe-und-Speiche-Topologien gegenüber M-zu-N-Netzen
- REL02-BP05 Erzwingen von sich nicht überschneidenden privaten IP-Adressbereichen in allen privaten Adressbereichen, in denen eine Verbindung besteht

REL02-BP01 Bereitstellen einer hochverfügbaren Netzwerkkonnektivität für öffentliche Endpunkte der Workload

Diese Endpunkte und das entsprechende Routing müssen hochverfügbar sein. Verwenden Sie dazu ein hochverfügbares DNS, Content Delivery Networks (CDNs), API Gateway, Load Balancing oder Reverse-Proxys.

Amazon Route 53, AWS Global Accelerator, Amazon CloudFront, Amazon API Gateway und Elastic Load Balancing (ELB) bieten allesamt hochverfügbare öffentliche Endpunkte. Sie können auch AWS Marketplace-Software-Appliances für Load Balancing und Proxyvorgänge auswerten.

Nutzer des Service, den Ihre Workload bereitstellt, unabhängig davon, ob es sich um Endbenutzer oder andere Services handelt, stellen Anfragen an diese Service-Endpunkte. Es stehen mehrere AWS-Ressourcen zur Verfügung, damit Sie hochverfügbare Endpunkte bereitstellen können.

Elastic Load Balancing bietet Load Balancing über Availability Zones hinweg, führt Layer 4 (TCP)- oder Layer 7-Routing (http/https) durch und lässt sich in AWS WAF und in AWS Auto Scaling integrieren, um eine Infrastruktur mit automatischer Fehlerbehebung zu erstellen und Datenverkehrssteigerungen zu absorbieren, während Ressourcen freigegeben werden, wenn der Datenverkehr abnimmt.

Amazon Route 53 ist ein skalierbares und hochverfügbares Domain Name System (DNS), das Benutzeranfragen auf effektive Weise mit über AWS bereitgestellter Infrastruktur verbindet – z. B. Amazon EC2-Instances, Elastic Load Balancing-Load Balancers oder Amazon S3-Buckets. Der Service kann außerdem zum Weiterleiten von Benutzern an Infrastruktur außerhalb von AWS eingesetzt werden.

AWS Global Accelerator ist ein Service auf Netzwerkebene, mit dem Sie Datenverkehr über das globale AWS-Netzwerk an optimale Endpunkte leiten können.

Distributed Denial of Service (DDoS)-Angriffe blockieren möglicherweise legitimen Datenverkehr und verringern die Verfügbarkeit für Ihre Benutzer. AWS Shield bietet automatischen Schutz gegen diese Angriffe ohne zusätzliche Kosten für AWS-Service-Endpunkte in Ihrer Workload. Sie können diese Funktionen zur Erfüllung Ihrer Anforderungen mit bei APN-Partnern und auf dem AWS Marketplace erhältlichen virtuellen Appliances erweitern.

Gängige Antimuster:

- Es werden öffentliche Internetadressen für Instances oder Container verwendet und die Verwaltung der Konnektivität zu ihnen erfolgt über DNS.
- Zum Suchen von Services werden IP-Adressen anstelle von Domänennamen verwendet.
- Inhalte (Webseiten, statische Komponenten, Mediendateien) werden in einem großen geografischen Bereich ohne ein CDN bereitgestellt.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie hochverfügbare Services in der Workload implementieren, haben Sie die Gewissheit, dass sie den Benutzern zur Verfügung steht.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Stellen Sie eine hochverfügbare Netzwerkkontinuität für die Benutzer der Workload sicher. Amazon Route 53, AWS Global Accelerator, Amazon CloudFront, Amazon API Gateway und Elastic Load Balancing (ELB) stellen allesamt hochverfügbare öffentliche Endpunkte bereit. Sie können auch AWS Marketplace-Software-Appliances für Load Balancing und Proxyvorgänge auswerten.

- Gewährleisten Sie eine hochverfügbare Verbindung zu den Benutzern.
- Verwenden Sie einen hochverfügbaren DNS zur Verwaltung der Domänennamen für die Anwendungsendpunkte.
  - Wenn die Benutzer über das Internet auf Ihre Anwendung zugreifen, sollten Sie mithilfe von Service-API-Vorgängen die korrekte Nutzung von Internet-Gateways überprüfen. Überprüfen Sie auch, ob die Einträge in den Routing-Tabellen für die Subnetze, die Ihre Anwendungsendpunkte hosten, korrekt sind.
    - DescribeInternetGateways
    - DescribeRouteTables
- Stellen Sie Ihrer Anwendung unbedingt einen hochverfügbaren Reverse-Proxy oder Load Balancer voran.
  - Wenn die Benutzer über Ihre On-Premises-Umgebung auf die Anwendung zugreifen, sollten Sie für eine hochverfügbare Verbindung zwischen AWS und der On-Premises-Umgebung sorgen.
  - Verwalten Sie Domänennamen mit Route 53.
    - Was ist Amazon Route 53?
  - Nutzen Sie einen externen DNS-Anbieter, der Ihre Anforderungen erfüllt.

- Nutzen Sie Elastic Load Balancing.
  - Was ist Elastic Load Balancing?
- Nutzen Sie eine AWS Marketplace-Appliance, die Ihren Anforderungen entspricht.

Ressourcen

Ähnliche Dokumente:

- APN-Partner: Partner, die Sie bei der Planung Ihres Netzwerks unterstützen können
- AWS Direct Connect Resiliency Recommendations (AWS Direct Connect-Resilienzempfehlungen)
- AWS Marketplace für Netzwerkinfrastruktur
- Amazon Virtual Private Cloud-Konnektivitätsoptionen – Whitepaper
- Hochverfügbare Netzwerkkonnektivität zwischen mehreren Rechenzentren
- Erste Schritte mit Direct Connect Resiliency Toolkit
- VPC-Endpunkte und VPC-Endpunktservices (AWS PrivateLink)
- Was ist AWS Global Accelerator?
- Was ist Amazon VPC?
- Was ist ein Transit-Gateway?
- Was ist Amazon CloudFront?
- Was ist Amazon Route 53?
- Was ist Elastic Load Balancing?
- Arbeiten mit Direct-Connect-Gateways

Ähnliche Videos:

- AWS re:Invent 2018: Erweitertes VPC-Design und neue Funktionen für Amazon VPC (NET303)
- AWS re:Invent 2019: AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs (NET406-R1)

REL02-BP02 Bereitstellen redundanter Konnektivität zwischen privaten Netzwerken in der Cloud und in On-Premises-Umgebungen

Verwenden Sie mehrere AWS Direct Connect-Verbindungen oder VPN-Tunnel zwischen separat bereitgestellten privaten Netzwerken. Verwenden Sie für eine hohe Verfügbarkeit mehrere Direct-

Connect-Standorte. Wenn Sie mehrere AWS-Regionen verwenden, stellen Sie in mindestens zwei davon Redundanz sicher. Erwägen Sie gegebenenfalls den Einsatz von AWS Marketplace-Appliances als Endpunkte von VPNs. Stellen Sie bei Verwendung von AWS Marketplace-Appliances redundante Instances bereit, um eine hohe Verfügbarkeit in verschiedenen Availability Zones zu gewährleisten.

Mit dem Cloud-Service AWS Direct Connect ist es einfach, eine dedizierte Netzwerkverbindung zwischen Ihrer On-Premises-Umgebung und AWS herzustellen. Mit Direct Connect Gateway kann Ihr On-Premises-Rechenzentrum mit mehreren AWS-VPCs verbunden werden, die über mehrere AWS-Regionen verteilt sind.

Diese Redundanz behebt mögliche Ausfälle, die sich auf die Ausfallsicherheit der Konnektivität auswirken:

- Wie können Sie sich gegen Fehler in Ihrer Topologie wappnen?
- Was passiert, wenn Sie etwas falsch konfigurieren oder die Konnektivität entfernen?
- Sind Sie in der Lage, eine unerwartete Erhöhung des Datenverkehrs bzw. der Nutzung Ihrer Services aufzufangen?
- Sind Sie in der Lage, den Versuch eines Distributed Denial of Service (DDoS)-Angriffs abzuwehren?

Berücksichtigen Sie bei der Verbindung Ihrer VPC mit Ihrem On-Premise-Rechenzentrum über VPN auch die Ausfallsicherheits- und Bandbreitenanforderungen, die Sie benötigen, wenn Sie den Anbieter und die Instance-Größe für die Ausführung der Appliance auswählen. Bei der Auswahl einer VPN-Appliance, die in ihrer Implementierung keine Ausfallsicherheit bietet, sollten Sie eine redundante Verbindung über eine zweite Appliance aufbauen. Bei all diesen Szenarios müssen Sie eine akzeptable Wiederherstellungszeit definieren und testen, um sicherzustellen, dass Sie diese Anforderungen erfüllen können.

Wenn Sie Ihre VPC über eine Direct-Connect-Verbindung mit Ihrem Rechenzentrum verbinden und diese Verbindung hochverfügbar sein muss, benötigen Sie redundante Direct-Connect-Verbindungen mit jedem Rechenzentrum. Die redundante Verbindung sollte eine zweite Direct-Connect-Verbindung von einem anderen Standort als der ersten verwenden. Wenn Sie mehrere Rechenzentren betreiben, stellen Sie sicher, dass Ihre Verbindungen an unterschiedlichen Orten enden. Verwenden Sie das Direct Connect Resiliency Toolkit, um dies einzurichten.

Wenn Sie sich für ein internetbasiertes Failover auf ein VPN mit einem AWS VPN entscheiden, ist es wichtig zu verstehen, dass es einen Datendurchsatz von bis zu 1,25 Gbit/s pro VPN-Tunnel bietet,

dass Equal Cost Multi Path (ECMP) für ausgehenden Datenverkehr jedoch nicht unterstützt wird, wenn mehrere von AWS verwaltete VPN-Tunnel auf demselben VGW enden. Wir raten davon ab, AWS Managed VPN als Sicherung für Direct-Connect-Verbindungen zu verwenden, es sei denn, Geschwindigkeiten von weniger als 1 Gbit/s während des Failovers stellen für Sie kein Problem dar.

Sie können VPC-Endpunkte auch verwenden, um Ihre VPC privat mit unterstützten AWS-Services und VPC-Endpunktservices zu verbinden, powered by AWS PrivateLink, ohne das öffentliche Internet zu durchlaufen. Endpunkte sind virtuelle Geräte. Sie sind horizontal skalierte, redundante und hochverfügbare VPC-Komponenten. Sie ermöglichen die Kommunikation zwischen Instances in Ihrer VPC und Ihren Services, ohne dass es zu Verfügbarkeitsrisiken oder Bandbreitenbeschränkungen für Ihren Netzwerkdatenverkehr kommt.

Gängige Antimuster:

- Einsatz nur eines Konnektivitätsanbieters zwischen dem lokalen Netzwerk und AWS.
- Die Konnektivitätsfunktionen der AWS Direct Connect-Verbindung werden genutzt, es gibt aber nur eine Verbindung.
- Es gibt nur einen Pfad für die VPN-Konnektivität.

Vorteile der Einführung dieser bewährten Methode: Durch die Implementierung redundanter Konnektivität zwischen Ihrer Cloud-Umgebung und Ihrer Unternehmens- bzw. On-Premises-Umgebung können Sie die sichere Kommunikation der abhängigen Services zwischen den beiden Umgebungen gewährleisten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Stellen Sie sicher, dass eine hochverfügbare Konnektivität zwischen AWS und der On-Premises-Umgebung vorhanden ist. Verwenden Sie mehrere AWS Direct Connect-Verbindungen oder VPN-Tunnel zwischen separat bereitgestellten privaten Netzwerken. Verwenden Sie für eine hohe Verfügbarkeit mehrere Direct-Connect-Standorte. Wenn Sie mehrere AWS-Regionen verwenden, stellen Sie in mindestens zwei davon Redundanz sicher. Erwägen Sie gegebenenfalls den Einsatz von AWS Marketplace-Appliances als Endpunkte von VPNs. Stellen Sie bei Verwendung von AWS Marketplace-Appliances redundante Instances bereit, um eine hohe Verfügbarkeit in verschiedenen Availability Zones zu gewährleisten.

- Stellen Sie sicher, dass eine redundante Verbindung zu Ihrer On-Premises-Umgebung besteht. Möglicherweise benötigen Sie redundante Verbindungen zu mehreren AWS-Regionen, um Ihre Verfügbarkeitsanforderungen zu erfüllen.
  - AWS Direct Connect Resiliency Recommendations (AWS Direct Connect-Resilienzempfehlungen)
  - Verwenden redundanter Site-to-Site-VPN-Verbindungen für Failover
    - Ermitteln Sie über die Service-API die ordnungsgemäße Nutzung von Direct-Connect-Verbindungen.
      - DescribeConnections
      - DescribeConnectionsOnInterconnect
      - DescribeDirectConnectGatewayAssociations
      - DescribeDirectConnectGatewayAttachments
      - DescribeDirectConnectGateways
      - DescribeHostedConnections
      - DescribeInterconnects
    - Wenn nur eine oder gar keine Direct-Connect-Verbindung besteht, richten Sie redundante VPN-Tunnel zu Ihren Virtual Private Gateways ein.
      - Was ist AWS-Site-to-Site VPN?
- Erfassen Sie die aktuelle Konnektivität (z. B. Direct Connect, Virtual Private Gateways, AWS Marketplace-Appliances).
  - Ermitteln Sie über die Service-API die Konfiguration von Direct-Connect-Verbindungen.
    - DescribeConnections
    - DescribeConnectionsOnInterconnect
    - DescribeDirectConnectGatewayAssociations
    - DescribeDirectConnectGatewayAttachments
    - DescribeDirectConnectGateways
    - DescribeHostedConnections
    - DescribeInterconnects
  - Erfassen Sie über die Service API die von Routing-Tabellen genutzten Virtual Private Gateways.
    - DescribeVpnGateways

    - DescribeRouteTables
  - Erfassen Sie über die Service-API die von Routing-Tabellen genutzten AWS Marketplace-Anwendungen.
    - DescribeRouteTables

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Planung Ihres Netzwerks unterstützen können
- AWS Direct Connect Resiliency Recommendations (AWS Direct Connect-Resilienzempfehlungen)
- AWS Marketplace für Netzwerkinfrastruktur
- Amazon Virtual Private Cloud-Konnektivitätsoptionen – Whitepaper
- Hochverfügbare Netzwerkkonnektivität zwischen mehreren Rechenzentren
- Verwenden redundanter Site-to-Site-VPN-Verbindungen für Failover
- Erste Schritte mit Direct Connect Resiliency Toolkit
- VPC-Endpunkte und VPC-Endpunktservices (AWS PrivateLink)
- Was ist Amazon VPC?
- Was ist ein Transit-Gateway?
- Was ist AWS-Site-to-Site VPN?
- Arbeiten mit Direct-Connect-Gateways

Relevante Videos:

- AWS re:Invent 2018: Erweitertes VPC-Design und neue Funktionen für Amazon VPC (NET303)
- AWS re:Invent 2019: AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs (NET406-R1)

REL02-BP03 Berücksichtigen von Erweiterungen und Verfügbarkeit bei der Zuweisung von IP-Adressen für Subnetze

Die IP-Adressbereiche für Amazon VPC müssen ausreichend groß sein, um die Anforderungen einer Workload zu erfüllen. Dabei sind zukünftige Erweiterungen und Zuweisungen von IP-Adressen zu

Subnetzen in verschiedenen Availability Zones zu berücksichtigen. Dies betrifft Load Balancer, EC2-Instances sowie containerbasierte Anwendungen.

Wenn Sie Ihre Netzwerktopologie planen, besteht der erste Schritt in der Definition des IP-Adressbereichs. Private IP-Adressbereiche (gemäß RFC 1918-Richtlinien) sollten jeder VPC zugewiesen werden. Berücksichtigen Sie im Rahmen dieses Prozesses die folgenden Anforderungen:

- Ermöglichen Sie einen IP-Adressbereich für mehr als eine VPC pro Region.
- Planen Sie innerhalb einer VPC Platz für mehrere Subnetze ein, die sich auf mehrere Availability Zones erstrecken.
- Lassen Sie für eine zukünftige Erweiterung stets Raum für nicht verwendete CIDR-Blöcke innerhalb einer VPC.
- Stellen Sie sicher, dass ein IP-Adressbereich vorhanden ist, um die Anforderungen von temporären EC2-Instances zu erfüllen, die Sie möglicherweise verwenden, z. B. Spot-Flotten für Machine Learning, Amazon EMR-Cluster oder Amazon Redshift-Cluster.
- Beachten Sie, dass die ersten vier IP-Adressen und die letzte IP-Adresse in jedem Subnetz-CIDR-Block reserviert und nicht für Sie verfügbar sind.
- Sie sollten die Bereitstellung großer VPC CIDR-Blöcke planen. Beachten Sie, dass der VPC CIDR-Block, der anfänglich Ihrer VPC zugewiesen war, nicht geändert oder gelöscht werden kann. Sie können der VPC jedoch zusätzliche, nicht überlappende CIDR-Blöcke hinzufügen. IPv4-CIDRs für Subnetze können nicht geändert werden, IPv6 CIDRs jedoch schon. Bedenken Sie, dass die Bereitstellung der größtmöglichen VPC (/16) mehr als 65 000 IP-Adressen zur Folge hat. Allein im IP-Adressbereich 10.x.x.x könnten Sie 255 solcher VPCs bereitstellen. Sie sollten daher eher auf eine zu große als eine zu kleine Lösung setzen, um die Verwaltung Ihrer VPCs zu vereinfachen.

Gängige Antimuster:

- Es werden kleine VPCs erstellt.
- Es werden kleine Subnetze erstellt und anschließend müssen beim Wachstum Subnetze zu Konfigurationen hinzugefügt werden.
- Es wird falsch eingeschätzt, wie viele IP-Adressen ein Elastic Load Balancer verwenden kann.
- Es werden viele Load Balancer mit hohem Datenverkehr in denselben Subnetzen bereitgestellt.

Vorteile der Einführung dieser bewährten Methode: So wird sichergestellt, dass Sie das Wachstum Ihrer Workloads bewältigen können und beim Hochskalieren weiterhin die entsprechende Verfügbarkeit bereitstellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Berücksichtigen Sie bei der Planung Ihres Netzwerks Ihr zukünftiges Wachstum, die Einhaltung gesetzlicher Vorschriften sowie die Kompatibilität mit anderen Netzwerken. Das Wachstum kann unterschätzt werden, gesetzliche Vorschriften können sich ändern, und bei Unternehmensübernahmen oder privaten Netzwerkverbindungen kann die Implementierung ohne fundierte Planung zur Herausforderung werden.
  - Wählen Sie relevante AWS-Konten und Regionen anhand von Serviceanforderungen, regulatorischen Anforderungen sowie Anforderungen für die Latenz und die Notfallwiederherstellung aus.
  - Identifizieren Sie Ihre Anforderungen bezüglich regionaler VPC-Bereitstellungen.
  - Ermitteln Sie die erforderliche Größe der VPCs.
    - Ermitteln Sie, ob Multi-VPC-Konnektivität bereitgestellt werden soll.
      - Was ist ein Transit-Gateway?
      - Multi-VPC-Konnektivität in einer Region
    - Ermitteln Sie, ob aufgrund von Compliance-Anforderungen getrennte Netzwerke erforderlich sind.
    - Legen Sie VPCs so groß wie möglich an. Der VPC-CIDR-Block, der anfänglich Ihrer VPC zugewiesen war, kann nicht geändert oder gelöscht werden. Sie können der VPC jedoch zusätzliche nicht überlappende CIDR-Blöcke hinzufügen. Dies kann jedoch zu einer Fragmentierung der Adressbereiche führen.
    - Legen Sie VPCs so groß wie möglich an. Der VPC-CIDR-Block, der anfänglich Ihrer VPC zugewiesen war, kann nicht geändert oder gelöscht werden. Sie können der VPC jedoch zusätzliche nicht überlappende CIDR-Blöcke hinzufügen. Dies kann jedoch zu einer Fragmentierung der Adressbereiche führen.

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Planung Ihres Netzwerks unterstützen können
- AWS Marketplace für Netzwerkinfrastruktur
- Amazon Virtual Private Cloud-Konnektivitätsoptionen – Whitepaper
- Hochverfügbare Netzwerkkonnektivität zwischen mehreren Rechenzentren
- Multi-VPC-Konnektivität in einer Region
- Was ist Amazon VPC?

Relevante Videos:

- AWS re:Invent 2018: Erweitertes VPC-Design und neue Funktionen für Amazon VPC (NET303)
- AWS re:Invent 2019: AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs (NET406-R1)

REL02-BP04 Vorziehen von Nabe-und-Speiche-Topologien gegenüber M-zu-N-Netzen

Wenn mehr als zwei Netzwerkadressbereiche (z. B. VPCs und On-Premises-Netzwerke) über VPC-Peering, AWS Direct Connect oder VPN verbunden sind, verwenden Sie ein Nabe-und-Speiche-Modell, wie es von AWS Transit Gateway bereitgestellt wird.

Wenn Sie nur zwei solche Netzwerke haben, können Sie sie einfach miteinander verbinden, doch wenn die Anzahl der Netzwerke zunimmt, ist die Komplexität derart vernetzter Verbindungen nicht mehr tragbar. AWS Transit Gateway bietet ein einfach zu wartendes Nabe-zu-Speiche-Modell, das die Weiterleitung des Datenverkehrs über mehrere Netzwerke ermöglicht.

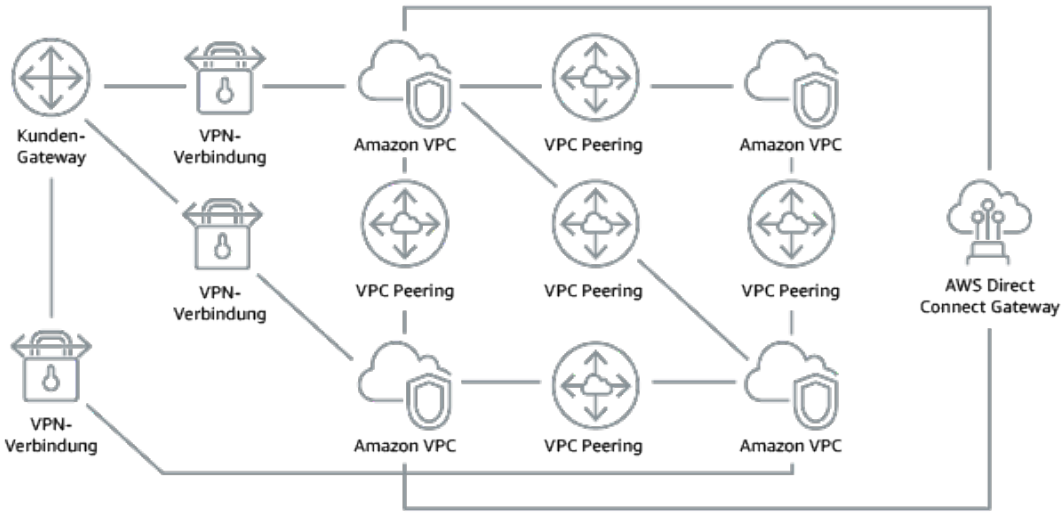


Abbildung 1: Ohne AWS Transit Gateway: Sie müssen jede Amazon VPC über eine VPN-Verbindung miteinander und mit jedem Standort verbinden. Bei der Skalierung kann dies sehr komplex werden.

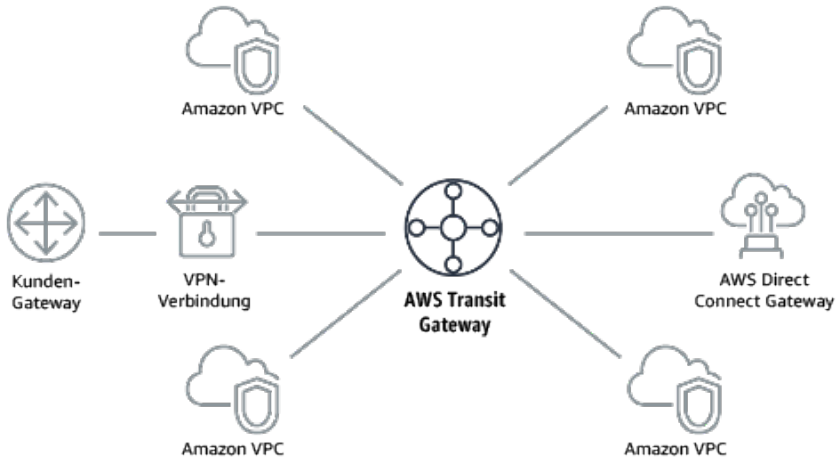


Abbildung 2: Mit AWS Transit Gateway: Sie verbinden einfach jede Amazon VPC oder jedes VPN mit dem AWS Transit Gateway und leiten den Datenverkehr zu und von jeder VPC oder VPN weiter.

Gängige Antimuster:

- Verbinden von mehr als zwei VPCs mit VPC-Peering.
- Es werden mehrere BGP-Sitzungen für jede VPC eingerichtet, um Konnektivität für mehrere Virtual Private Clouds (VPCs) in mehreren AWS-Regionen herzustellen.

Vorteile der Einführung dieser bewährten Methode: Mit der zunehmenden Anzahl der Netzwerke wird die Komplexität solcher verflochtenen Verbindungen immer größer. AWS Transit Gateway bietet ein einfach zu wartendes Nabe-und-Speiche-Modell, das die Weiterleitung des Datenverkehrs über mehrere Netzwerke ermöglicht.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Ziehen Sie Nabe-und-Speiche-Topologien gegenüber M-zu-N-Netzen vor. Wenn mehr als zwei Netzwerkadressbereiche (VPCs, On-Premises-Netzwerke) über VPC-Peering, AWS Direct Connect oder VPN verbunden sind, verwenden Sie ein Nabe-und-Speiche-Modell, wie es von AWS Transit Gateway bereitgestellt wird.

- Bei nur zwei derartigen Netzwerken können Sie sie einfach miteinander verbinden, doch mit der zunehmenden Anzahl der Netzwerke wird die Komplexität solcher verflochtenen Verbindungen immer größer. AWS Transit Gateway bietet ein einfach zu wartendes Nabe-und-Speiche-Modell, das die Weiterleitung des Datenverkehrs über mehrere Netzwerke ermöglicht.
  - Was ist ein Transit-Gateway?

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Planung Ihres Netzwerks unterstützen können
- AWS Marketplace für Netzwerkinfrastruktur
- Hochverfügbare Netzwerkkonnektivität zwischen mehreren Rechenzentren
- VPC-Endpunkte und VPC-Endpunktservices (AWS PrivateLink)
- Was ist Amazon VPC?
- Was ist ein Transit-Gateway?

Relevante Videos:

- AWS re:Invent 2018: Erweitertes VPC-Design und neue Funktionen für Amazon VPC (NET303)
- AWS re:Invent 2019: AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs (NET406-R1)

REL02-BP05 Erzwingen von sich nicht überschneidenden privaten IP-Adressbereichen in allen privaten Adressbereichen, in denen eine Verbindung besteht

Die IP-Adressbereiche Ihrer VPCs dürfen sich nicht überschneiden, wenn sie per Peering oder über VPN verbunden sind. Ebenso müssen Sie IP-Adresskonflikte zwischen einer VPC und lokalen Umgebungen oder anderen verwendeten Cloud-Anbietern vermeiden. Sie müssen bei Bedarf auch die Möglichkeit haben, private IP-Adressbereiche zuzuweisen.

Ein IP-Adressenverwaltungssystem (IPAM) kann dabei helfen. Im AWS Marketplace stehen mehrere IPAMs zur Verfügung.

Gängige Antimuster:

- Verwenden Sie denselben IP-Bereich in Ihrer VPC wie im lokalen Netzwerk oder in Ihrem Unternehmensnetzwerk.
- Keine Verfolgung von IP-Bereichen von VPCs, die zur Bereitstellung der Workloads verwendet werden.

Vorteile der Einführung dieser bewährten Methode: Mit der aktiven Planung des Netzwerks stellten Sie sicher, dass dieselbe IP-Adresse in miteinander verbunden Netzwerken nicht mehrmals vorkommt. So wird verhindert, dass Routing-Probleme in Teilen der Workload auftreten, die die verschiedenen Anwendungen verwenden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Überwachen und verwalten Sie die CIDR-Nutzung. Bewerten Sie die potenzielle Nutzung in AWS, fügen Sie vorhandenen VPCs CIDR-Bereiche hinzu und erstellen Sie neue VPCs, um das geplante Wachstum abzudecken.
  - Ermitteln Sie den aktuellen CIDR-Umfang (z. B. VPCs, Subnetze).
    - Erfassen Sie über die Service-API den aktuellen CIDR-Umfang.
  - Erfassen Sie die aktuelle Subnetzauslastung.
    - Ermitteln Sie über die Service-API die in jeder Region pro VPC vorhandenen Subnetze.
      - DescribeSubnets
    - Zeichnen Sie die aktuelle Auslastung auf.
    - Prüfen Sie, ob sich IP-Bereiche überschneiden.
    - Berechnen Sie die freie Kapazität.
    - Identifizieren Sie sich überschneidende IP-Bereiche. Sie können wahlweise zu einem neuen Adressbereich migrieren oder NAT-Appliances (Network and Port Translation) aus AWS Marketplace verwenden, wenn Sie die sich überschneidenden Bereiche verbinden müssen.

Ressourcen

Ähnliche Dokumente:

- APN-Partner: Partner, die Sie bei der Planung Ihres Netzwerks unterstützen können
- AWS Marketplace für Netzwerkinfrastruktur

- Amazon Virtual Private Cloud-Konnektivitätsoptionen – Whitepaper
- Hochverfügbare Netzwerkkonnektivität zwischen mehreren Rechenzentren
- Was ist Amazon VPC?
- Was ist IPAM?

Ähnliche Videos:

- AWS re:Invent 2018: Erweitertes VPC-Design und neue Funktionen für Amazon VPC (NET303)
- AWS re:Invent 2019: AWS Transit Gateway-Referenzarchitekturen für verschiedene VPCs (NET406-R1)

Workload-Architektur

Fragen

- ZUV 3 Wie entwerfen Sie Ihre Workload-Service-Architektur?
- ZUV 4 Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle vermieden werden?
- ZUV 5 Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle abgemildert oder bewältigt werden?

ZUV 3 Wie entwerfen Sie Ihre Workload-Service-Architektur?

Erstellen Sie hoch skalierbare und zuverlässige Workloads mithilfe einer serviceorientierten Architektur (SOA) oder einer Microservices-Architektur. Eine serviceorientierte Architektur (SOA) hat zum Ziel, Softwarekomponenten über Service-Schnittstellen wiederverwendbar zu machen. Die Microservices-Architektur geht noch weiter, um Komponenten kleiner und einfacher zu machen.

Bewährte Methoden

- REL03-BP01 Segmentierung Ihres Workloads
- REL03-BP02 Entwickeln von Services, die sich auf bestimmte Geschäftsbereiche und Funktionen konzentrieren
- REL03-BP03 Bereitstellen von Serviceverträgen pro API

REL03-BP01 Segmentierung Ihres Workloads

Die Workload-Segmentierung ist wichtig, wenn es um die Festlegung der Resilienzanforderungen Ihrer Anwendung geht. Eine monolithische Architektur sollte vermieden werden, wann immer möglich. Stattdessen sollten Sie sorgfältig überlegen, welche Anwendungskomponenten in Microservices aufgeteilt werden können. Abhängig von den Anforderungen Ihrer Anwendung könnte es sich im Endergebnis um eine Kombination aus einer serviceorientierten Architektur (SOA) und Microservices handeln, wenn dies möglich ist. Workloads, die zustandslos sein können, können eher als Microservices bereitgestellt werden.

Gewünschtes Ergebnis: Workloads sollten unterstützbar, skalierbar und so lose miteinander verbunden sein wie möglich.

Wiegen Sie bei Entscheidungen zur Segmentierung von Workloads die Vorteile und die Komplexitäten miteinander ab. Was für ein neues Produkt richtig ist, das gerade auf dem Markt eingeführt wird, unterscheidet sich von den Anforderungen eines Workloads, der von Anfang an skalierbar sein muss. Bei einem Faktorwechsel für einen vorhandenen Monolith müssen Sie berücksichtigen, wie gut dieser aufgeteilt und in zustandslose Anwendungen transformiert werden kann. Die Aufteilung von Services in kleinere Teile ermöglicht kleinen, klar definierten Teams, diese weiterzuentwickeln und zu verwalten. Kleinere Services können jedoch Komplexitäten wie eine möglicherweise erhöhte Latenz, ein komplexeres Debugging und einen erhöhten operativen Aufwand einführen.

Typische Anti-Muster:

- Der Microservice Death Star ist eine Situation, in der die einzelnen Komponenten so stark voneinander abhängig werden, dass der Ausfall einer einzigen Komponente einen wesentlich größeren Ausfall bewirkt. Das bedeutet, dass die Komponenten so starr und anfällig wie ein Monolith sind.

Vorteile der Einrichtung dieser Best Practice:

- Spezifischere Segmente führen zu einer größeren Agilität, zu organisatorischer Flexibilität und zu Skalierbarkeit.
- Die Auswirkungen von Service-Unterbrechungen werden reduziert.
- Die einzelnen Komponenten einer Anwendung besitzen möglicherweise unterschiedliche Anforderungen an die Verfügbarkeit, die von einer stärkeren Segmentierung besser unterstützt werden können.

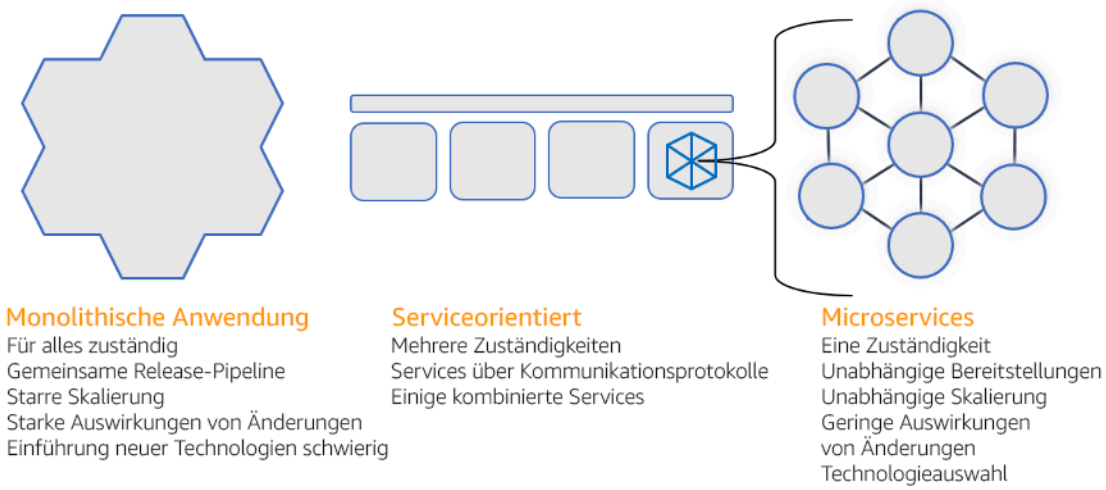
- Die Verantwortlichkeiten der Teams, die den Workload unterstützen, sind klar definiert.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Wählen Sie Ihren Architekturtyp basierend auf der Segmentierung Ihres Workloads aus. Wählen Sie eine serviceorientierte Architektur (SOA) oder eine Microservices-Architektur aus. (In seltenen Fällen ist möglicherweise auch eine monolithische Architektur geeignet.) Auch wenn Sie mit einer monolithischen Architektur beginnen möchten, müssen Sie sicherstellen, dass diese modular ist und zu einer SOA oder zu Microservices weiterentwickeln werden kann, wenn Ihr Produkt aufgrund der zunehmenden Einführung durch Benutzer skaliert wird. SOA und Microservices ermöglichen eine kleinteiligere Segmentierung, die als moderne skalierbare und zuverlässige Architektur bevorzugt wird. Es gibt jedoch auch Nachteile, die besonders bei der Bereitstellung einer Microservice-Architektur berücksichtigt werden sollten.

Aufgrund ihrer verteilten Computing-Architektur kann es schwieriger sein, die Latenzanforderungen von Benutzern zu erfüllen. Außerdem sind das Debugging und die Nachverfolgung von Benutzerinteraktionen komplexer. Zur Lösung dieses Problems können Sie AWS X-Ray verwenden. Ein weiterer Effekt ist die erhöhte operative Komplexität, da die Anzahl der von Ihnen verwalteten Anwendungen zunimmt. In der Folge müssen Sie eine größere Zahl voneinander unabhängiger Komponenten bereitstellen.



Monolithische, serviceorientierte und Microservice-Architekturen

Implementierungsschritte

- Ermitteln Sie die richtige Architektur für den Faktorwechsel oder die Entwicklung Ihrer Anwendung. SOA und Microservices bieten eine jeweils kleinere Segmentierung, die als moderne skalierbare und zuverlässige Architektur bevorzugt wird. SOA kann ein guter Kompromiss für das Erreichen einer kleineren Segmentierung sein, während die Komplexität von Microservices zum Teil vermieden wird. Weitere Informationen finden Sie in Kompromisse bei Microservices.
- Wenn Ihre Workload für sie zugänglich ist und Ihre Organisation sie unterstützen kann, sollten Sie eine Microservices-Architektur verwenden, um die beste Agilität und Zuverlässigkeit zu erzielen. Weitere Informationen finden Sie in Implementieren von Microservices in AWS.
- Sie sollten das Muster mit der Bezeichnung Strangler Fig („Würgefeige“) verwenden, um einen Faktorwechsel für einen Monolithen durchzuführen, bei dem Sie diesen in kleinere Komponenten aufteilen. Dies umfasst die schrittweise Ersetzung spezifischer Anwendungskomponenten durch neue Anwendungen und Services. AWS Migration Hub Refactor Spaces dient als Ausgangspunkt für den inkrementellen Faktorwechsel. Weitere Informationen finden Sie in Nahtlose Integration ältere On-Premises-Workloads unter Anwendung eines Strangler-Fig-Musters.
- Die Implementierung von Microservices erfordert möglicherweise einen Mechanismus für die Entdeckung von Services, damit diese verteilten Services miteinander kommunizieren können. AWS App Mesh kann mit serviceorientierten Architekturen verwendet werden, um eine zuverlässige Erkennung von Services und den Zugriff auf sie zu unterstützen. AWS Cloud Map kann für die dynamische, DNS-basierte Serviceerkennung verwendet werden.
- Wenn Sie von einem Monolithen zur SOA migrieren, kann Amazon MQ helfen, als Service-Bus die Lücke zu überbrücken, wenn Sie ältere Anwendungen in der Cloud neu entwerfen.
- Im Fall vorhandener Monolithen mit einer einzigen, geteilten Datenbank müssen Sie entscheiden, wie Sie die Daten neu in kleineren Segmenten organisieren. Dabei kann es sich um Geschäftsbereiche, Zugriffsmuster oder Datenstrukturen handeln. An diesem Punkt des Faktorwechsel-Prozesses sollten Sie entscheiden, ob Sie eine relationale oder eine nicht relationale (NoSQL) Datenbank verwenden. Weitere Informationen finden Sie in Von SQL zu NoSQL.

Aufwand für den Implementierungsplan: Hoch

Ressourcen

Zugehörige bewährte Methoden:

- REL03-BP02 Entwickeln von Services, die sich auf bestimmte Geschäftsbereiche und Funktionen konzentrieren

Zugehörige Dokumente:

- Amazon API Gateway: Konfigurieren einer REST-API mit OpenAPI
- Was ist eine serviceorientierte Architektur?
- Bounded Context (Begrenzter Kontext) (ein zentrales Muster im domänengesteuerten Design)
- Implementieren von Microservices in AWS
- Kompromisse bei Microservices
- Microservices – eine Definition dieses neuen Architekturbegriffs
- Microservices in AWS
- Was ist AWS App Mesh?

Zugehörige Beispiele:

- Workshop für die iterative App-Modernisierung

Zugehörige Videos:

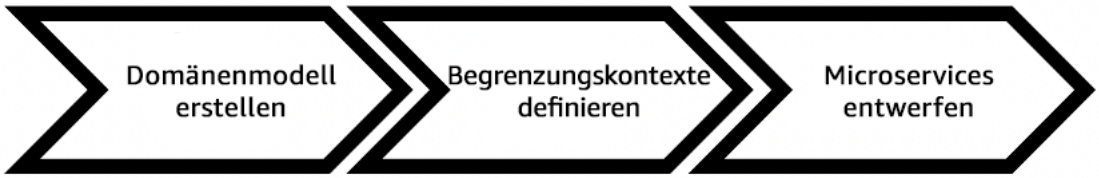
- Kompetenz mit Microservices in AWS

REL03-BP02 Entwickeln von Services, die sich auf bestimmte Geschäftsbereiche und Funktionen konzentrieren

Eine serviceorientierte Architektur (SOA) entwickelt Services mit genau abgegrenzten Funktionen, die von Geschäftsanforderungen definiert werden. Microservices verwenden Domänenmodelle und begrenzten Kontext, um dies weiter einzuschränken, sodass jeder Service nur eine Aufgabe erledigt. Wenn Sie sich auf bestimmte Funktionen konzentrieren, können Sie die Zuverlässigkeitsanforderungen verschiedener Services differenzieren und Investitionen genauer ausrichten. Ein präzises Geschäftsproblem und ein kleines Team, das mit jedem Service verbunden ist, ermöglichen auch eine einfachere Skalierung der Organisation.

Beim Entwerfen einer Microservice-Architektur ist es hilfreich, das Domain-Driven Design (DDD) zu verwenden, um das Geschäftsproblem mithilfe von Entitäten zu modellieren. Für die Website Amazon.com können Entitäten beispielsweise Pakete, Zustellung, Zeitplan, Preise, Rabatte und

Währung enthalten. Anschließend wird das Modell mit begrenztem Kontextweiter in kleinere Modelle unterteilt, wobei Entities mit ähnlichen Funktionen und Attributen in Gruppen sortiert werden. Beim Beispiel des Amazon.com-Pakets wären Lieferung und Zeitplan Teil des Versandkontexts, während Preise, Rabatte und Währung Teil des Preiskontexts sind. Wenn das Modell in Kontexte unterteilt ist, entsteht eine Vorlage für die Begrenzung von Microservices.



Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Entwerfen Sie Ihre Workload basierend auf Ihren Unternehmensdomänen und deren jeweiliger Funktionalität. Wenn Sie sich auf bestimmte Funktionen konzentrieren, können Sie die Zuverlässigkeitsanforderungen verschiedener Services differenzieren und Investitionen genauer ausrichten. Ein präzises Geschäftsproblem und ein kleines Team, das mit jedem Service verbunden ist, ermöglichen auch eine einfachere Skalierung der Organisation.
  - Führen Sie eine Domänenanalyse durch, um Ihrer Workload ein domänengesteuertes Design (DDD) zuzuordnen. Anschließend können Sie einen Architekturtyp auswählen, der die Anforderungen Ihrer Workload erfüllt.
    - How to break a Monolith into Microservices (Aufschlüsseln eines Monolithen in Microservices)
    - Getting Started with DDD when Surrounded by Legacy Systems (Erste Schritte mit DDD, wenn die Umgebung aus Legacy-Systemen besteht)
    - „Domain-Driven Design: Tackling Complexity in the Heart of Software“ von Eric Evans („Domänengesteuertes Design: Bewältigung der Komplexität im Herzen der Software“)
    - Implementieren von Microservices in AWS
- Zerlegen Sie Ihre Services in kleinstmögliche funktionale Einheiten. Mit der Microservices-Architektur können Sie Ihre Workload in Komponenten mit minimaler Funktionalität aufteilen, um Skalierung und Agilität der Organisation zu ermöglichen.
  - Definieren Sie die API für die Workload und ihre Designziele, Limits und sonstigen Nutzungsanforderungen.
    - Definieren Sie die API.
      - Lassen Sie in der API-Definition Raum für Wachstum und weitere Parameter.

- Definieren Sie die gewünschten Verfügbarkeiten.
  - Sie können für Ihre API mehrere Designziele für unterschiedliche Funktionen haben.
- Limits festlegen
  - Definieren Sie mithilfe von Tests die Limits Ihrer Workload-Funktionen.

Ressourcen

Relevante Dokumente:

- Amazon API Gateway: Konfigurieren einer REST-API mit OpenAPI
- Bounded Context (Begrenzter Kontext) (ein zentrales Muster im domänengesteuerten Design)
- „Domain-Driven Design: Tackling Complexity in the Heart of Software“ von Eric Evans („Domänengesteuertes Design: Bewältigung der Komplexität im Herzen der Software“)
- Getting Started with DDD when Surrounded by Legacy Systems (Erste Schritte mit DDD, wenn die Umgebung aus Legacy-Systemen besteht)
- How to break a Monolith into Microservices (Aufschlüsseln eines Monolithen in Microservices)
- Implementieren von Microservices in AWS
- Kompromisse bei Microservices
- Microservices – eine Definition dieses neuen Architekturbegriffs
- Microservices in AWS

REL03-BP03 Bereitstellen von Serviceverträgen pro API

Serviceverträge sind dokumentierte Vereinbarungen zur Service-Integration zwischen Teams und enthalten eine maschinenlesbare API-Definition, Ratenlimits und Leistungserwartungen. Eine Versionsverwaltungs-Strategie ermöglicht es Ihren Clients, die vorhandene API weiter zu verwenden und ihre Anwendungen auf die neuere API zu migrieren, wenn sie bereit sind. Die Bereitstellung kann jederzeit erfolgen, solange der Vertrag nicht verletzt wird. Das Serviceanbieterteam kann den Technologie-Stack seiner Wahl verwenden, um den API-Vertrag zu erfüllen. Ebenso kann der Service-Verbraucher seine eigene Technologie verwenden.

Microservices nutzen das Konzept einer serviceorientierten Architektur (SOA) und erstellen Services mit minimalem Funktionsumfang. Jeder Service veröffentlicht eine API und Designziele, Limits und andere Überlegungen zur Nutzung des Services. Damit entsteht ein Vertrag mit aufrufenden Anwendungen. Dies bietet die folgenden drei Vorteile:

- Der Service muss eine Lösung für ein konkretes Geschäftsproblem bieten und verfügt über ein kleines Team, das Eigentümer des Geschäftsproblems ist. Dieser Ansatz ermöglicht eine bessere unternehmerische Skalierung.
- Das Team kann jederzeit Bereitstellungen durchführen, solange es seine API- und weitere vertragsbasierte Anforderungen erfüllt.
- Das Team kann einen beliebigen Technologie-Stack verwenden, solange es seine API- und weitere vertragsbasierte Anforderungen erfüllt.

Amazon API Gateway ist ein vollständig verwalteter Service, der es Entwicklern erleichtert, APIs in jeder Größenordnung zu erstellen, zu veröffentlichen, zu warten, zu überwachen und zu sichern. API Gateway übernimmt alle Aufgaben, die mit der Annahme und Verarbeitung von mitunter Hunderttausenden gleichzeitigen API-Aufrufen verbunden sind. Hierzu zählen die Verwaltung des Datenverkehrs, die Autorisierung, die Zugriffskontrolle, die Überwachung sowie die API-Versionsverwaltung. Mit der OpenAPI-Spezifikation (OAS), früher als Swagger-Spezifikation bezeichnet, können Sie Ihren API-Vertrag definieren und in API Gateway importieren. Mit API Gateway können Sie die APIs versionieren und bereitstellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Stellen Sie Serviceverträge pro API bereit. Serviceverträge sind dokumentierte Vereinbarungen zur Service-Integration zwischen Teams und enthalten eine maschinenlesbare API-Definition, Ratenlimits und Leistungserwartungen.
  - Amazon API Gateway: Konfigurieren einer REST-API mit OpenAPI
    - Eine Versioning-Strategie ermöglicht es Clients, die vorhandene API weiter zu verwenden und ihre Anwendungen auf die neuere API zu migrieren, wenn sie bereit sind.
    - Amazon API Gateway ist ein vollständig verwalteter Service, der es Entwicklern leicht macht, APIs beliebiger Größe zu erstellen. Mit der OpenAPI-Spezifikation (OAS), früher als Swagger-Spezifikation bezeichnet, können Sie Ihren API-Vertrag definieren und in API Gateway importieren. Mit API Gateway können Sie die APIs versionieren und bereitstellen.

Ressourcen

Relevante Dokumente:

- Amazon API Gateway: Konfigurieren einer REST-API mit OpenAPI

- Bounded Context (Begrenzter Kontext) (ein zentrales Muster im domänengesteuerten Design)
- Implementieren von Microservices in AWS
- Kompromisse bei Microservices
- Microservices – eine Definition dieses neuen Architekturbegriffs
- Microservices in AWS

ZUV 4 Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle vermieden werden?

Verteilte Systeme nutzen Kommunikationsnetzwerke, um Komponenten wie Server oder Services miteinander zu verbinden. Ihre Workload muss trotz Datenverlust oder höherer Latenz in diesen Netzwerken zuverlässig ausgeführt werden. Komponenten des verteilten Systems müssen so funktionieren, dass sie keine negativen Auswirkungen auf andere Komponenten oder die Workload haben. Diese bewährten Methoden verhindern Ausfälle und verbessern die mittlere Zeit zwischen Ausfällen (MTBF).

Bewährte Methoden

- REL04-BP01 Bestimmen, welches verteilte System erforderlich ist
- REL04-BP02 Implementieren lose gekoppelter Abhängigkeiten
- REL04-BP03 Konstante Ausführung
- REL04-BP04 Festlegen aller Reaktionen als idempotent

REL04-BP01 Bestimmen, welches verteilte System erforderlich ist

Harte verteilte Echtzeitsysteme erfordern synchrone und schnelle Antworten, während bei weichen Echtzeitsystemen ein großzügigeres Zeitfenster von Minuten (oder mehr) für Antworten besteht. Offline-Systeme verarbeiten Antworten über Stapelverarbeitung oder asynchrone Verarbeitung. Harte verteilte Echtzeitsysteme haben die strengsten Zuverlässigkeitsanforderungen.

Die schwierigsten Herausforderungen mit verteilten Systemen gelten für die harten verteilten Echtzeitsysteme, die auch als Anfrage-/Antwortservices bezeichnet werden. Die Schwierigkeiten entstehen dadurch, dass Anfragen unvorhersehbar eingehen und schnelle Antworten ausgegeben werden müssen (z. B. weil der Kunde aktiv auf die Antwort wartet). Beispiele sind Frontend-Webserver, die Auftragspipeline, Kreditkartentransaktionen, jede AWS-API und Telefonie.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Bestimmen Sie, welches verteilte System erforderlich ist. Zu den Herausforderungen verteilter Systeme gehörten die Latenz, die Skalierung, das Verständnis von Netzwerk-APIs, das Marshalling und Unmarshalling von Daten sowie die Komplexität von Algorithmen wie Paxos. Angesichts des zunehmenden Wachstums und Verteilungsgrads von Systemen werden theoretische Edge-Fälle zu regelmäßigen Ereignissen.
  - Die Amazon Builders' Library: Herausforderungen bei verteilten Systemen
    - In Echtzeit verteilte Systeme erfordern synchrone und schnelle Antworten.
    - Bei weichen Echtzeitsystemen besteht ein großzügigeres Zeitfenster von Minuten (oder mehr) für Antworten.
    - Offline-Systeme verarbeiten Antworten über Stapelverarbeitung oder asynchrone Verarbeitung.
    - Harte verteilte Echtzeitsysteme haben die strengsten Zuverlässigkeitsanforderungen.

Ressourcen

Relevante Dokumente:

- Amazon EC2: Idempotenz sicherstellen
- Die Amazon Builders' Library: Herausforderungen bei verteilten Systemen
- Die Amazon Builders' Library: Zuverlässigkeit, stetige Ausführung und eine gute Tasse Kaffee
- Was ist Amazon EventBridge?
- Was ist Amazon Simple Queue Service?

Relevante Videos:

- AWS New York Summit 2019: Einführung in ereignisgesteuerte Architekturen und Amazon EventBridge (MAD205)
- AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über Systeme übernimmt – große und kleine ARC337 (umfasst lose Verkoppelung, konstante Ausführung, statische Stabilität)
- AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)

REL04-BP02 Implementieren lose gekoppelter Abhängigkeiten

Abhängigkeiten etwa zwischen Warteschlangensystemen, Streaming-Systemen, Workflows und Load Balancern sind lose gekoppelt. Eine lose Verkoppelung hilft, das Verhalten einer Komponente von anderen Komponenten zu isolieren, die von ihr abhängig sind. Dies verbessert Resilienz und Agilität.

Wenn Änderungen an einer Komponente bewirken, dass andere abhängige Komponenten ebenfalls geändert werden, sind sie eng gekoppelt. Die lose Kopplung unterbricht diese Abhängigkeit, sodass abhängige Komponenten nur die versionierte und veröffentlichte Schnittstelle kennen müssen. Die Implementierung einer losen Kopplung zwischen Abhängigkeiten isoliert einen Ausfall. So wird verhindert, dass er sich auf andere Komponenten auswirkt.

Die lose Kopplung ermöglicht Ihnen, einer Komponente zusätzlichen Code oder Funktionen hinzuzufügen und gleichzeitig das Risiko für Komponenten zu minimieren, die von ihr abhängig sind. Außerdem wird die Skalierbarkeit verbessert, da Sie die zugrunde liegende Implementierung der Abhängigkeit aufskalieren oder sogar ändern können.

Um die Ausfallsicherheit durch lose Kopplung weiter zu verbessern, legen Sie Komponenten-Interaktionen nach Möglichkeit als asynchron fest. Dieses Modell eignet sich für jede Interaktion, bei der keine sofortige Antwort benötigt wird, sondern die Bestätigung ausreicht, dass eine Anfrage registriert wurde. Es umfasst eine Komponente, die Ereignisse generiert, und eine andere Komponente, die sie konsumiert. Die beiden Komponenten lassen sich nicht durch direkte Punkt-zu-Punkt-Interaktion integrieren, sondern in der Regel über eine temporäre, robuste Speicherschicht, z. B. eine SQS-Warteschlange oder eine Streaming-Datenplattform wie Amazon Kinesis oder AWS Step Functions.

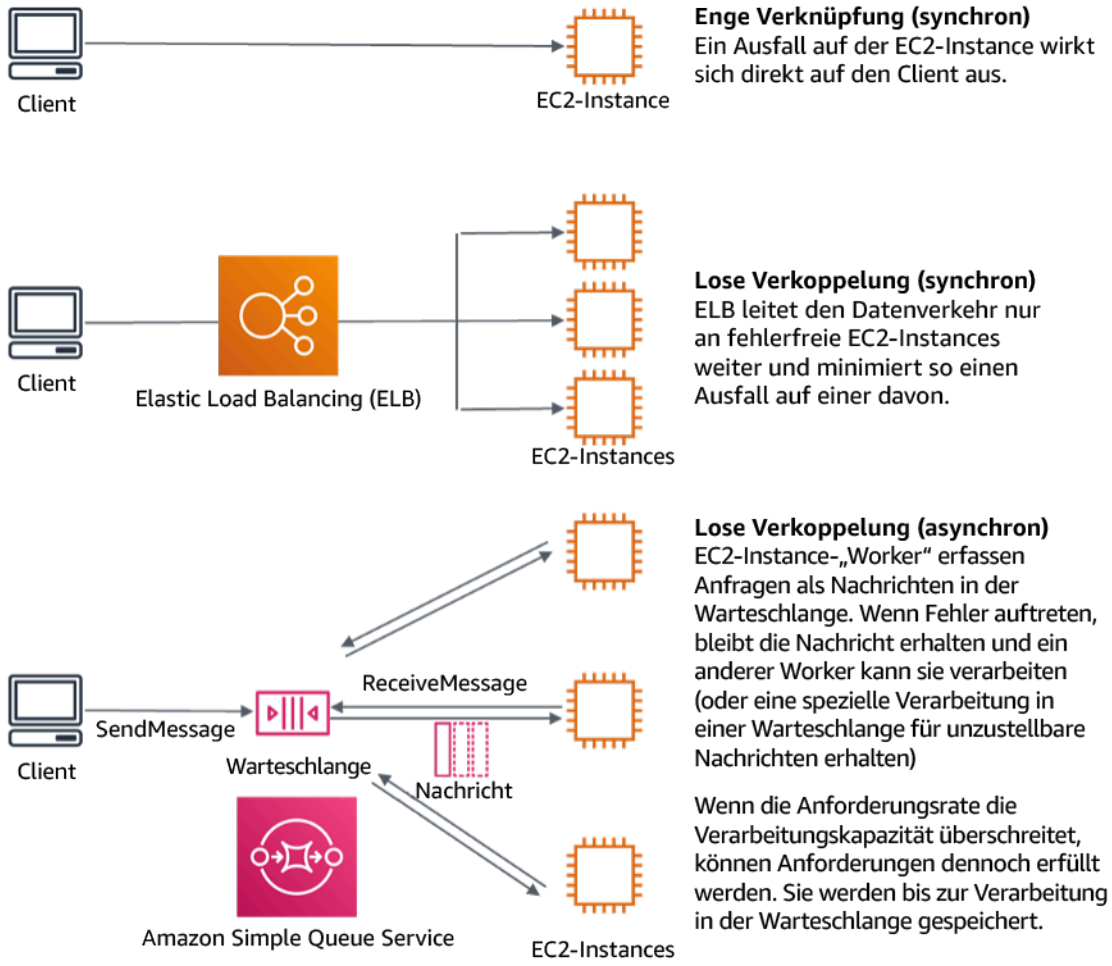


Abbildung 4: Abhängigkeiten etwa zwischen Warteschlangensystemen und Load Balancer sind lose gekoppelt

Amazon SQS-Warteschlangen und Elastic Load Balancers sind nur zwei Möglichkeiten, um eine Zwischenschicht für lose Kopplung hinzuzufügen. Ereignisgesteuerte Architekturen können auch in der AWS Cloud mithilfe von Amazon EventBridge erstellt werden, was Clients (Ereignisproduzenten) von den Services abstrahieren kann, auf die sie sich verlassen (Ereignisverbraucher). Amazon Simple Notification Service (Amazon SNS) ist eine effektive Lösung, wenn Sie Push-basiertes M-zu-N-Messaging mit hohem Durchsatz benötigen. Mithilfe von Amazon SNS-Themen können Ihre Publisher-Systeme Nachrichten zur parallelen Verarbeitung an eine große Anzahl von Abonnenten-Endpunkten senden.

Warteschlangen bieten zwar mehrere Vorteile, doch Anfragen, die älter als ein Schwellenwert sind (oft Sekunden), sollten in den meisten harten Echtzeitsystemen als veraltet betrachtet (der Client hat aufgegeben und wartet nicht mehr auf eine Antwort) und nicht verarbeitet werden. Auf diese Weise können stattdessen neuere (und wahrscheinlich noch gültige Anfragen) verarbeitet werden.

Gängige Antimuster:

- Bereitstellen eines Singletons im Rahmen einer Workload.
- APIs werden zwischen Workload-Ebenen direkt aufgerufen, ohne Möglichkeit eines Failovers oder einer asynchronen Verarbeitung der Anfrage.

Vorteile der Einführung dieser bewährten Methode: Eine lose Verkoppelung hilft, das Verhalten einer Komponente von anderen Komponenten zu isolieren, die von ihr abhängig sind. Dies verbessert Resilienz und Agilität. Fehler in einer Komponente sind von anderen isoliert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren Sie lose gekoppelte Abhängigkeiten. Abhängigkeiten etwa zwischen Warteschlangensystemen, Streaming-Systemen, Workflows und Load Balancern sind lose gekoppelt. Eine lose Verkoppelung hilft, das Verhalten einer Komponente von anderen Komponenten zu isolieren, die von ihr abhängig sind. Dies verbessert Resilienz und Agilität.
  - AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)
  - Was ist Amazon EventBridge?
  - Was ist Amazon Simple Queue Service?
    - Mit Amazon EventBridge können Sie ereignisgesteuerte Architekturen entwickeln, die lose verkoppelt und verteilt sind.
      - AWS New York Summit 2019: Einführung in ereignisgesteuerte Architekturen und Amazon EventBridge (MAD205)
    - Wenn Änderungen für eine Komponente Änderungen für andere Komponenten auslöst, die von ihr abhängig sind, sind sie eng verkoppelt. Die lose Kopplung hebt diese Abhängigkeit auf, sodass abhängige Komponenten nur die versionierte und veröffentlichte Schnittstelle kennen müssen.
    - Gestalten Sie die Interaktionen zwischen Komponenten möglichst als asynchrone Interaktionen. Dieses Modell ist für Interaktionen geeignet, die keine sofortigen Reaktionen erfordern und für die die Bestätigung der Registrierung einer Anfrage ausreichend ist.
      - AWS re:Invent 2019: Scalable serverless event-driven applications using Amazon SQS and Lambda (API304) (Skalierbare serverlose ereignisgesteuerte Anwendungen, die Amazon SQS und Lambda nutzen)

Ressourcen

Relevante Dokumente:

- AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)
- Amazon EC2: Idempotenz sicherstellen
- Die Amazon Builders' Library: Herausforderungen für verteilte Systeme
- Die Amazon Builders' Library: Zuverlässigkeit, stetige Ausführung und eine gute Tasse Kaffee
- Was ist Amazon EventBridge?
- Was ist Amazon Simple Queue Service?

Relevante Videos:

- AWS New York Summit 2019: Einführung in ereignisgesteuerte Architekturen und Amazon EventBridge (MAD205)
- AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über Systeme übernimmt – große und kleine ARC337 (umfasst lose Verkoppelung, konstante Ausführung, statische Stabilität)
- AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)
- AWS re:Invent 2019: Scalable serverless event-driven applications using Amazon SQS and Lambda (API304) (Skalierbare serverlose ereignisgesteuerte Anwendungen, die Amazon SQS und Lambda nutzen)

REL04-BP03 Konstante Ausführung

Bei größeren, schnellen Lastveränderungen können Systeme ausfallen. Wenn Ihre Workload beispielsweise eine Zustandsprüfung ausführt, die den Zustand vieler tausend Server überwacht, sollte sie jedes Mal die gleiche Nutzlast senden (einen vollständigen Snapshot des aktuellen Status). Unabhängig davon, ob keine Server oder alle Server ausfallen, führt das System für die Zustandsprüfung die Aufgaben stetig und ohne große, schnelle Änderungen aus.

Wenn das Zustandsprüfungssystem beispielsweise 100 000 Server überwacht, ist die Last darauf angesichts der normalerweise geringen Serverausfallrate nominal. Wenn jedoch ein großes

Ereignis die Hälfte dieser Server fehlerhaft macht, wäre das Zustandsprüfungssystem überfordert, wenn es versucht, Benachrichtigungssysteme zu aktualisieren und den Status an seine Clients zu kommunizieren. Stattdessen sollte das Zustandsprüfungssystem jedes Mal den vollständigen Snapshot des aktuellen Status senden. 100 000 Server-Zustände, die jeweils durch ein Bit dargestellt werden, entsprächen nur eine Nutzlast von 12,5 KB. Unabhängig davon, ob keine oder alle Server ausfallen – das System für die Zustandsprüfung erledigt seine Arbeit konstant und große, schnelle Änderungen stellen keine Bedrohung für die Systemstabilität dar. Auf diese Weise führt Amazon Route 53 Zustandsprüfungen für Endpunkte (wie z. B. IP-Adressen) durch, um zu ermitteln, wie Endbenutzer an diese weitergeleitet werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Führen Sie Aufgaben konstant aus, sodass auch bei großen, schnellen Lastveränderungen keine Fehler auf Systemen auftreten.
- Implementieren Sie lose gekoppelte Abhängigkeiten. Abhängigkeiten etwa zwischen Warteschlangensystemen, Streaming-Systemen, Workflows und Load Balancern sind lose gekoppelt. Eine lose Verkoppelung hilft, das Verhalten einer Komponente von anderen Komponenten zu isolieren, die von ihr abhängig sind. Dies verbessert Resilienz und Agilität.
  - Die Amazon Builders' Library: Zuverlässigkeit, stetige Ausführung und eine gute Tasse Kaffee
  - AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über große und kleine Systeme übernimmt ARC337 (umfasst konstante Ausführung)
    - Beispiel: Zustandsprüfungssystem, das 100.000 Server überwacht: Entwickeln Sie die Workloads so, dass die Nutzlastgrößen unabhängig von der Anzahl der Erfolge oder Ausfälle konstant bleiben.

Ressourcen

Ähnliche Dokumente:

- Amazon EC2: Idempotenz sicherstellen
- Die Amazon Builders' Library: Herausforderungen für verteilte Systeme
- Die Amazon Builders' Library: Zuverlässigkeit, stetige Ausführung und eine gute Tasse Kaffee

Ähnliche Videos:

- AWS New York Summit 2019: Einführung in ereignisgesteuerte Architekturen und Amazon EventBridge (MAD205)
- AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über große und kleine Systeme übernimmt ARC337 (umfasst konstante Ausführung)
- AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über Systeme übernimmt – große und kleine ARC337 (umfasst lose Verkoppelung, konstante Ausführung, statische Stabilität)
- AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)

REL04-BP04 Festlegen aller Reaktionen als idempotent

Ein idempotenter Service garantiert, dass jede Anfrage genau einmal abgeschlossen wird. Das bedeutet, dass das Senden mehrerer identischer Anfragen den gleichen Effekt hat wie das Senden einer einzelnen Anfrage. Ein idempotenter Service erleichtert es einem Client, Wiederholungen zu implementieren. So muss nicht befürchtet werden, dass eine Anfrage fälschlicherweise mehrfach verarbeitet wird. Zu diesem Zweck können Clients API-Anfragen mit einem Idempotenz-Token ausgeben. Das gleiche Token wird verwendet, wenn die Anfrage wiederholt wird. Eine idempotente Service-API gibt mithilfe des Tokens eine Antwort zurück, die identisch mit der Antwort ist, die beim ersten Abschluss der Anfrage zurückgegeben wurde.

In einem verteilten System ist es einfach, eine Aktion höchstens einmal (der Client stellt nur eine Anforderung) oder mindestens einmal (Anforderung so lange, bis der Client erfolgreich ist) durchzuführen. Es ist jedoch schwer zu gewährleisten, dass eine Aktion idempotent ist, was bedeutet, dass sie genau einmal ausgeführt wird, sodass das Erstellen mehrerer identischer Anfragen den gleichen Effekt hat wie das Erstellen einer einzelnen Anfrage. Durch die Verwendung von idempotenten Tokens in APIs können Services einmal oder mehrmals eine sich verändernde Anfrage erhalten, ohne dass doppelte Datensätze erstellt werden oder sonstige Probleme entstehen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Legen Sie alle Reaktionen als idempotent fest. Ein idempotenter Service garantiert, dass jede Anfrage genau einmal abgeschlossen wird. Das bedeutet, dass das Senden mehrerer identischer Anfragen den gleichen Effekt hat wie das Senden einer einzelnen Anfrage.
  - Clients können API-Anfragen mit einem Idempotenz-Token ausgeben. Das gleiche Token wird bei einer Wiederholung der Anfrage verwendet. Eine idempotente Service-API gibt mithilfe des

Tokens eine Antwort zurück, die identisch mit der Antwort ist, die beim ersten Abschluss der Anfrage zurückgegeben wurde.

- Amazon EC2: Idempotenz sicherstellen

Ressourcen

Ähnliche Dokumente:

- Amazon EC2: Idempotenz sicherstellen
- Die Amazon Builders' Library: Herausforderungen bei verteilten Systemen
- Die Amazon Builders' Library: Zuverlässigkeit, stetige Ausführung und eine gute Tasse Kaffee

Ähnliche Videos:

- AWS New York Summit 2019: Einführung in ereignisgesteuerte Architekturen und Amazon EventBridge (MAD205)
- AWS re:Invent 2018: Kreisläufe schließen & aufgeschlossen sein: Wie man die Kontrolle über Systeme übernimmt – große und kleine ARC337 (umfasst lose Verkoppelung, konstante Ausführung, statische Stabilität)
- AWS re:Invent 2019: Moving to event-driven architectures (SVS308) (Umstieg auf ereignisgesteuerte Architekturen)

ZUV 5 Wie lassen sich Interaktionen in einem verteilten System so gestalten, dass Ausfälle abgemildert oder bewältigt werden?

Verteilte Systeme nutzen Kommunikationsnetzwerke, um Komponenten (wie Server oder Services) miteinander zu verbinden. Ihre Workload muss trotz Datenverlust oder höherer Latenz in diesen Netzwerken zuverlässig ausgeführt werden. Komponenten des verteilten Systems müssen so funktionieren, dass sie keine negativen Auswirkungen auf andere Komponenten oder die Workload haben. Mit den folgenden bewährten Methoden können Workloads Belastungen oder Ausfällen standhalten, schneller wiederhergestellt werden und die Auswirkungen solcher Beeinträchtigungen verringern. Das Ergebnis ist eine verbesserte mittlere Reparaturzeit (MTTR).

Bewährte Methoden

- REL05-BP01 Implementieren einer ordnungsgemäßen Funktionsminderung, um harte Abhängigkeiten in weiche zu ändern

- REL05-BP02 Drosselung von Anfragen
- REL05-BP03 Steuern und Einschränken von Wiederholungsaufrufen
- REL05-BP04 Schnelles Scheitern und Begrenzen von Warteschlangen
- REL05-BP05 Festlegen von Client-Zeitüberschreitungen
- REL05-BP06 Erstellen zustandsloser Anwendungen
- REL05-BP07 Implementieren von Nothebeln

REL05-BP01 Implementieren einer ordnungsgemäßen Funktionsminderung, um harte Abhängigkeiten in weiche zu ändern

Wenn die Abhängigkeiten einer Komponente fehlerhaft sind, kann die Komponente selbst weiterhin funktionieren, wenn auch in eingeschränkter Weise. Wenn beispielsweise ein Abhängigkeitsaufruf fehlschlägt, erfolgt ein Failover auf eine vordefinierte statische Antwort.

Ziehen Sie einen Service B in Betracht, der von Service A aufgerufen wird und wiederum Service C aufruft.

Abbildung 5: Service C schlägt fehl, wenn er von Service B aufgerufen wird. Service B gibt eine schlechtere Antwort an Service A zurück.

Wenn Service B Service C aufruft, hat er eine Fehlermeldung oder eine Zeitüberschreitung von ihm erhalten. Service B gibt ohne Antwort von Service C (und den darin enthaltenen Daten) stattdessen zurück, was möglich ist. Dies kann der letzte zwischengespeicherte Wert sein oder Service B kann eine vordefinierte statische Antwort für den Wert ersetzen, den er von Service C erhalten hätte. Er kann dann eine verminderte Antwort an seinen Aufrufer, Service A zurückgeben. Ohne diese statische Antwort würde der Fehler in Service C durch Service B zu Service A kaskadieren, was zu einem Verlust der Verfügbarkeit führt.

Gemäß dem multiplikativen Faktor in der Verfügbarkeitsgleichung für harte Abhängigkeiten (siehe Berechnen der Verfügbarkeit mit harten Abhängigkeiten) wirkt sich jeder Rückgang der Verfügbarkeit von C erheblich auf die effektive Verfügbarkeit von B aus. Durch die Rückgabe des

statischen Antwortservice B wird der Fehler in C verringert und lässt die Verfügbarkeit von Service C wie eine hundertprozentige Verfügbarkeit aussehen (vorausgesetzt, dass er die statische Antwort unter Fehlerbedingungen zuverlässig zurückgibt). Beachten Sie, dass die statische Antwort eine einfache Alternative zur Rückgabe eines Fehlers darstellt und kein Versuch ist, die Antwort mit anderen Mitteln neu zu berechnen. Solche Versuche mit einem völlig anderen Mechanismus, um dasselbe Ergebnis zu erzielen, werden als Fallback-Verhalten bezeichnet und sind ein zu vermeidendes Antimuster.

Ein weiteres Beispiel für eine ordnungsgemäße Verschlechterung ist der Schaltkreisunterbrecher. Wiederholungsstrategien sollten verwendet werden, wenn der Fehler vorübergehend ist. Wenn dies nicht der Fall ist und der Vorgang wahrscheinlich fehlschlägt, verhindert der Unterbrecher, dass der Client eine Anfrage ausführt, die wahrscheinlich fehlschlägt. Wenn die Anfragen normal verarbeitet werden, wird der Unterbrecher geschlossen, und die Anfragen laufen durch. Wenn das Remote-System Fehler zurückgibt oder eine hohe Latenz aufweist, wird der Unterbrecher geöffnet und die Abhängigkeit wird ignoriert oder die Ergebnisse werden durch einfachere aber weniger umfassende Antworten ersetzt (was einfach ein Antwort-Cache sein kann). Das System versucht regelmäßig, die Abhängigkeit aufzurufen, um zu ermitteln, ob sie wiederhergestellt wurde. In diesem Fall wird der Unterbrecher geschlossen.

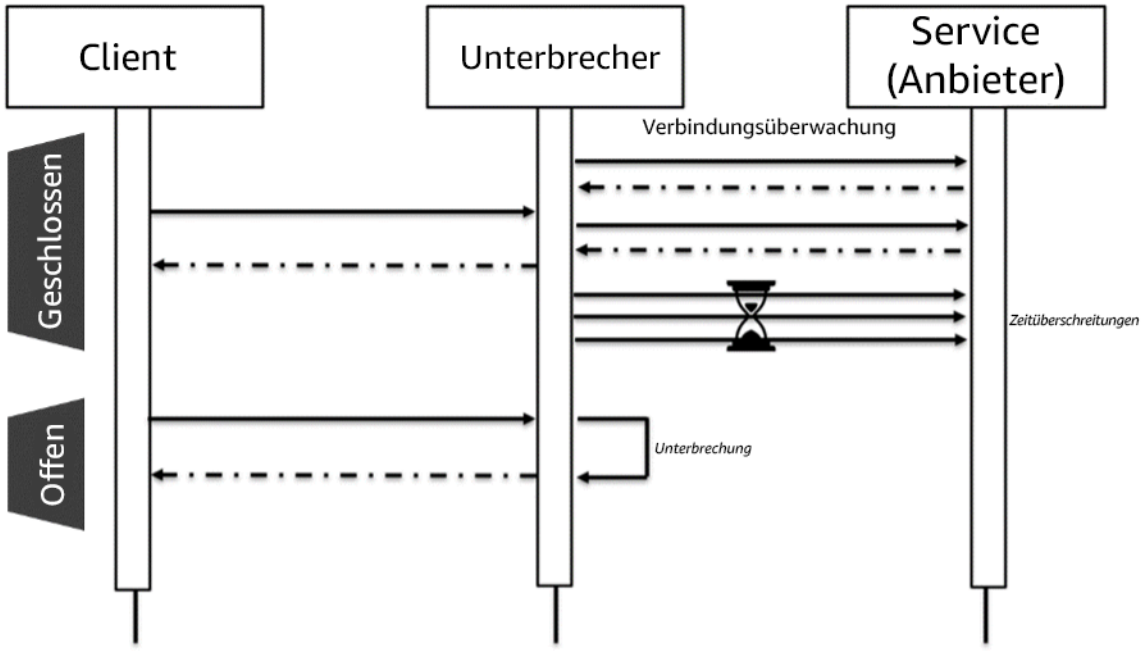


Abbildung 6: Unterbrecher mit dem Status "Geschlossen" und "Offen".

Zusätzlich zu den im Diagramm gezeigten Zuständen "Geschlossen" und "Offen" kann der Unterbrecher nach einem konfigurierbaren Zeitraum im offenen Zustand in "Halboffen" übergehen.

In diesem Zustand versucht er in regelmäßigen Abständen, den Service mit einer viel geringeren Rate als normal aufzurufen. Auf diese Weise wird der Zustand des Service überprüft. Nach einigen Erfolgen im halb geöffneten Zustand wechselt der Unterbrecher zu "closed" und die Anfragen werden normal fortgesetzt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren Sie eine ordnungsgemäße Funktionsminderung, um harte Abhängigkeiten in weiche zu ändern. Wenn die Abhängigkeiten einer Komponente fehlerhaft sind, kann die Komponente selbst weiterhin funktionieren, wenn auch in eingeschränkter Weise. Wenn beispielsweise ein Abhängigkeitsaufruf fehlschlägt, erfolgt ein Failover auf eine vordefinierte statische Antwort.
  - Durch Rückgabe einer statischen Antwort grenzt die Workload Fehler in ihren Abhängigkeiten ein.
    - Well-Architected Lab: Level 300: Implementieren von Zustandsprüfungen und Verwalten von Abhängigkeiten zur Verbesserung der Zuverlässigkeit
  - Erkennen Sie, wann eine Wiederholungsoperation wahrscheinlich fehlschlägt, und verhindern Sie mit dem Unterbrecher die Wiederholung fehlgeschlagener Aufrufe durch den Client.
    - CircuitBreaker

Ressourcen

Relevante Dokumente:

- Amazon API Gateway: Drosselung von API-Anfragen für höheren Durchsatz
- CircuitBreaker (Zusammenfassung des Circuit Breaker aus dem Buch „Release It!“)
- Fehlerwiederholungen und exponentielles Backoff in AWS
- Michael Nygard, „Release It!“ Design and Deploy Production-Ready Software“
- Die Amazon Builders' Library: Vermeiden von Fallback in verteilten Systemen
- Die Amazon Builders' Library: Vermeiden von nicht mehr aufholbaren Warteschlangen-Rückständen
- Die Amazon Builders' Library: Herausforderungen und Strategien für das Caching
- Die Amazon Builders' Library: Timeouts, Wiederholungen und Backoff mit Jitter

Relevante Videos:

- Wiederholung, Backoff und Jitter: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)

Ähnliche Beispiele:

- Well-Architected Lab: Level 300: Implementieren von Zustandsprüfungen und Verwalten von Abhängigkeiten zur Verbesserung der Zuverlässigkeit

REL05-BP02 Drosselung von Anfragen

Die Drosselung von Anfragen ist ein Abschwächungsmuster, um auf einen unerwarteten Anstieg der Nachfrage zu reagieren. Einige Anfragen werden berücksichtigt, doch solche, die über ein definiertes Limit hinausgehen, werden abgelehnt. Zudem wird eine entsprechende Meldung über deren Drosselung zurückgegeben. Dabei wird erwartet, dass Clients die Anfragen abbrechen oder es mit einer langsameren Rate erneut versuchen.

Ihre Services sollten so konzipiert werden, dass jeder Knoten und jede Zelle eine bekannte Anzahl von Anfragen verarbeiten kann. Diese Kapazität kann mit Lasttests erreicht werden. Anschließend müssen Sie die Eingangsrate von Anfragen verfolgen, und wenn die vorübergehende Eingangsrate dieses Limit überschreitet, muss aus der entsprechenden Antwort hervorgehen, dass die Anfrage gedrosselt wurde. Damit kann der Benutzer es bei einem anderen Knoten oder einer anderen Zelle, der/die ggf. über Kapazität verfügt, erneut versuchen. Amazon API Gateway bietet Methoden zum Drosseln von Anfragen. Amazon SQS und Amazon Kinesis können Anfragen puffern, die Anfragerate glätten und die Notwendigkeit einer Drosselung für asynchrone Anfragen verringern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Drosseln Sie Anfragen. Hierbei handelt es sich um ein Abmilderungsmuster, mit dem auf einen unerwarteten Anstieg der Nachfrage reagiert werden kann. Einige Anfragen werden berücksichtigt, doch solche, die über ein definiertes Limit hinausgehen, werden abgelehnt. Zudem wird eine entsprechende Meldung über deren Drosselung zurückgegeben. Dabei wird erwartet, dass Clients die Anfragen abbrechen oder es mit einer langsameren Rate erneut versuchen.
  - Nutzen Sie Amazon API Gateway
    - Drosseln Sie API-Anfragen, um einen höheren Durchsatz zu erzielen.

Ressourcen

Relevante Dokumente:

- Amazon API Gateway: Drosselung von API-Anfragen für höheren Durchsatz
- Fehlerwiederholungen und exponentielles Backoff in AWS
- Die Amazon Builders' Library: Vermeiden von Fallback in verteilten Systemen
- Die Amazon Builders' Library: Vermeiden von nicht mehr aufholbaren Warteschlangen-Rückständen
- Die Amazon Builders' Library: Timeouts, Wiederholungen und Backoff mit Jitter
- Drosseln Sie API-Anfragen, um einen höheren Durchsatz zu erzielen.

Relevante Videos:

- Wiederholung, Backoff und Jitter: AWS re:Invent 2019: Einführung in die Amazon Builders’ Library (DOP328)

REL05-BP03 Steuern und Einschränken von Wiederholungsaufrufen

Verwenden Sie ein exponentielles Backoff, um Aufrufe nach zunehmend längeren Intervallen zu wiederholen. Nutzen Sie Jitter, um die Wiederholungsintervalle zu randomisieren, und legen Sie ein Limit für die Zahl der Wiederholungen fest.

Typische Komponenten in einem verteilten Softwaresystem sind Server, Load Balancer, Datenbanken und DNS-Server. Im Betrieb und bei Ausfällen kann jede dieser Komponenten mit dem Generieren von Fehlern beginnen. Die Standardmethode für den Umgang mit Fehlern besteht darin, Wiederholungen auf Clientseite zu implementieren. Diese Methode erhöht die Zuverlässigkeit und Verfügbarkeit der Anwendung. In großem Umfang – und wenn Clients versuchen, den fehlgeschlagenen Vorgang zu wiederholen, sobald ein Fehler auftritt – kann das Netzwerk schnell mit neuen und nicht mehr aktiven Anfragen überfordert werden, wobei jede Anfrage um die Netzwerkbandbreite konkurriert. Dies kann zu einer Wiederholungsflut führen, die die Verfügbarkeit des Service reduziert. Dieses Muster setzt sich möglicherweise fort, bis ein vollständiger Systemausfall auftritt.

Um solche Szenarien zu vermeiden, sollten Backoff-Algorithmen wie das gängige exponentielle Backoff verwendet werden. Algorithmen für exponentielles Backoff verringern schrittweise die Rate, mit der Wiederholungen durchgeführt werden, und verhindern dadurch Netzwerküberlastungen.

Viele SDKs und Softwarebibliotheken, auch die von AWS, implementieren eine Version dieser Algorithmen. Sie sollten jedoch nie davon ausgehen, dass ein Backoff-Algorithmus vorhanden ist. Dies muss stets im Voraus getestet und überprüft werden.

Ein einfaches Backoff alleine reicht nicht aus, da in verteilten Systemen alle Clients gleichzeitig einen Backoff durchführen und Anhäufungen von Wiederholungsaufrufen erstellen können. Marc Brooker erklärt in seinem Blogbeitrag Exponentielles Backoff und Jitter, wie die Funktion „wait()“ im exponentiellen Backoff geändert werden kann, um Wiederholungsaufruf-Cluster zu verhindern. Die Lösung besteht darin, Jitter der Funktion „wait()“ hinzuzufügen. Um einen zu langen Wiederholungsversuch zu vermeiden, sollten Implementierungen das Backoff auf einen maximalen Wert begrenzen.

Schließlich ist es wichtig, eine maximale Anzahl von Wiederholungen oder die verstrichene Zeit zu konfigurieren, nach der Wiederholversuche einfach fehlschlagen. AWS SDKs implementieren dies als konfigurierbaren Standard. Für Services, die sich weiter unten im Stack befinden, kann ein maximales Wiederholungslimit von null oder eins das Risiko begrenzen und ist dennoch wirksam, da Wiederholungen an Services delegiert werden, die sich höher im Stack befinden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Steuern und begrenzen Sie Wiederholungsaufrufe. Verwenden Sie ein exponentielles Backoff, um Aufrufe nach zunehmend längeren Intervallen zu wiederholen. Nutzen Sie Jitter, um die Wiederholungsintervalle zu randomisieren, und legen Sie ein Limit für die Zahl der Wiederholungen fest.
  - Fehlerwiederholungen und exponentielles Backoff in AWS
    - Mit Amazon SKDs werden Wiederholungen und exponentielles Backoff standardmäßig implementiert. Implementieren Sie in die Abhängigkeitsebene zum Aufrufen Ihrer eigenen abhängigen Services eine ähnliche Logik. Legen Sie entsprechend Ihrem Anwendungsfall Zeitüberschreitungen fest und geben Sie an, wann Wiederholversuche gestoppt werden sollen.

Ressourcen

Relevante Dokumente:

- Amazon API Gateway: Drosselung von API-Anfragen für höheren Durchsatz
- Fehlerwiederholungen und exponentielles Backoff in AWS

- Die Amazon Builders' Library: Vermeiden von Fallback in verteilten Systemen
- Die Amazon Builders' Library: Vermeiden von nicht mehr aufholbaren Warteschlangen-Rückständen
- Die Amazon Builders' Library: Herausforderungen und Strategien für das Caching
- Die Amazon Builders' Library: Timeouts, Wiederholungen und Backoff mit Jitter

Relevante Videos:

- Wiederholung, Backoff und Jitter: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)

REL05-BP04 Schnelles Scheitern und Begrenzen von Warteschlangen

Wenn eine Workload auf eine Anfrage nicht erfolgreich antworten kann, sollte sie per Fail-Fast beendet werden. So können Ressourcen freigegeben werden, die einer Anfrage zugeordnet sind, und der Service kann entlastet werden, wenn die Ressourcen zur Neige gehen. Wenn die Workload erfolgreich antworten kann, aber die Rate der Anfragen zu hoch ist, sollten Sie die Anfragen mithilfe einer Warteschlange puffern. Lassen Sie jedoch keine langen Warteschlangen zu. Sie können dazu führen, dass veraltete Anfragen verarbeitet werden, die der Client bereits aufgegeben hat.

Diese bewährte Methode gilt für den Server oder den Empfänger der Anfrage.

Beachten Sie, dass Warteschlangen auf mehreren Ebenen eines Systems erstellt werden können und die Möglichkeit einer schnellen Wiederherstellung möglicherweise erheblich beeinträchtigt wird, da ältere veraltete Anfragen (die keine Antwort mehr benötigen) vor neueren Anfragen verarbeitet werden. Machen Sie sich mit den Orten vertraut, an denen Warteschlangen vorhanden sind. Sie verbergen sich häufig in Workflows oder in Daten, die in einer Datenbank aufgezeichnet werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Implementieren Sie schnelles Scheitern und begrenzen Sie Warteschlangen. Wenn eine Workload auf eine Anfrage nicht erfolgreich antworten kann, sollte sie per schnellem Scheitern beendet werden. So können Ressourcen freigegeben werden, die einer Anfrage zugeordnet sind, und der Service kann entlastet werden, wenn die Ressourcen zur Neige gehen. Wenn die Workload erfolgreich antworten kann, aber die Rate der Anfragen zu hoch ist, sollten Sie die Anfragen mithilfe

einer Warteschlange puffern. Lassen Sie jedoch keine langen Warteschlangen zu. Sie können dazu führen, dass veraltete Anfragen verarbeitet werden, die der Client bereits aufgegeben hat.

- Implementieren Sie schnelles Scheitern bei hoher Belastung eines Service.
  - Schnell scheitern
- Begrenzen Sie Warteschlagen. Wenn in einem warteschlangenbasierten System die Verarbeitung gestoppt wird, aber weiterhin Nachrichten eintreffen, kann ein großer Rückstand an unverarbeiteten Nachrichten entstehen, was die Verarbeitungszeit erhöht. Die Verarbeitung kann so spät abgeschlossen werden, dass die Ergebnisse nicht mehr nützlich sind. Dies führt zur Beeinträchtigung der Verfügbarkeit, die mit der Warteschlange eigentlich aufrecht erhalten werden sollte.
  - Die Amazon Builders' Library: Vermeiden von nicht mehr aufzuholenden Rückständen

Ressourcen

Ähnliche Dokumente:

- Fehlerwiederholungen und exponentielles Backoff in AWS
- Schnell scheitern
- Die Amazon Builders' Library: Vermeiden von Fallback in verteilten Systemen
- Die Amazon Builders' Library: Vermeiden von nicht mehr aufholbaren Warteschlangen-Rückständen
- Die Amazon Builders' Library: Herausforderungen und Strategien für das Caching
- Die Amazon Builders' Library: Timeouts, Wiederholungen und Backoff mit Jitter

Ähnliche Videos:

- Wiederholung, Backoff und Jitter: AWS re:Invent 2019: Einführung in die Amazon Builders’ Library (DOP328)

REL05-BP05 Festlegen von Client-Zeitüberschreitungen

Legen Sie Zeitbeschränkungen entsprechend fest, überprüfen Sie sie systematisch und verlassen Sie sich nicht auf Standardwerte, da diese in der Regel zu hoch eingestellt sind.

Diese bewährte Methode gilt für den Client oder den Absender der Anfrage.

Legen Sie sowohl eine Zeitbeschränkung für Verbindungen als auch eine für Anforderungen für jeden Remote-Aufruf und in der Regel für jeden Prozess-übergreifenden Aufruf fest. Viele Frameworks bieten integrierte Zeitbeschränkungsfunktionen, deren Standardwerte jedoch oft unendlich oder zu hoch sind. Ein zu hoher Wert reduziert die Nützlichkeit der Zeitbeschränkung, da Ressourcen weiterhin verbraucht werden, während der Client auf das Einsetzen der Zeitbeschränkung wartet. Ein zu niedriger Wert kann zu erhöhtem Datenverkehr im Backend und zu erhöhter Latenz führen, da zu viele Anfragen wiederholt werden. In einigen Fällen kann dies zu vollständigen Ausfällen führen, da alle Anfragen wiederholt werden.

Weitere Informationen dazu, wie Amazon Zeitüberschreitungen, Wiederholungen und Backoff mit Jitter verwendet, finden Sie in der Builders' Library unter Zeitüberschreitungen, Wiederholungsversuche und Backoff mit Jitter.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Legen Sie sowohl eine Zeitbeschränkung für Verbindungen als auch eine für Anforderungen für jeden Remote-Aufruf und in der Regel für jeden Prozess-übergreifenden Aufruf fest. Viele Frameworks bieten integrierte Zeitbeschränkungsfunktionen, deren Standardwerte jedoch oft unendlich oder zu hoch sind. Ein zu hoher Wert reduziert die Nützlichkeit der Zeitbeschränkung, da Ressourcen weiterhin verbraucht werden, während der Client auf das Einsetzen der Zeitbeschränkung wartet. Ein zu niedriger Wert kann zu erhöhtem Datenverkehr im Backend und zu erhöhter Latenz führen, da zu viele Anfragen wiederholt werden. In einigen Fällen kann dies zu vollständigen Ausfällen führen, da alle Anfragen wiederholt werden.
  - AWS SDK: Wiederholungen und Zeitüberschreitungen

Ressourcen

Relevante Dokumente:

- AWS SDK: Wiederholungen und Zeitüberschreitungen
- Amazon API Gateway: Drosselung von API-Anfragen für höheren Durchsatz
- Fehlerwiederholungen und exponentielles Backoff in AWS
- Die Amazon Builders' Library: Timeouts, Wiederholungen und Backoff mit Jitter

Relevante Videos:

- Wiederholung, Backoff und Jitter: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)

REL05-BP06 Erstellen zustandsloser Anwendungen

Services sollten entweder keinen Zustand erfordern oder ihn so auslagern, dass zwischen verschiedenen Client-Anfragen keine Abhängigkeit von lokal gespeicherten Daten auf der Festplatte und im Arbeitsspeicher besteht. Auf diese Weise können Server nach Belieben ersetzt werden, ohne dass dies Auswirkungen auf die Verfügbarkeit hat. Amazon ElastiCache oder Amazon DynamoDB sind gute Ziele für den ausgelagerte Zustand.

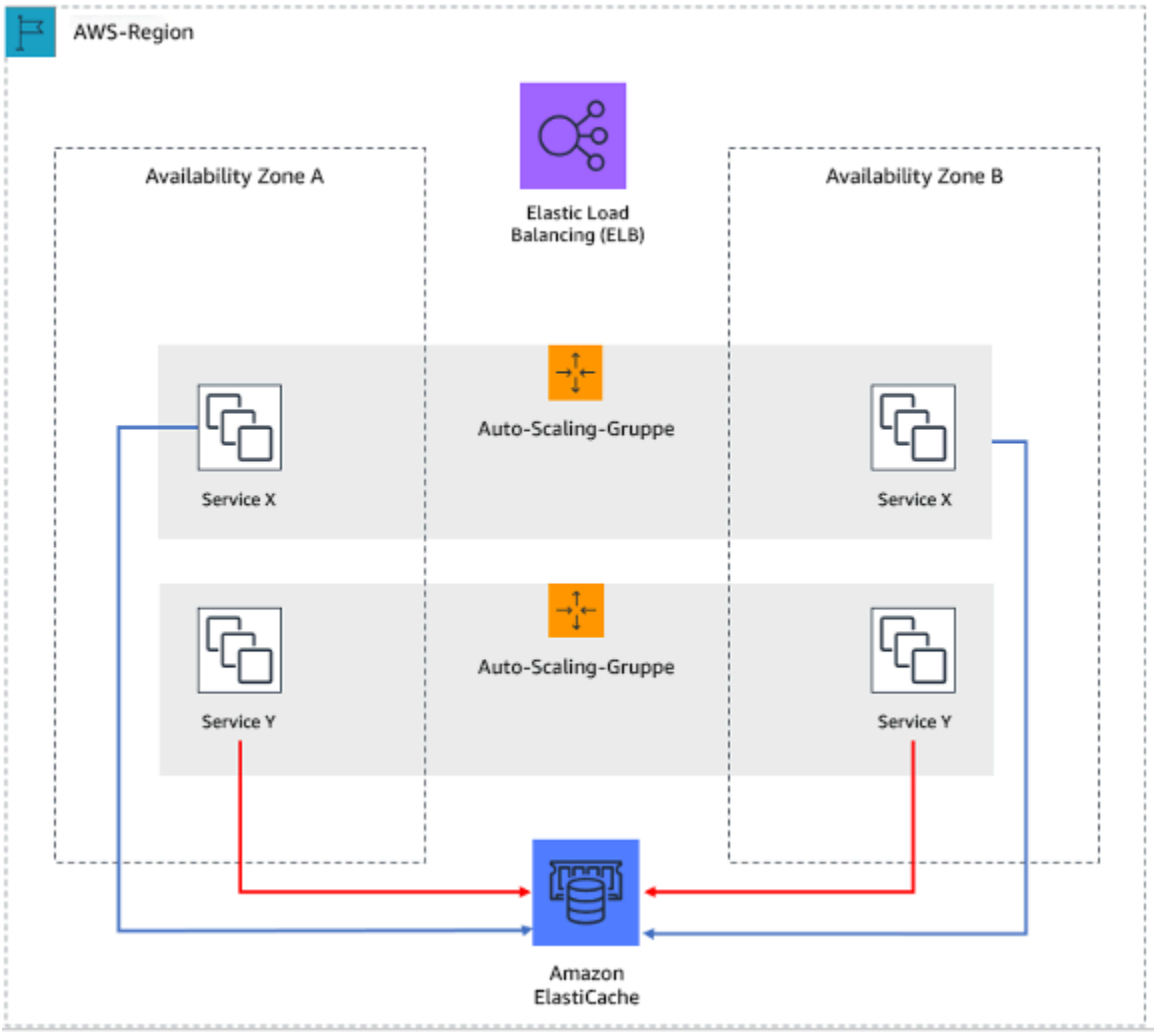


Abbildung 7: In dieser zustandslosen Webanwendung wird der Sitzungsstatus in Amazon ElastiCache ausgelagert.

Wenn Benutzer oder Services mit einer Anwendung interagieren, führen sie häufig eine Reihe von Interaktionen aus, die eine Sitzung bilden. Bei einer Sitzung handelt es sich um eindeutige Daten für

Benutzer, die zwischen Anfragen bestehen bleiben, während sie die Anwendung verwenden. Eine zustandslose Anwendung ist eine Anwendung, die keine Informationen zu früheren Interaktionen benötigt und keine Sitzungsinformationen speichert.

Sobald eine Anwendung als zustandslos entwickelt wurde, können Sie serverlose Compute-Services wie AWS Lambda oder AWS Fargate verwenden.

Neben dem Serverersatz besteht ein weiterer Vorteil zustandsloser Anwendungen darin, dass sie horizontal skaliert werden können, da alle verfügbaren Compute-Ressourcen (z. B. EC2-Instances und AWS Lambda-Funktionen) jede Anfrage bearbeiten können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Erstellen Sie zustandslose Anwendungen. Zustandslose Anwendungen ermöglichen eine horizontale Skalierung und sind gegenüber dem Ausfall eines einzelnen Knotens tolerant.
  - Entfernen Sie Zustände, die tatsächlich in Anfrageparametern gespeichert werden können.
  - Nachdem Sie untersucht haben, ob der Zustand erforderlich ist, verschieben Sie die gesamte Zustandsverfolgung in einen ausfallsicheren Multizonen-Cache oder Datenspeicher wie Amazon ElastiCache, Amazon RDS, Amazon DynamoDB oder in die verteilte Datenlösung eines Drittanbieters. Speichern Sie nicht verlagerbare Zustände in ausfallsicheren Datenspeichern.
    - Manche Daten (wie Cookies) können in Headern oder Abfrageparametern übergeben werden.
    - Entfernen Sie Zustände, die sich schnell in Anfragen übergeben lassen.
    - Einige Daten sind möglicherweise nicht für jede Anfrage erforderlich, sondern können bei Bedarf abgerufen werden.
    - Entfernen Sie asynchron abrufbare Daten.
    - Wählen Sie einen Datenspeicher, der die Anforderungen eines erforderlichen Zustands erfüllt.
    - Ziehen Sie für nichtrelationale Daten eine NoSQL-Datenbank in Erwägung.

Ressourcen

Ähnliche Dokumente:

- Die Amazon Builders' Library: Vermeiden von Fallback in verteilten Systemen
- Die Amazon Builders' Library: Vermeiden von nicht mehr aufholbaren Warteschlangen-Rückständen

- Die Amazon Builders' Library: Herausforderungen und Strategien für das Caching

REL05-BP07 Implementieren von Nothebeln

Nothebel sind schnelle Prozesse, die die Auswirkungen auf die Verfügbarkeit Ihrer Workload mindern können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Implementieren von Nothebeln. Hierbei handelt es sich um schnelle Prozesse, die die Auswirkungen auf die Verfügbarkeit Ihrer Workload mindern können. Sie können ausgeführt werden, wenn keine Ursache vorliegt. Ein idealer Nothebel reduziert die kognitive Belastung der Resolver auf null, indem er vollständig deterministische Aktivierungs- und Deaktivierungskriterien bereitstellt. Hebel müssen oft manuell ausgeführt werden, können aber auch automatisiert werden.
  - Beispiele für Nothebel:
    - Blockieren des gesamten Roboterdatenverkehrs
    - Anbieten von statischen Seiten anstelle von dynamischen Seiten
    - Reduzieren der Häufigkeit von Aufrufen für eine Abhängigkeit
    - Drosselung der Aufrufe von Abhängigkeiten
  - Tipps zur Implementierung und Verwendung von Nothebeln
    - Wenn die Hebel aktiviert sind, sollten Sie WENIGER machen, nicht mehr.
    - Halten Sie die Dinge einfach, vermeiden Sie bimodales Verhalten.
    - Testen Sie die Hebel regelmäßig.
  - Nachfolgend finden Sie Beispiele für Aktionen, die KEINE Nothebel sind:
    - Kapazität hinzufügen
    - Servicebesitzer von Clients anrufen, die von Ihrem Service abhängig sind, und sie bitten, Aufrufe zu reduzieren
    - Code ändern und freigeben

Änderungsverwaltung

Fragen

- ZUV 6 Was ist bei der Überwachung von Workload-Ressourcen zu beachten?

- ZUV 7 Wie lässt sich der Workload so gestalten, dass er sich an Bedarfsänderungen anpasst?
- ZUV 8 Wie implementieren Sie Änderungen?

ZUV 6 Was ist bei der Überwachung von Workload-Ressourcen zu beachten?

Protokolle und Metriken sind wertvolle Tools, um einen Einblick in den Zustand Ihrer Workloads zu gewinnen. Sie können Ihre Workload so konfigurieren, dass Protokolle und Metriken überwacht und bei Über- oder Unterschreiten von Schwellenwerten oder wichtigen Ereignissen Benachrichtigungen gesendet werden. Dank der Überwachung kann die Workload erkennen, wenn Schwellenwerte für eine niedrige Leistung unterschritten werden oder Ausfälle auftreten, sodass als Reaktion drauf eine automatische Wiederherstellung erfolgen kann.

Bewährte Methoden

- REL06-BP01 Überwachen aller Komponenten der Workload (Generierung)
- REL06-BP02 Definieren und Berechnen von Metriken (Aggregierung)
- REL06-BP03 Senden von Benachrichtigungen (Verarbeitung und Benachrichtigung in Echtzeit)
- REL06-BP04 Automatisieren von Antworten (Verarbeitung und Benachrichtigung in Echtzeit)
- REL06-BP05 Analysen
- REL06-BP06 Regelmäßiges Durchführen von Prüfungen
- REL06-BP07 Überwachen der gesamten Nachverfolgung von Anfragen im System

REL06-BP01 Überwachen aller Komponenten der Workload (Generierung)

Überwachen Sie die Komponenten der Workload mit Amazon CloudWatch oder Tools von Drittanbietern. Überwachen Sie AWS-Services mit dem AWS Health Dashboard.

Alle Komponenten Ihrer Workload sollten überwacht werden, einschließlich Frontend, Geschäftslogik und Speicherstufen. Definieren Sie Schlüsselmetriken, beschreiben Sie, wie Sie diese gegebenenfalls aus Protokollen extrahieren, und legen Sie Schwellenwerte für das Auslösen entsprechender Alarmereignisse fest. Stellen Sie sicher, dass die Metriken für die wichtigen Leistungskennzahlen (KPIs) Ihrer Workload relevant sind und verwenden Sie Metriken und Protokolle, um frühe Warnzeichen einer Serviceverschlechterung zu identifizieren. Beispielsweise kann eine mit Geschäftsergebnissen zusammenhängende Metrik wie etwa die Anzahl der pro Minute erfolgreich verarbeiteten Bestellungen schneller auf Workload-Probleme hinweisen als eine technische Metrik wie etwa die CPU-Auslastung. Verwenden Sie das AWS Health Dashboard für

eine personalisierte Ansicht der Leistung und Verfügbarkeit der AWS-Services, die Ihren AWS-Ressourcen zugrunde liegen.

Die Überwachung in der Cloud bietet neue Möglichkeiten. Die meisten Cloudanbieter haben anpassbare Hooks entwickelt und können Einblicke liefern, mit denen Sie mehrere Ebenen Ihrer Workload überwachen können. AWS-Services wie Amazon CloudWatch wenden statistische und Machine-Learning-Algorithmen an, um Metriken von Systemen und Anwendungen kontinuierlich zu analysieren, normale Basiswerte zu erkennen und Oberflächenanomalien anhand eines minimalen Benutzereingriffs aufzudecken. Algorithmen zur Erkennung von Anomalien berücksichtigen saisonale Schwankungen und Trendänderungen von Metriken.

AWS stellt zahlreiche Überwachungs- und Protokollinformationen bereit, die genutzt werden können, um workload-spezifische Metriken und Bedarfsänderungsprozesse zu definieren und Machine-Learning-Verfahren unabhängig von der ML-Erfahrung einzuführen.

Zudem können Sie auch all Ihre externen Endpunkte überwachen, um sicherzustellen, dass diese von Ihrer Basisimplementierung unabhängig sind. Diese aktive Überwachung kann anhand von synthetischen Transaktionen erfolgen (auch Benutzer-Canariesgenannt, jedoch nicht zu verwechseln mit Canary-Bereitstellungen). Diese führen regelmäßig eine Reihe gängiger Aufgaben aus, die mit Aktionen übereinstimmen, die von Clients der Workload durchgeführt werden. Diese Aufgaben sollten nicht zu lang sein und Sie sollten darauf achten, Ihre Workload beim Testen nicht zu überlasten. Mit Amazon CloudWatch Synthetics können Sie synthetische Canaries erstellen, um Ihre Endpunkte und APIs zu überwachen. Sie können die synthetischen Canary-Client-Knoten auch mit der AWS X-Ray-Konsole kombinieren, um zu bestimmen, bei welchen synthetischen Canaries im ausgewählten Zeitraum Probleme mit Fehlern, Störungen oder Drosselungsraten auftreten.

Gewünschtes Ergebnis:

Erfassen und Nutzen kritischer Metriken aus allen Komponenten der Workload, um die Workload-Zuverlässigkeit und eine optimale Benutzererfahrung sicherzustellen. Zu erkennen, dass eine Workload keine Geschäftsergebnisse erzielt, ermöglicht es Ihnen, schnell einen Systemausfall zu deklarieren und das System nach einem Vorfall wiederherzustellen.

Gängige Antimuster:

- Es werden nur externe Schnittstellen zur Workload überwacht.
- Es werden keine workload-spezifischen Metriken erzeugt und Sie verlassen sich nur auf Metriken, die Ihnen von den AWS-Services, die Ihre Workload verwendet, bereitgestellt werden.

- Es werden nur technische Metriken in Ihrer Workload verwendet und es werden keinerlei Metriken im Zusammenhang mit nicht-technischen KPIs, zu denen die Workload beiträgt, überwacht.
- Sie verlassen sich auf den Produktionsdatenverkehr und einfache Zustandsprüfungen für die Überwachung und Bewertung des Workload-Status.

Vorteile der Einführung dieser bewährten Methode: Durch die Überwachung aller Ebenen Ihrer Workload können Sie Probleme in den darin enthaltenen Komponenten schneller vorhersehen und beheben.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

1. Aktivieren Sie die Protokollierung, wann immer verfügbar. Von allen Workload-Komponenten sollten Überwachungsdaten erzielt werden. Aktivieren Sie eine zusätzliche Protokollierung, wie etwa S3 Access Logs, und ermöglichen Sie es Ihrer Workload, die workload-spezifischen Daten zu protokollieren. Erfassen Sie Metriken für die Durchschnittswerte zu CPU, Netzwerk-E/A und Laufwerk-E/A von Services wie Amazon ECS, Amazon EKS, Amazon EC2, Elastic Load Balancing, AWS Auto Scaling und Amazon EMR. Unter AWS-Services, die CloudWatch-Metriken veröffentlichen finden Sie eine Liste an AWS-Services, die Metriken in CloudWatch veröffentlichen.
2. Sehen Sie sich alle Standardmetriken an, um mehr über mögliche Datenerfassungslücken zu erfahren. Jeder Service generiert Standardmetriken. Durch die Erfassung von Standardmetriken erhalten Sie ein besseres Verständnis über die Abhängigkeiten zwischen Workload-Komponenten und darüber, wie die Komponentenzuverlässigkeit und -leistung die Workload beeinträchtigen. Sie können auch Ihre eigenen Metriken in CloudWatch unter Verwendung der AWS CLI oder einer API erstellen und veröffentlichen. Dies
3. Bewerten Sie alle Metriken, um zu entscheiden, für welche eine Warnmeldung für jeden AWS-Service in Ihrer Workload eingerichtet werden soll. Sie können eine Metriken-Untergruppe auswählen, die eine höhere Auswirkung auf die Workload-Zuverlässigkeit hat. Wenn Sie sich auf kritische Metriken und Schwellenwerte konzentrieren, können Sie die Anzahl an Warnmeldungen genauer definieren und so Falschmeldungen reduzieren.
4. Definieren Sie Warnungen und den Wiederherstellungsprozess für Ihre Workload nach dem Auslösen der Warnmeldung. Das Definieren von Warnmeldungen ermöglicht es Ihnen, schnell zu benachrichtigen, zu eskalieren und die für die Wiederherstellung nach einem Vorfall erforderlichen Schritte durchzuführen, um so Ihren festgelegten Recovery Time Objective (RTO) zu erfüllen. Sie können Amazon CloudWatch-Alarme für das Aufrufen von automatisierten Workflows und

die Initiierung von Wiederherstellungsverfahren basierend auf definierten Schwellenwerten verwenden.

5. Erfahren Sie mehr über die Verwendung von synthetischen Transaktionen für das Erfassen relevanter Daten zum Workload-Status. Die synthetische Überwachung folgt denselben Routen und führt dieselben Aktionen aus wie ein Kunde. Dadurch haben Sie die Möglichkeit, die Kundenerfahrung kontinuierlich zu überprüfen, selbst, wenn Sie keinen Kundendatenverkehr auf Ihren Workloads haben. Durch die Verwendung von synthetischen Transaktionenkönnen Sie Probleme erkennen, bevor Ihre Kunden dies tun.

Ressourcen

Relevante bewährte Methoden:

- REL11-BP03 Automatisieren der Reparatur auf allen Ebenen

Relevante Dokumente:

- Getting started with your AWS Health Dashboard – Your account health (Erste Schritte mit Ihrem AWS Health-Dashboard – Der Zustand Ihres Kontos)
- AWS-Services, die CloudWatch-Metriken veröffentlichen
- Zugriffsprotokolle für Ihren Network Load Balancer
- Zugriffsprotokolle für Ihre Application Load Balancer
- Zugriff auf Amazon CloudWatch Logs für AWS Lambda
- Protokollierung von Amazon S3-Serverzugriffen
- Aktivieren Sie Zugriffsprotokolle für Ihren Classic Load Balancer.
- Exportieren von Protokolldaten zu Amazon S3
- Installieren des CloudWatch-Agenten
- Veröffentlichen benutzerdefinierter Metriken
- Verwenden von Amazon CloudWatch-Dashboards
- Verwenden von Amazon CloudWatch-Metriken
- Verwenden von Synthetic Monitoring
- Was sind Amazon CloudWatch Logs?

Benutzerhandbücher:

- Erstellen eines Trails
- Überwachen von Arbeitsspeicher- und Datenträgermetriken für Amazon EC2 Linux-Instances
- Verwenden von CloudWatch Logs mit Container-Instances
- VPC Flow Logs
- Was ist Amazon DevOps Guru?
- Was ist AWS X-Ray?

Ähnliche Blogs:

- Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray

Ähnliche Beispiele und Workshops:

- AWS Well-Architected Labs: Operational Excellence - Dependency Monitoring (AWS Well-Architected Labs: Operative Exzellenz – Überwachung von Abhängigkeiten)
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- Workshop zur Beobachtbarkeit

REL06-BP02 Definieren und Berechnen von Metriken (Aggregierung)

Speichern Sie Protokolldaten und wenden Sie gegebenenfalls Filter an, um Metriken zu berechnen. Dazu gehören z. B. die Anzahl eines bestimmten Protokollereignisses oder die Latenz, die aus den Zeitstempeln des Protokollereignisses berechnet wird.

Amazon CloudWatch und Amazon S3 dienen als primäre Aggregierungs- und Speicherebenen. Bei einigen Services wie AWS Auto Scaling und Elastic Load Balancing werden Standardkennzahlen für die CPU-Last oder die durchschnittliche Anfragelatenz eines Clusters oder einer Instance bereitgestellt. Für Streaming-Services wie VPC Flow Logs und AWS CloudTrail werden Ereignisdaten an CloudWatch Logs weitergeleitet und Sie müssen Filter definieren und anwenden, um Metriken aus diesen Ereignisdaten zu extrahieren. Auf diese Weise erhalten Sie Zeitreihendaten, die als Eingaben für CloudWatch-Alarme dienen können, die Sie zum Auslösen von Warnungen definieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Definieren und berechnen Sie Metriken (Aggregierung). Speichern Sie Protokolldaten und wenden Sie gegebenenfalls Filter an, um Metriken zu berechnen. Dazu gehören z. B. die Anzahl eines bestimmten Protokollereignisses oder die Latenz, die aus den Zeitstempeln des Protokollereignisses berechnet wird.
  - Metrikfilter definieren die Begriffe und Muster, die in Protokolldaten zu suchen sind, wenn diese an CloudWatch Logs gesendet werden. CloudWatch Logs verwendet diese Metrikfilter, um Protokolldaten in numerische CloudWatch-Metriken umzuwandeln, die Sie grafisch darstellen oder für die Sie einen Alarm einrichten können.
    - Suchen und Filtern von Protokolldaten
  - Verwenden Sie einen vertrauenswürdigen Drittanbieter für die Protokollaggregierung.
    - Befolgen Sie die Anweisungen des Drittanbieters. Die meisten Produkte von Drittanbietern lassen sich in CloudWatch und Amazon S3 integrieren.
  - Einige AWS-Services können Protokolle direkt in Amazon S3 veröffentlichen. Wenn die Speicherung von Protokollen in Amazon S3 die wichtigste Anforderung ist, kann der Protokoll-Service die Protokolle direkt an Amazon S3 senden, ohne dass eine zusätzliche Infrastruktur eingerichtet werden muss.
    - Senden von Protokollen direkt an Amazon S3

Ressourcen

Relevante Dokumente:

- Amazon CloudWatch Logs Insights-Beispielabfragen
- Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray
- Workshop zur Beobachtbarkeit
- Suchen und Filtern von Protokolldaten
- Senden von Protokollen direkt an Amazon S3
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen

REL06-BP03 Senden von Benachrichtigungen (Verarbeitung und Benachrichtigung in Echtzeit)

Sorgen Sie dafür, dass bei wichtigen Ereignissen die entsprechenden Organisationen benachrichtigt werden.

Warnungen können an Amazon Simple Notification Service (Amazon SNS)-Themen gesendet und anschließend an eine beliebige Anzahl von Abonnenten weitergeleitet werden. Beispiel: Amazon SNS kann Warnungen an einen E-Mail-Alias weiterleiten, sodass das technische Personal reagieren kann.

Gängige Antimuster:

- Alarme werden mit einem zu niedrigen Schwellenwert konfiguriert, wodurch zu viele Benachrichtigungen gesendet werden.
- Keine Archivierung von Alarmen für künftige Untersuchungen.

Vorteile der Einführung dieser bewährten Methode: Durch Benachrichtigungen zu Ereignissen (auch solche, auf die reagiert werden kann und die sich automatisch lösen lassen) können Sie Ereignisse aufzeichnen und sie unter Umständen in Zukunft anders behandeln.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Führen Sie Verarbeitung und Alarme in Echtzeit aus. Sorgen Sie dafür, dass bei wichtigen Ereignissen die entsprechenden Organisationen benachrichtigt werden.
  - Amazon CloudWatch-Dashboards sind anpassbare Startseiten in der CloudWatch-Konsole für die Überwachung Ihrer Ressourcen in einer einzigen Ansicht, auch wenn sie über verschiedene Regionen verteilt sind.
    - Verwenden von Amazon CloudWatch-Dashboards
  - Lassen Sie einen Alarm auslösen, wenn die Metrik einen Grenzwert überschreitet.
    - Verwenden von Amazon CloudWatch-Alarmen

Ressourcen

Relevante Dokumente:

- Workshop zur Beobachtbarkeit

- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- Verwenden von Amazon CloudWatch-Alarmen
- Verwenden von Amazon CloudWatch-Dashboards
- Verwenden von Amazon CloudWatch-Metriken

REL06-BP04 Automatisieren von Antworten (Verarbeitung und Benachrichtigung in Echtzeit)

Automatisieren Sie bei Erkennung von Ereignissen die erforderlichen Maßnahmen, wie etwa den Austausch fehlerhafter Komponenten.

Alarme können AWS Auto Scaling-Ereignisse auslösen, sodass Cluster auf Bedarfsänderungen reagieren können. Warnungen können an Amazon Simple Queue Service (Amazon SQS) gesendet werden, das als Integrationspunkt für Ticketsysteme externer Anbieter dienen kann. Auch AWS Lambda kann Warnungen abonnieren und Benutzern so ein asynchrones serverloses Modell bereitstellen, das dynamisch auf Änderungen reagiert. AWS Config überwacht und zeichnet Ihre AWS-Ressourcenkonfigurationen kontinuierlich auf und kann AWS Systems Manager Automation auslösen, um Probleme zu beheben.

Amazon DevOps Guru kann Anwendungsressourcen automatisch auf anormale Verhaltensweisen überwachen und gezielte Empfehlungen für eine schnellere Problemidentifizierung und Fehlerbehebung bereitstellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Verwenden Sie Amazon DevOps Guru, um automatisierte Aktionen auszuführen. Amazon DevOps Guru kann Anwendungsressourcen automatisch auf anormale Verhaltensweisen überwachen und gezielte Empfehlungen für eine schnellere Problemidentifizierung und Fehlerbehebung bereitstellen.
  - Was ist Amazon DevOps Guru?
- Verwenden Sie AWS Systems Manager, um automatisierte Aktionen auszuführen. AWS Config überwacht und zeichnet Ihre AWS-Ressourcenkonfigurationen kontinuierlich auf und kann zur Behebung von Problemen AWS Systems Manager Automation auslösen.
  - AWS Systems Manager Automation

    - Erstellen und verwenden Sie Systems-Manager-Automation-Dokumente. Darin sind die Maßnahmen definiert, die Systems Manager in den verwalteten Instances und anderen AWS-Ressourcen durchführt, wenn ein Automatisierungslauf ausgeführt wird.
    - Arbeiten mit Automation-Dokumenten (Playbooks)
- Amazon CloudWatch sendet Änderungsereignisse für den Alarmstatus an Amazon EventBridge. Erstellen Sie EventBridge-Regeln zur Automatisierung von Antworten.
    - Erstellen einer EventBridge-Regel, die durch ein Ereignis aus einer AWS-Ressource ausgelöst wird
- Erstellen Sie einen Plan für die Automatisierung von Antworten und führen Sie ihn aus.
    - Inventarisieren Sie alle Verfahren zur Reaktion auf Warnungen. Sie müssen die Reaktionen auf Warnungen planen, bevor Sie die Aufgaben nach Rang einstufen.
    - Inventarisieren Sie alle Aufgaben mit spezifischen Maßnahmen, die durchgeführt werden müssen. Die meisten dieser Maßnahmen sind in Runbooks dokumentiert. Sie müssen außerdem über Playbooks für Warnungen zu unerwarteten Ereignissen verfügen.
    - Suchen Sie in den Runbooks und Playbooks nach allen automatisierbaren Maßnahmen. Wenn eine Maßnahme definiert werden kann, lässt sie sich in der Regel auch automatisieren.
    - Ordnen Sie zunächst die fehleranfälligen oder zeitaufwändigen Aktivitäten in einer Rangfolge ein. Es ist äußerst nützlich, Fehlerquellen zu entfernen und die Zeit bis zur Lösung zu verkürzen.
    - Erstellen Sie einen Plan, um die Automatisierung abzuschließen. Verwalten Sie einen aktiven Plan zur Automatisierung und aktualisieren Sie die Automatisierung.
    - Untersuchen Sie die manuellen Anforderungen auf Automatisierungsmöglichkeiten. Hinterfragen Sie Ihren manuellen Prozess und suchen Sie nach Automatisierungsmöglichkeiten.

Ressourcen

Ähnliche Dokumente:

- AWS Systems Manager Automation
- Erstellen einer EventBridge-Regel, die durch ein Ereignis aus einer AWS-Ressource ausgelöst wird
- Workshop zur Beobachtbarkeit
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- Was ist Amazon DevOps Guru?
- Arbeiten mit Automation-Dokumenten (Playbooks)

REL06-BP05 Analysen

Erfassen Sie Protokolldateien und Metrikverläufe und analysieren Sie diese, um allgemeine Trends zu erkennen und Workload-Einblicke zu erhalten.

Amazon CloudWatch Logs Insights unterstützt eine einfache und dennoch leistungsstarke Abfragesprache, mit der Sie Protokolldaten analysieren können. Amazon CloudWatch Logs unterstützt auch Abonnements, mit denen Daten nahtlos nach Amazon S3 fließen können, wo Sie sie nutzen oder Amazon Athena verwenden können, um die Daten abzufragen. Abfragen für eine große Auswahl von Formaten werden ebenfalls unterstütz. Unter Unterstützte SerDes- und Datenformate im Amazon Athena-Benutzerhandbuch finden Sie weitere Informationen dazu. Für die Analyse riesiger Protokolldateisätze können Sie einen Amazon EMR-Cluster ausführen, um Analysen im Petabyte-Bereich auszuführen.

Es gibt es eine Reihe von Werkzeugen von AWS-Partnern und externen Anbietern, die Aggregierung, Verarbeitung, Speicherung und Analyse ermöglichen. Dazu gehören u. a. die Tools New Relic, Splunk, Loggly, Logstash, CloudHealth und Nagios. Die Generierung außerhalb von System- und Anwendungsprotokollen weicht jedoch bei jedem Cloud-Anbieter und häufig sogar bei den einzelnen Services ab.

Ein häufig übersehener Teil des Überwachungsprozesses ist die Datenverwaltung. Sie müssen Aufbewahrungsanforderungen für die Überwachung von Daten definieren und anschließend entsprechende Lebenszyklusrichtlinien anwenden. Amazon S3 unterstützt die Lebenszyklusverwaltung auf der Ebene von S3-Buckets. Diese Lebenszyklusverwaltung kann auf unterschiedliche Weise auf verschiedene Pfade im Bucket angewendet werden. Gegen Ende des Lebenszyklus können Sie die Daten zur Langzeitspeicherung an Amazon S3 Glacier weiterleiten und nach Ablauf der Aufbewahrungsperiode die Speicherung beenden. Die S3 Intelligent-Tiering-Speicherklasse wurde entwickelt, um die Kosten zu optimieren. Daten werden automatisch in die kostengünstigste Zugriffsstufe verschoben, ohne Auswirkungen auf die Leistung oder höheren Betriebsaufwand.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Mit CloudWatch Logs Insights können Sie Protokolldaten in Amazon CloudWatch Logs interaktiv durchsuchen und analysieren.
  - Analysieren von Protokolldaten mit CloudWatch Logs Insights
  - Amazon CloudWatch Logs Insights-Beispielabfragen

- Verwenden Sie Amazon CloudWatch Logs, um Protokolle an Amazon S3 zu senden, wo Sie sie nutzen oder Amazon Athena verwenden können, um die Abfrage der Daten nutzen können.
  - Wie analysiere ich meine Amazon S3-Serverzugriffsprotokolle mit Athena?
    - Erstellen Sie eine S3-Lebenszyklusrichtlinie für Ihren Bucket mit den Serverzugriffsprotokollen. Konfigurieren Sie die Richtlinie so, dass Protokolldateien regelmäßig entfernt werden. Dies reduziert die Datenmenge, die Athena für die einzelnen Abfragen analysiert.
      - Wie erstelle ich eine Lebenszyklusrichtlinie für einen S3-Bucket?

Ressourcen

Relevante Dokumente:

- Amazon CloudWatch Logs Insights-Beispielabfragen
- Analysieren von Protokolldaten mit CloudWatch Logs Insights
- Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray
- Wie erstelle ich eine Lebenszyklusrichtlinie für einen S3-Bucket?
- Wie analysiere ich meine Amazon S3-Serverzugriffsprotokolle mit Athena?
- Workshop zur Beobachtbarkeit
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen

REL06-BP06 Regelmäßiges Durchführen von Prüfungen

Prüfen Sie regelmäßig, wie die Workload-Überwachung implementiert ist, und aktualisieren Sie sie auf Grundlage wichtiger Ereignisse und Änderungen.

Eine effektive Überwachung basiert auf wichtigen Geschäftsmetriken. Stellen Sie sicher, dass diese Metriken in Ihrer Workload berücksichtigt werden, wenn sich geschäftliche Prioritäten ändern.

Durch die Prüfung Ihrer Überwachung stellen Sie sicher, dass Sie wissen, wann eine Anwendung die eigenen Verfügbarkeitsziele erfüllt. Für die Durchführung von Ursachenanalysen ist es erforderlich, bei Ausfällen ermitteln zu können, was passiert ist. AWS bietet Services, mit denen Sie den Status Ihrer Services während eines Vorfalls nachverfolgen können.

- Amazon CloudWatch Logs: Sie können Ihre Protokolle in diesem Service speichern und die Inhalte überprüfen.

- Amazon CloudWatch Logs Insights: Ein vollständig verwalteter Service, mit dem Sie umfangreiche Protokolle innerhalb von Sekunden analysieren können. Es bietet Ihnen schnelle, interaktive Abfragen und Visualisierungen.
- AWS Config: Sie können sehen, welche AWS-Infrastruktur zu verschiedenen Zeitpunkten verwendet wurde.
- AWS CloudTrail: Mit diesem Service können Sie erkennen, welche AWS-APIs zu welchem Zeitpunkt und durch welchen Prinzipal aufgerufen wurden.

Bei AWS werden wöchentliche Meetings abgehalten, um die Produktionsleistung zu prüfen und Erkenntnisse mit anderen Teams zu teilen. Da es so viele Teams in AWS gibt, haben wir Das Rad entwickelt, um zufällig eine zu überprüfende Workload auszuwählen. Der Aufbau einer Struktur mit regelmäßigen Überprüfungen der betrieblichen Leistung und mit Wissensaustausch verbessert Ihre Fähigkeit, höhere Leistungen bei Ihren Betriebsteams zu erzielen.

Gängige Antimuster:

- Es werden nur Standardmetriken erfasst.
- Es wird eine Überwachungsstrategie festgelegt, aber nie überprüft.
- Bei Bereitstellung größerer Änderungen wird die Überwachung nicht erörtert.

Vorteile der Einführung dieser bewährten Methode: Durch die regelmäßige Prüfung der Überwachung können Sie mögliche Probleme vorhersehen, statt nur zu reagieren, wenn ein Problem tatsächlich auftritt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Erstellen Sie mehrere Dashboards für die Workload. Ein übergeordnetes Dashboard mit den wichtigsten Geschäftsmetriken ist unverzichtbar. Es sollte zudem die technischen Metriken enthalten, die Sie für den prognostizierten Zustand der Workload bei variabler Nutzung als die relevantesten eingestuft haben. Dashboards für verschiedene Anwendungsebenen und Abhängigkeiten, die untersucht werden können, sind ebenfalls empfehlenswert.
  - Verwenden von Amazon CloudWatch-Dashboards
- Planen und prüfen Sie die Workload-Dashboards regelmäßig. Führen Sie regelmäßige Untersuchungen der Dashboards durch. Was die Gründlichkeit der Untersuchungen angeht, sind unterschiedliche Intervalle denkbar.

- Spüren Sie Trends in den Metriken auf. Vergleichen Sie die Metrikwerte mit Werten aus der Vergangenheit, um Trends zu erkennen, die darauf hinweisen könnten, dass etwas untersucht werden muss. Beispiele hierfür: ansteigende Latenz, Nachlassen der primären Geschäftsfunktion und zunehmende Anzahl von Reaktionen auf Fehler.
- Spüren Sie Ausreißer/Anomalien in den Metriken auf. Ausreißer sind anhand von Durchschnitts- oder Mittelwerten oder Anomalien nicht unbedingt erkennbar. Sehen Sie sich die höchsten und niedrigsten Werte in einem bestimmten Zeitraum an und untersuchen Sie die Ursachen für die extremen Werte. Beseitigen Sie nach und nach die Ursachen und legen Sie dabei einen engeren Maßstab für die Definition von Extremwerten an. So können Sie die Beständigkeit der Workload-Leistung weiter erhöhen.
- Spüren Sie plötzliche Änderungen im Verhalten auf. Eine plötzliche Veränderung in der Menge oder Richtung einer Metrik kann auf eine Änderung in der Anwendung hindeuten. Sie kann aber auch ein Hinweis auf externe Faktoren sein, für deren Verfolgung Sie möglicherweise weitere Metriken hinzufügen müssen.

Ressourcen

Ähnliche Dokumente:

- Amazon CloudWatch Logs Insights-Beispielabfragen
- Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray
- Workshop zur Beobachtbarkeit
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- Verwenden von Amazon CloudWatch-Dashboards

REL06-BP07 Überwachen der gesamten Nachverfolgung von Anfragen im System

Verwenden Sie AWS X-Ray oder Tools von Drittanbietern, damit Entwickler verteilte Systeme einfacher analysieren und debuggen können, um Einblicke in die Leistung der Anwendungen und der zugrunde liegenden Services zu erhalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Überwachen Sie die gesamte Nachverfolgung von Anfragen im System. AWS X-Ray ist ein Service, der Daten zu Anfragen erfasst, die von Ihrer Anwendung verarbeitet werden. Zudem stellt er Tools bereit, mit denen Sie diese Daten anzeigen, filtern und auswerten können, um Probleme und Verbesserungsmöglichkeiten zu ermitteln. Sie können für jede nachverfolgte Anfrage an die Anwendung detaillierte Informationen zu Anfrage und Antwort anzeigen. Informationen zu Aufrufen, die Ihre Anwendung für nachgelagerte AWS-Ressourcen, Microservices, Datenbanken und Web-APIs ausführt, werden ebenfalls aufgeführt.
  - Was ist AWS X-Ray?
  - Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray

Ressourcen

Relevante Dokumente:

- Debugging mit Amazon CloudWatch Synthetics und AWS X-Ray
- Workshop zur Beobachtbarkeit
- Die Amazon Builders' Library: Verteilte Systeme instrumentieren, um betriebliche Transparenz zu erzielen
- Verwenden von Synthetic Monitoring
- Was ist AWS X-Ray?

ZUV 7 Wie lässt sich der Workload so gestalten, dass er sich an Bedarfsänderungen anpasst?

Eine skalierbare Workload bietet die Elastizität, Ressourcen automatisch entsprechend dem aktuellen Bedarf hinzuzufügen oder zu entfernen.

Bewährte Methoden

- REL07-BP01 Automatisches Abrufen und Skalieren von Ressourcen:
- REL07-BP02 Abrufen von Ressourcen bei Erkennen einer Beeinträchtigung einer Workload
- REL07-BP03 Abrufen von Ressourcen bei Feststellung, dass für eine Workload mehr Ressourcen benötigt werden
- REL07-BP04 Durchführen von Lasttests für die Workload

REL07-BP01 Automatisches Abrufen und Skalieren von Ressourcen:

Wenn Sie beeinträchtigte Ressourcen ersetzen oder Ihre Workload skalieren, können Sie den Prozess mithilfe von verwalteten AWS-Services wie Amazon S3 und AWS Auto Scaling automatisieren. Sie können die Skalierung auch mit Tools von Drittanbietern und AWS SDKs automatisieren.

Zu den verwalteten AWS-Services gehören Amazon S3, Amazon CloudFront, AWS Auto Scaling, AWS Lambda, Amazon DynamoDB, AWS Fargate und Amazon Route 53.

Mit AWS Auto Scaling können Sie beeinträchtigte Instances erkennen und ersetzen. Außerdem können Sie Skalierungspläne für Ressourcen erstellen, unter anderem für Amazon EC2 -Instances und Spot-Flotten, Amazon ECS -Aufgaben, Amazon DynamoDB -Tabellen und -Indizes sowie für Amazon Aurora -Replicas.

Bei der Skalierung von EC2-Instances sollten Sie mehrere Availability Zones nutzen (mindestens drei) und Kapazität hinzufügen oder entfernen, um ein Gleichgewicht über diese Availability Zones hinweg zu gewährleisten. ECS-Aufgaben oder Kubernetes-Pods (bei Verwendung von Amazon Elastic Kubernetes Service) sollten ebenfalls über mehrere Availability Zones hinweg verteilt werden.

Bei Verwendung von AWS Lambda werden Instances automatisch skaliert. Jedes Mal, wenn eine Ereignisbenachrichtigung für Ihre Funktion eingeht, ermittelt AWS Lambda schnell freie Kapazität innerhalb seiner Compute-Flotte und führt Ihren Code bis zur zugeteilten Gleichzeitigkeit aus. Sie müssen sicherstellen, dass die erforderliche Gleichzeitigkeit auf dem spezifischen Lambda und in Ihrem Service Quotas konfiguriert ist.

Amazon S3 wird automatisch skaliert, um hohe Anfrageraten zu verarbeiten. Beispielsweise kann Ihre Anwendung mindestens 3 500 PUT/COPY/POST/DELETE- oder 5 500 GET/HEAD-Anfragen pro Sekunde pro Präfix in einem Bucket erreichen. Für die Anzahl der Präfixe in einem Bucket gibt es keine Beschränkungen. Sie können Ihre Lese- oder Schreibleistung erhöhen, indem Sie Lesevorgänge parallelisieren. Wenn Sie beispielsweise 10 Präfixe in einem Amazon S3-Bucket erstellen, können Sie die Leseleistung auf 55 000 Leseanfragen pro Sekunde skalieren, um die Lesevorgänge zu parallelisieren.

Konfigurieren und nutzen Sie Amazon CloudFront oder ein vertrauenswürdiges Content Delivery Network (CDN). Ein CDN kann Antwortzeiten für Endbenutzer verkürzen und Anfragen für Inhalte aus dem Cache verarbeiten. Dadurch wird die Notwendigkeit zur Skalierung Ihrer Workload verringert.

Gängige Antimuster:

- Es werden Auto-Scaling-Gruppen für die automatisierte Reparatur implementiert, aber keine Elastizität.
- Als Reaktion auf stark ansteigenden Datenverkehr wird automatisch skaliert.
- Es werden hochgradig zustandsbehaftete Anwendungen bereitgestellt, wodurch die Option der Elastizität entfällt.

Vorteile der Einführung dieser bewährten Methode: Durch die Automatisierung entfällt die Gefahr manueller Fehler bei der Bereitstellung und Außerbetriebnahme von Ressourcen. Durch die Automatisierung entfällt das Risiko von Kostenüberschreitungen und Dienstverweigerungen (Denial of Service) aufgrund der langsamen Reaktion auf Bedürfnisse bezüglich der Bereitstellung oder Außerbetriebnahme von Ressourcen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Konfigurieren und nutzen Sie AWS Auto Scaling. Hiermit erfolgt eine Überwachung der Anwendungen und eine automatische Anpassung der Kapazität, um eine stabile, vorhersehbare Leistung zu möglichst niedrigen Kosten aufrechtzuerhalten. Mit AWS Auto Scaling lässt sich die Anwendungsskalierung für mehrere Ressourcen in mehreren Services einrichten.
  - Was ist AWS Auto Scaling?
    - Konfigurieren Sie Auto Scaling nach Bedarf in Ihren Amazon EC2-Instances und Spot-Flotten, Amazon ECS-Aufgaben, Amazon DynamoDB-Tabellen und -Indizes, Amazon Aurora-Replikaten und AWS Marketplace-Appliances.
      - Automatische Verwaltung der Durchsatzkapazität mit DynamoDB Auto Scaling
        - Legen Sie über die Service-API Alarme, Skalierungsrichtlinien sowie Aufwärm- und Abkühlungszeiten fest.
- Nutzen Sie Elastic Load Balancing. Load Balancer können die Last nach Pfaden oder Netzwerkkonnektivität verteilen.
  - Was ist Elastic Load Balancing?
    - Application Load Balancers kann Lasten nach Pfaden verteilen.
      - Was ist ein Application Load Balancer?
        - Konfigurieren Sie einen Application Load Balancer, um Datenverkehr basierend auf dem Pfad unter dem Domänennamen auf verschiedene Workloads zu verteilen.

        - Mit Application Load Balancers können Sie Lasten entsprechend dem AWS Auto Scaling verteilen, um den Bedarf zu verwalten.
            - Nutzen eines Load Balancer mit einer Auto-Scaling-Gruppe
    - Network Load Balancer können Lasten nach Verbindungen verteilen.
        - Was ist ein Network Load Balancer?
            - Konfigurieren Sie einen Network Load Balancer, um Datenverkehr auf verschiedene Workloads mit TCP zu verteilen oder einen konstanten Satz von IP-Adressen für die Workload festzulegen.
            - Mit Network Load Balancern können Sie Lasten entsprechend dem AWS Auto Scaling verteilen, um den Bedarf zu verwalten.
- Nutzen Sie einen hochverfügbaren DNS-Anbieter. Mithilfe von DNS-Namen können Ihre Benutzer anstelle von IP-Adressen Namen eingeben, um auf Ihre Workloads zuzugreifen. Diese Informationen werden innerhalb einer definierten Reichweite (meist weltweit) für Benutzer der Workload verteilt.
    - Nutzen Sie Amazon Route 53 oder einen vertrauenswürdigen DNS-Anbieter.
        - Was ist Amazon Route 53?
    - Mit Route 53 können Sie Ihre CloudFront-Verteilungen und Load Balancer verwalten.
        - Ermitteln Sie die zu verwaltenden Domänen und Subdomänen.
        - Erstellen Sie entsprechende Datensätze mithilfe von ALIAS- oder CNAME-Datensätzen.
            - Arbeiten mit Datensätzen
- Nutzen Sie das globale AWS-Netzwerk, um den Pfad von den Benutzern zu Ihren Anwendungen zu optimieren. AWS Global Accelerator überwacht kontinuierlich den Zustand der Anwendungsendpunkte und leitet den Datenverkehr in weniger als 30 Sekunden an fehlerfreie Endpunkte um.
    - Bei AWS Global Accelerator handelt es sich um einen Service, der die Verfügbarkeit und Leistung der Anwendungen bei lokalen oder weltweiten Benutzern verbessert. Er stellt statische IP-Adressen bereit, die als fester Einstiegspunkt zu den Anwendungsendpunkten in einer einzelnen oder in mehreren AWS-Regionen fungieren, z. B. Application Load Balancers, Network Load Balancer oder Amazon EC2-Instances.
        - Was ist AWS Global Accelerator?
- Konfigurieren und nutzen Sie Amazon CloudFront oder ein vertrauenswürdiges Content Delivery Network (CDN). Ein Content Delivery Network kann Antwortzeiten für Endbenutzer verkürzen und Anfragen für Inhalte verarbeiten, die zu einer unnötigen Skalierung Ihrer Workloads führen könnten.

- Was ist Amazon CloudFront?
  - Konfigurieren Sie Amazon CloudFront-Verteilungen für Ihre Workloads oder verwenden Sie das CDN eines Drittanbieters.
    - Sie können festlegen, dass die Workloads nur über CloudFront zugänglich sind. Legen Sie hierfür die IP-Bereiche für CloudFront in den Sicherheitsgruppen oder Zugriffsrichtlinien der Endpunkte fest.

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Ihnen beim Erstellen automatisierter Datenverarbeitungslösungen helfen können
- AWS Auto Scaling: Funktionsweise von Skalierungsplänen
- AWS Marketplace: Für Auto Scaling geeignete Produkte
- Automatische Verwaltung der Durchsatzkapazität mit DynamoDB Auto Scaling
- Nutzen eines Load Balancer mit einer Auto-Scaling-Gruppe
- Was ist AWS Global Accelerator?
- Was ist Amazon EC2 Auto Scaling?
- Was ist AWS Auto Scaling?
- Was ist Amazon CloudFront?
- Was ist Amazon Route 53?
- Was ist Elastic Load Balancing?
- Was ist ein Network Load Balancer?
- Was ist ein Application Load Balancer?
- Arbeiten mit Datensätzen

REL07-BP02 Abrufen von Ressourcen bei Erkennen einer Beeinträchtigung einer Workload

Skalieren Sie Ressourcen bei Bedarf reaktiv, wenn die Verfügbarkeit beeinträchtigt ist, um die Verfügbarkeit der Workload wiederherzustellen.

Sie müssen zunächst Zustandsprüfungen und die Kriterien für diese Prüfungen konfigurieren, um anzugeben, wann die Verfügbarkeit durch fehlende Ressourcen beeinträchtigt wird. Dann können

Sie entweder das entsprechende Personal informieren, die Ressource manuell zu skalieren, oder Sie lösen die Automatisierung aus, damit die Skalierung automatisch erfolgt.

Die Skalierung kann manuell an Ihre Workload angepasst werden, z. B. indem Sie die Anzahl der EC2-Instances in einer Auto-Scaling-Gruppe ändern oder den Durchsatz einer DynamoDB-Tabelle über die AWS Management Console oder AWS CLI. Nach Möglichkeit sollten Sie jedoch die Automatisierung verwenden (siehe Automatisiertes Abrufen oder Skalieren von Ressourcen).

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Rufen Sie Ressourcen bei der Erkennung einer Beeinträchtigung einer Workload ab. Skalieren Sie Ressourcen bei Bedarf reaktiv, wenn die Verfügbarkeit beeinträchtigt ist, um die Verfügbarkeit der Workload wiederherzustellen.
  - Nutzen Sie Skalierungspläne, bei denen es sich um die Kernkomponente von AWS Auto Scaling handelt, um eine Reihe von Anweisungen für das Skalieren Ihrer Ressourcen zu konfigurieren. Wenn Sie mit AWS CloudFormation arbeiten oder AWS-Ressourcen Tags hinzufügen, können Sie pro Anwendung Skalierungspläne für verschiedenen Ressourcengruppen einrichten. AWS Auto Scaling bietet Empfehlungen für an jede Ressource angepasste Skalierungsstrategien. Nachdem Sie einen Skalierungsplan erstellt haben, kombiniert AWS Auto Scaling zur Unterstützung Ihrer Skalierungsstrategie Methoden für die dynamische und prädiktive Skalierung.
    - AWS Auto Scaling: Funktionsweise von Skalierungsplänen
  - Mit Amazon EC2 Auto Scaling können Sie sicherstellen, dass Ihnen die richtige Anzahl von Amazon EC2-Instances zur Verfügung steht, um die Anwendungslast zu bewältigen. Sie erstellen Sammlungen von EC2-Instances, die als Auto-Scaling-Gruppen bezeichnet werden. In jeder Auto-Scaling-Gruppe können Sie die Mindestanzahl von Instances angeben. Amazon EC2 Auto Scaling stellt dann sicher, dass die Gruppe diese Größe nie unterschreitet. In jeder Auto-Scaling-Gruppe können Sie die maximale Anzahl von Instances angeben. Amazon EC2 Auto Scaling stellt dann sicher, dass die Gruppe diese Größe nie überschreitet.
    - Was ist Amazon EC2 Auto Scaling?
  - Bei der automatischen Skalierung von Amazon DynamoDB wird der AWS-Application-Auto-Scaling-Service genutzt, um die bereitgestellte Durchsatzkapazität in Ihrem Auftrag dynamisch an die Muster des tatsächlichen Datenverkehrs anzupassen. So kann eine Tabelle oder ein globaler Sekundärindex die bereitgestellte Lese- und Schreibkapazität erhöhen, um einen plötzlichen Anstieg des Datenverkehrs ohne Drosselung zu bewältigen.

- Automatische Verwaltung der Durchsatzkapazität mit DynamoDB Auto Scaling

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Ihnen beim Erstellen automatisierter Datenverarbeitungslösungen helfen können
- AWS Auto Scaling: Funktionsweise von Skalierungsplänen
- AWS Marketplace: Für Auto Scaling geeignete Produkte
- Automatische Verwaltung der Durchsatzkapazität mit DynamoDB Auto Scaling
- Was ist Amazon EC2 Auto Scaling?

REL07-BP03 Abrufen von Ressourcen bei Feststellung, dass für eine Workload mehr Ressourcen benötigt werden

Skalieren Sie Ressourcen proaktiv, um den Bedarf zu erfüllen und Auswirkungen auf die Verfügbarkeit zu vermeiden.

Viele AWS-Services werden automatisch dem Bedarf entsprechend skaliert. Wenn Sie Amazon EC2-Instances oder Amazon ECS-Cluster verwenden, können Sie die automatische Skalierung dieser Instances auf der Grundlage von Nutzungsmetriken konfigurieren, die dem Bedarf Ihrer Workload entsprechen. Für Amazon EC2 können Sie die durchschnittliche CPU-Auslastung, die Anzahl der Load Balancer-Anfragen oder die Netzwerkbandbreite verwenden, um EC2-Instances zu skalieren. Für Amazon ECS können Sie die durchschnittliche CPU-Auslastung, die Anzahl der Load-Balancer-Anfragen und die Speichernutzung verwenden, um ECS-Aufgaben auf- oder abzuskalieren. Mit Target Auto Scaling auf AWS fungiert der Autoscaler wie ein Haushaltsthermostat, der Ressourcen hinzufügt oder entfernt, um den von Ihnen angegebenen Zielwert (z. B. 70 % CPU-Auslastung) beizubehalten.

AWS Auto Scaling kann auch Predictive Auto Scalingdurchführen. Dabei wird Machine Learning verwendet, um die bisherige Workload jeder Ressource zu analysieren und regelmäßig die zukünftige Last für die nächsten zwei Tage zu prognostizieren.

Das Gesetz von Little hilft beim Berechnen der Anzahl von Compute-Instances, die Sie benötigen (EC2-Instances, gleichzeitige Lambda-Funktionen usw.).

$L = \lambda W$

L = Anzahl der Instances (oder mittlere Gleichzeitigkeit im System)

λ = mittlere Rate des Eingangs von Anfragen (Anfrage/Sekunde)

W = mittlere Zeit, die jede Anfrage im System verbringt (Sekunden)

Wenn beispielsweise bei 100 RPS die Verarbeitung jeder Anfrage 0,5 Sekunden dauert, benötigen Sie 50 Instances, um mit dem Bedarf Schritt zu halten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Rufen Sie Ressourcen ab, wenn Sie feststellen, dass für eine Workload mehr Ressourcen benötigt werden. Skalieren Sie Ressourcen proaktiv, um den Bedarf zu erfüllen und Auswirkungen auf die Verfügbarkeit zu vermeiden.
  - Berechnen Sie, wie viele Rechenressourcen Sie benötigen (Gleichzeitigkeit der Datenverarbeitung), um eine bestimmte Anfragerate zu verarbeiten.
    - Berichte über das Gesetz von Little
  - Wenn Sie über ein Verlaufsmuster für die Nutzung verfügen, richten Sie die geplante Skalierung für Amazon EC2 ein.
    - Geplante Skalierung für Amazon EC2 Auto Scaling
  - Verwenden Sie die vorausschauende Skalierung von AWS.
    - Prädiktive Skalierung für EC2, unterstützt von Machine Learning

Ressourcen

Relevante Dokumente:

- AWS Auto Scaling: Funktionsweise von Skalierungsplänen
- AWS Marketplace: Für Auto Scaling geeignete Produkte
- Automatische Verwaltung der Durchsatzkapazität mit DynamoDB Auto Scaling
- Prädiktive Skalierung für EC2, unterstützt von Machine Learning
- Geplante Skalierung für Amazon EC2 Auto Scaling
- Berichte über das Gesetz von Little
- Was ist Amazon EC2 Auto Scaling?

REL07-BP04 Durchführen von Lasttests für die Workload

Messen Sie anhand von Lasttests, ob die Skalierung den Workload-Anforderungen gerecht wird.

Es ist wichtig, regelmäßige Lasttests durchzuführen. Mit diesen Tests können Sie die Belastungsgrenze Ihrer Workload ermitteln und deren Leistung prüfen. AWS erleichtert das Einrichten temporärer Testumgebungen, die den Umfang Ihrer Produktions-Workload modellieren. Sie können in der Cloud bei Bedarf eine Testumgebung in Produktionsgröße einrichten, Ihre Tests abschließen und die Ressourcen dann wieder stilllegen. Weil Sie für die Testumgebung nur dann zahlen, wenn sie genutzt wird, können Sie Ihre Live-Umgebung zu einem Bruchteil der Kosten testen, die Sie an einem lokalen Standort hätten.

Lasttests in der Produktion sollten auch im Rahmen von Ernstfallübungen durchgeführt werden, bei denen das Produktionssystem in einem Zeitraum mit geringer Kundennutzung stark belastet wird. Alle Mitarbeiter sollten an dieser Übung beteiligt sein, die Ergebnisse gemeinsam interpretieren und auftretende Probleme beheben.

Gängige Antimuster:

- Es werden Lasttests für Bereitstellungen durchgeführt, die nicht mit der Konfiguration der Produktionsumgebung übereinstimmen.
- Lasttests werden nur für einzelne Teile, nicht aber für die gesamte Workload durchgeführt.
- Es werden Lasttests mit einer Teilmenge von Anfragen durchgeführt, aber nicht mit einer repräsentativen Gruppe tatsächlicher Anfragen.
- Es werden Lasttests mit einem kleinen Sicherheitsfaktor durchgeführt, der über der erwarteten Last liegt.

Vorteile der Einführung dieser bewährten Methode: Sie wissen, welche Komponenten in der Architektur unter Last ausfallen, und können die zu überwachenden Metriken festlegen, die rechtzeitig auf die Annäherung an die Belastungsgrenze hinweisen, damit Sie das Problem beheben und entsprechende Auswirkungen vermeiden können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Bestimmen Sie anhand von Lasttests, welcher Aspekt der Workload angeben soll, dass Kapazität hinzugefügt oder entfernt werden muss. Bei Lasttests sollte ein repräsentativer Datenverkehr zum Einsatz kommen, der dem in der Produktion ähnelt. Erhöhen Sie unter Beobachtung der

instrumentierten Metriken die Last, um zu bestimmen, welche Metrik angibt, wann Ressourcen hinzugefügt oder entfernt werden müssen.

- Verteilte Lasttests auf AWS: Simulation Tausender verbundener Benutzer
  - Ermitteln Sie die Zusammensetzung von Anfragen. Möglicherweise haben Sie unterschiedliche Zusammensetzungen von Anfragen. Daher sollten Sie sich bei der Ermittlung der Zusammensetzung des Datenverkehrs verschiedene Zeiträume ansehen.
  - Implementieren Sie einen Lasttreiber. Zum Implementieren eines Lasttreibers können Sie Software mit eigenem Code, Open-Source-Software oder kommerzielle Software verwenden.
  - Führen Sie Lasttests zunächst mit geringer Kapazität durch. Schon bei der Erhöhung der Last für eine Einheit mit geringerer Kapazität, etwa einer einzelnen Instance oder einem einzelnen Container, stellen Sie unmittelbare Auswirkungen fest.
  - Führen Sie Lasttests mit größerer Kapazität durch. Bei einer verteilten Last sehen die Auswirkungen anders aus. Daher müssen Sie bei Tests Bedingungen herstellen, die der Produktionsumgebung möglichst nahekommen.

Ressourcen

Relevante Dokumente:

- Verteilte Lasttests auf AWS: Simulation Tausender verbundener Benutzer

ZUV 8 Wie implementieren Sie Änderungen?

Kontrollierte Änderungen sind erforderlich, um neue Funktionen bereitzustellen und um sicherzustellen, dass die Workloads und die Betriebsumgebung bekannte Software ausführen und auf vorhersagbare Weise durch Patches aktualisiert oder ersetzt werden können. Wenn diese Änderungen nicht kontrolliert stattfinden, ist es schwierig, ihre Auswirkungen vorherzusagen oder daraus entstehende Probleme zu beheben.

Bewährte Methoden

- REL08-BP01 Verwenden von Runbooks für Standardaktivitäten wie die Bereitstellung
- REL08-BP02 Integrieren von Funktionstests in die Bereitstellung
- REL08-BP03 Integrieren von Ausfallsicherheitstests in die Bereitstellung
- REL08-BP04 Bereitstellung mit einer unveränderlichen Infrastruktur
- REL08-BP05 Automatisieren von Änderungen

REL08-BP01 Verwenden von Runbooks für Standardaktivitäten wie die Bereitstellung

Runbooks sind vordefinierte Verfahren, die ein bestimmtes Ergebnis verfolgen. Verwenden Sie Runbooks, um Standardaktivitäten manuell oder automatisch durchzuführen. Beispiele für solche Aktivitäten sind etwa die Bereitstellung und das Patchen einer Workload oder das Vornehmen von DNS-Änderungen.

Sie können z. B. Prozesse einrichten, um bei Bereitstellungen die Rollback-Sicherheit zu gewährleisten. Wenn Sie eine Bereitstellung ohne Unterbrechung für Ihre Kunden zurücksetzen können, steigert das die Zuverlässigkeit Ihres Service.

Für Runbook-Verfahren sollten Sie mit einem gültigen, effektiven manuellen Prozess beginnen, diesen in Code implementieren und ggf. die automatische Ausführung auslösen.

Selbst bei anspruchsvollen Workloads mit umfassender Automatisierung sind Runbooks nützlich, um Ernstfallübungen auszuführen oder strenge Berichterstellungs- und Auditing-Anforderungen zu erfüllen.

Playbooks werden als Reaktion auf bestimmte Vorfälle verwendet und mit Runbooks sollen bestimmte Ergebnisse erzielt werden. Häufig werden Runbooks für Routineaktivitäten genutzt, während Playbooks für die Reaktion auf außerplanmäßige Ereignisse verwendet werden.

Gängige Antimuster:

- Durchführen ungeplanter Änderungen an der Konfiguration in der Produktion.
- Überspringen von Schritten in Ihrem Plan, um schneller bereitzustellen, was dann jedoch zum Fehlschlagen der Bereitstellung führt.
- Vornehmen von Änderungen, ohne die Umkehrung der Änderung zu testen.

Vorteile der Einführung dieser bewährten Methode: Die effektive Änderungsplanung erhöht Ihre Fähigkeit, die Änderung erfolgreich auszuführen, da Sie sich über alle betroffenen Systeme bewusst sind. Die Validierung Ihrer Änderungen in Testumgebungen erhöht Ihre Sicherheit.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Unterstützen Sie konsistente und schnelle Reaktionen auf gut bekannte Ereignisse, indem Sie Verfahren in Runbooks dokumentieren.
  - AWS-Well-Architected-Framework: Konzepte: Runbook

- Verwenden Sie zur Definition Ihrer Infrastruktur den Grundsatz „Infrastructure as Code“. Wenn Sie Ihre Infrastruktur mit AWS CloudFormation oder dem vertrauenswürdigen Tool eines Drittanbieters definieren, können Sie Änderungen mithilfe einer Versionskontrollsoftware versionieren und nachverfolgen.
  - Nutzen Sie zur Definition Ihrer Infrastruktur AWS CloudFormation (oder das vertrauenswürdige Tool eines Drittanbieters).
    - Was ist AWS CloudFormation?
  - Erstellen Sie unter Anwendung guter Grundsätze für das Softwaredesign Vorlagen, die getrennt und entkoppelt sind.
    - Ermitteln Sie die für die Implementierung erforderlichen Berechtigungen, Vorlagen und zuständigen Parteien.
      - Zugriffssteuerung mit AWS Identity and Access Management
    - Verwenden Sie zur Versionskontrolle eine Quellkontrolle wie AWS CodeCommit oder das vertrauenswürdige Tool eines Drittanbieters.
      - Was ist AWS CodeCommit?

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie beim Erstellen automatisierter Bereitstellungslösungen unterstützen können
- AWS Marketplace: Produkte zur Automatisierung Ihrer Bereitstellungen
- AWS-Well-Architected-Framework: Konzepte: Runbook
- Was ist AWS CloudFormation?
- Was ist AWS CodeCommit?

Ähnliche Beispiele:

- Automating operations with Playbooks and Runbooks (Vorgänge mit Playbooks und Runbooks automatisieren)

REL08-BP02 Integrieren von Funktionstests in die Bereitstellung

Funktionstests werden im Rahmen der automatisierten Bereitstellung ausgeführt. Wenn die Erfolgskriterien nicht erfüllt sind, wird die Pipeline angehalten oder rückgängig gemacht.

Diese Tests werden in einer Vorproduktionsumgebung ausgeführt, die vor der Produktion in der Pipeline bereitgestellt wird. Idealerweise erfolgt dies im Rahmen einer Bereitstellungspipeline.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Integrieren Sie Funktionstests in Ihre Bereitstellung. Funktionstests werden im Rahmen der automatisierten Bereitstellung ausgeführt. Wenn die Erfolgskriterien nicht erfüllt sind, wird die Pipeline angehalten oder rückgängig gemacht.
  - Rufen Sie AWS CodeBuild während der „Testaktion“ Ihrer in AWS CodePipeline modellierten Software-Release-Pipelines auf. Mit dieser Funktion können Sie ganz einfach verschiedene Tests für Ihren Code ausführen, z. B. Komponententests, statische Code-Analysen und Integrationstests.
    - AWS CodePipeline fügt Unterstützung für Komponententests und angepasste Integrationstests mit AWS CodeBuild hinzu.
  - Verwenden Sie AWS Marketplace-Lösungen, um als Teil Ihrer Softwarebereitstellungs-Pipeline automatisierte Tests auszuführen.
    - Automatisierung von Softwaretests

Ressourcen

Relevante Dokumente:

- AWS CodePipeline fügt Unterstützung für Komponententests und angepasste Integrationstests mit AWS CodeBuild hinzu.
- Automatisierung von Softwaretests
- Was ist AWS CodePipeline?

REL08-BP03 Integrieren von Ausfallsicherheitstests in die Bereitstellung

Ausfallsicherheitstests (unter Anwendung der Grundlagen des Chaos-Engineering) werden als Teil der automatisierten Bereitstellungs-Pipeline in einer Vorproduktionsumgebung ausgeführt.

Diese Tests werden in einer Vorproduktionsumgebung in der Pipeline bereitgestellt und ausgeführt. Sie sollten auch in der Produktion ausgeführt werden, aber im Rahmen von Ernstfallübungen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Integrieren Sie Ausfallsicherheitstests in Ihre Bereitstellung. Verwenden Sie Chaos-Engineering, die Disziplin des Experimentierens an einer Workload, um Vertrauen in die Fähigkeit der Workload aufzubauen, turbulente Bedingungen in der Produktion zu bewältigen.
  - Ausfallsicherheitstests schleusen Fehler oder die Verschlechterung von Ressourcen ein, um zu bewerten, ob Ihre Workload mit der vorgesehenen Resilienz reagiert.
    - Well-Architected Lab: Level 300: Testen auf Resilienz von EC2 RDS and S3
  - Diese Tests können regelmäßig für automatisierte Bereitstellungs-Pipelines in Vorproduktionsumgebungen ausgeführt werden.
  - Sie sollten auch in der Produktion als Teil geplanter Ernstfallübungen ausgeführt werden.
  - Entwickeln Sie unter Verwendung von Grundsätzen des Chaos-Engineering Hypothesen zur Leistung Ihrer Workload bei verschiedenen Beeinträchtigungen. Testen Sie dann Ihre Hypothesen mithilfe von Resilienztests.
    - Grundlagen des Chaos-Engineering

Ressourcen

Relevante Dokumente:

- Grundlagen des Chaos-Engineering
- Was ist AWS Fault Injection Simulator?

Ähnliche Beispiele:

- Well-Architected Lab: Level 300: Testen auf Resilienz von EC2 RDS and S3

REL08-BP04 Bereitstellung mit einer unveränderlichen Infrastruktur

Eine unveränderliche Infrastruktur sieht vor, dass Updates, Sicherheits-Patches oder Konfigurationsänderungen nicht direkt in Produktions-Workloads durchgeführt werden. Wenn eine Änderung erforderlich ist, wird die Architektur auf einer neuen Infrastruktur eingerichtet und für die Produktion bereitgestellt.

Die häufigste Implementierung des unveränderlichen Infrastrukturparadigmas ist der unveränderlicher Server.. Wenn ein Update erforderlich ist oder Fehler behoben werden müssen, werden neue

Server bereitgestellt, statt die bereits verwendeten Server zu aktualisieren. Statt sich also über SSH beim Server anzumelden und die Softwareversion zu aktualisieren, beginnt jede Änderung in der Anwendung mit einer Push-Verteilung der Software an das Code-Repository, z. B. git push. Da Änderungen in einer unveränderlichen Infrastruktur nicht zulässig sind, ist Ihnen der Status des bereitgestellten Systems immer bekannt. Unveränderliche Infrastrukturen sind grundsätzlich konsistenter, zuverlässiger und berechenbarer und vereinfachen viele Aspekte der Softwareentwicklung und des Betriebs.

Verwenden Sie eine Canary- oder Blue/Green-Bereitstellung, wenn Sie Anwendungen in unveränderlichen Infrastrukturen bereitstellen.

Canary-Bereitstellung wird eine kleine Anzahl Ihrer Kunden zur neuen Version weitergeleitet, die in der Regel auf einer einzelnen Service-Instance (dem Canary) ausgeführt wird. Anschließend überprüfen Sie sämtliche Verhaltensänderungen oder Fehler, die generiert werden. Sie können Datenverkehr aus der Canary-Umgebung entfernen, wenn kritische Probleme auftreten, und die Benutzer auf die vorherige Version zurücksetzen. Wenn die Bereitstellung erfolgreich verläuft, können Sie das gewünschte Tempo beibehalten und die Änderungen auf Fehler überwachen, bis der Bereitstellungsvorgang vollständig abgeschlossen ist. Sie können AWS CodeDeploy mit einer Bereitstellungskonfiguration konfigurieren, die eine Canary-Bereitstellung ermöglicht.

Blue/Green-Bereitstellungen verhalten sich ähnlich wie Canary-Bereitstellungen. Allerdings wird die vollständige Flotte der Anwendung parallel bereitgestellt. Sie können Ihre Bereitstellungen über die zwei Stacks (blau und grün) alternieren. Auch hier können Sie Datenverkehr an die neue Version senden und einen Failback auf die alte Version durchführen, wenn bei der Bereitstellung Probleme auftreten. Normalerweise wird der gesamte Datenverkehr auf einmal umgeschaltet. Sie können Ihren Datenverkehr aber auch auf die Versionen verteilen, um die Einführung der neuen Version mithilfe der gewichteten DNS-Routing-Funktionen von Amazon Route 53 durchzuführen. Sie können AWS CodeDeploy und AWS Elastic Beanstalk mit einer Bereitstellungskonfiguration konfigurieren, die eine Blau/Grün-Bereitstellung ermöglicht.

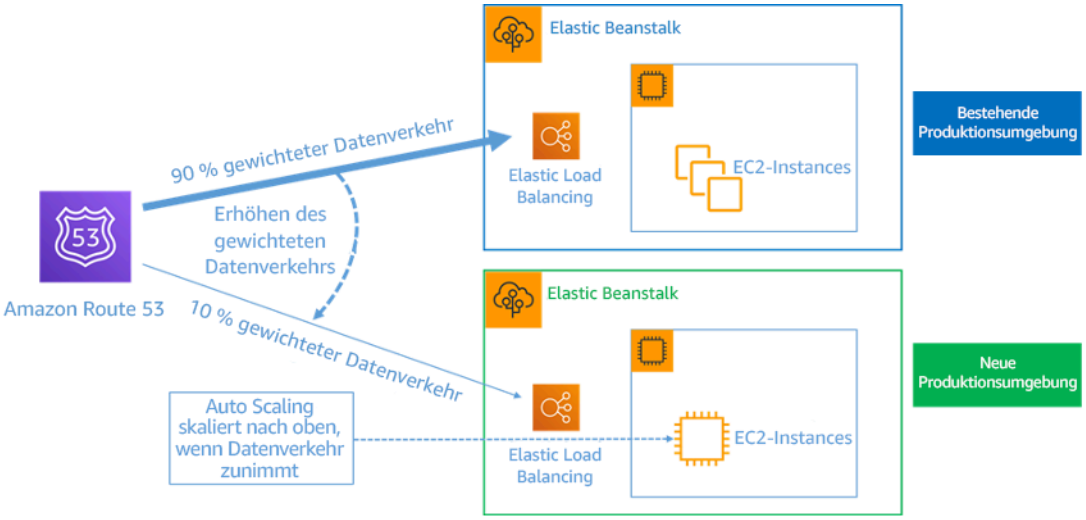


Abbildung 8: Blau/Grün-Bereitstellung mit AWS Elastic Beanstalk und Amazon Route 53

Vorteile einer unveränderlichen Infrastruktur:

- Reduzierung der Konfigurationsabweichungen: Wenn Sie Server häufig mit bekannten, versionsgesteuerten und Basiskonfigurationen austauschen, wird die Infrastruktur in einen bekannten Zustand zurückgesetzt. Dadurch werden Konfigurationsabweichungen vermieden.
- Vereinfachte Bereitstellungen: Bereitstellungen werden vereinfacht, da sie keine Upgrades unterstützen müssen. Upgrades sind einfach neue Bereitstellungen.
- Zuverlässige atomare Bereitstellungen: Bereitstellungen werden entweder erfolgreich abgeschlossen oder es werden keine Änderungen vorgenommen. Das erhöht die Zuverlässigkeit des Bereitstellungsprozesses.
- Sicherere Bereitstellungen mit schnellen Rollback- und Wiederherstellungsprozessen: Bereitstellungen sind sicherer, da die vorherige funktionierende Version nicht geändert wird. Sie können einen Rollback zur vorherigen Version durchführen, wenn Fehler erkannt werden.
- Konsistente Test- und Debugging-Umgebungen: Da alle Server dasselbe Image verwenden, gibt es keine Unterschiede zwischen Umgebungen. Ein Build wird in mehreren Umgebungen bereitgestellt. So werden außerdem inkonsistente Umgebungen verhindert und das Testen und Debuggen wird vereinfacht.
- Erhöhte Skalierbarkeit: Da Server ein Basis-Image verwenden, konsistent und wiederholbar sind, ist die automatische Skalierung sehr einfach.
- Vereinfachte Toolkette: Die Toolkette ist vereinfacht, da Sie für die Verwaltung von Produktionssoftware-Upgrades keine Konfigurationsmanagement-Tools mehr benötigen. Auf

Servern sind keine zusätzlichen Tools oder Agents installiert. Änderungen werden am Basis-Image vorgenommen, getestet und bereitgestellt.

- Erhöhte Sicherheit: Wenn Sie alle Änderungen an den Servern ablehnen, können Sie SSH auf Instances deaktivieren und Schlüssel entfernen. Dadurch wird der Angriffsvektor reduziert und die Sicherheitslage Ihres Unternehmens verbessert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Stellen Sie eine unveränderliche Infrastruktur bereit. Eine unveränderliche Infrastruktur ist ein Modell, bei dem Updates, Sicherheits-Patches oder Konfigurationsänderungen nicht direkt in Produktions-Workloads durchgeführt werden. Wenn eine Änderung erforderlich ist, wird eine neue Version der Architektur entwickelt und in der Produktion bereitgestellt.
  - Übersicht über eine Blue/Green-Bereitstellung
  - Schrittweise Bereitstellung von Serverless-Anwendungen
  - Unveränderliche Infrastruktur: Zuverlässigkeit, Konsistenz und Vertrauen durch Unveränderlichkeit
  - Canary-Release

Ressourcen

Relevante Dokumente:

- Canary-Release
- Schrittweise Bereitstellung von Serverless-Anwendungen
- Unveränderliche Infrastruktur: Zuverlässigkeit, Konsistenz und Vertrauen durch Unveränderlichkeit
- Übersicht über eine Blue/Green-Bereitstellung
- Die Amazon Builders' Library: Rollback-Sicherheit bei Bereitstellungen gewährleisten

REL08-BP05 Automatisieren von Änderungen

Bereitstellungen und Patches werden automatisiert, um negative Auswirkungen zu vermeiden.

Änderungen an Produktionssystemen gehören in vielen Unternehmen zu den größten Risikofaktoren. Neben den geschäftlichen Problemen, die durch die Software behoben werden, betrachten wir

Bereitstellungen als vorrangiges Problem, das es zu lösen gilt. Heutzutage bedeutet das, wenn immer möglich und sinnvoll, Vorgänge zu automatisieren. Dazu gehören Tests und die Bereitstellung von Änderungen, das Hinzufügen oder Entfernen von Kapazität und das Migrieren von Daten. Mit AWS CodePipeline können Sie die erforderlichen Schritte für die Freigabe Ihrer Workload verwalten. Dies umfasst einen Bereitstellungsstatus in AWS CodeDeploy, um die Bereitstellung von Anwendungscode für Amazon EC2-Instances, On-Premise-Instances, serverlose Lambda-Funktionen oder Amazon ECS-Services zu automatisieren.

> Empfehlung
>
> Obwohl die gängige Meinung vorherrscht, dass es sinnvoll ist, Menschen bei den komplexesten betrieblichen Abläufen in die Vorgänge zu integrieren, empfehlen wir, die komplexesten Abläufe aus genau diesem Grund zu automatisieren.

Gängige Antimuster:

- Manuelles Durchführen von Änderungen.
- Überspringen von Schritten in Ihrer Automatisierung durch Notfallarbeitsabläufe.
- Entspricht nicht Ihren Plänen.

Vorteile der Einführung dieser bewährten Methode: Die Verwendung der Automatisierung zum Bereitstellen aller Änderungen verringert das Risiko menschlicher Fehler und ermöglicht die Durchführung von Tests vor Produktionsänderung, um sicherzustellen, dass Ihre Pläne vollständig sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren Sie Ihre Bereitstellungs-Pipeline. Mit Bereitstellungs-Pipelines können Sie Tests und die Entdeckung von Anomalien automatisieren und die Pipeline an einem bestimmten Schritt vor der Bereitstellung in der Produktion anhalten oder eine Änderung automatisch zurückführen.
  - Die Amazon Builders' Library: Rollback-Sicherheit bei Bereitstellungen gewährleisten
  - Die Amazon Builders' Library: Schneller mit kontinuierlicher Bereitstellung
    - Verwenden Sie AWS CodePipeline oder das vertrauenswürdige Produkt eines Drittanbieters), um Ihre Pipelines zu definieren und auszuführen.

- Legen Sie fest, dass die Pipeline startet, sobald in Ihrem Code-Repository eine Änderung festgeschrieben wird.
  - Was ist AWS CodePipeline?
- Verwenden Sie Amazon Simple Notification Service (Amazon SNS) und Amazon Simple Email Service (Amazon SES), um Benachrichtigungen bezüglich Pipeline-Problemen zu senden, oder integrieren Sie diese in ein Team-Chat-Tool wie Amazon Chime.
  - Was ist Amazon Simple Notification Service?
  - Was ist Amazon SES?
  - Was ist Amazon Chime?
  - Automatisieren Sie Chat-Nachrichten mit Webhooks.

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie beim Erstellen automatisierter Bereitstellungslösungen unterstützen können
- AWS Marketplace: Produkte zur Automatisierung Ihrer Bereitstellungen
- Automatisieren Sie Chat-Nachrichten mit Webhooks.
- Die Amazon Builders' Library: Rollback-Sicherheit bei Bereitstellungen gewährleisten
- Die Amazon Builders' Library: Schneller mit kontinuierlicher Bereitstellung
- Was ist AWS CodePipeline?
- Was ist CodeDeploy?
- AWS Systems Manager Patch Manager
- Was ist Amazon SES?
- Was ist Amazon Simple Notification Service?

Relevante Videos:

- AWS Summit 2019: CI/CD on AWS (AWS Summit 2019: CI/CD auf AWS)

Fehlerverwaltung

Fragen

- ZUV 9 Was ist bei der Sicherung von Daten zu beachten?
- ZUV 10 Wie schützen Sie Ihren Workload mithilfe der Fehlerisolierung?
- ZUV 11 Wie lassen sich Workloads so gestalten, dass sie Komponentenausfälle verkraften?
- ZUV 12 Wie lässt sich die Zuverlässigkeit testen?
- ZUV 13 Was ist bei der Planung der Notfallwiederherstellung zu beachten?

ZUV 9 Was ist bei der Sicherung von Daten zu beachten?

Sichern Sie Daten, Anwendungen und Konfigurationen, um die Anforderungen im Hinblick auf das Recovery Time Objective (RTO, Wiederherstellungsdauer) und das Recovery Point Objective (RPO, Wiederherstellungszeitpunkt) zu erfüllen.

Bewährte Methoden

- REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen
- REL09-BP02 Schützen und Verschlüsseln von Backups
- REL09-BP03 Automatische Daten-Backups
- REL09-BP04 Verifizieren der Sicherungsintegrität und -verfahren durch regelmäßiges Wiederherstellen der Daten

REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen

Alle AWS-Datenspeicher bieten Backup-Möglichkeiten. Services wie Amazon RDS und Amazon DynamoDB unterstützen zusätzlich ein automatisiertes Backup, das eine zeitpunktbezogene Wiederherstellung (PITR) ermöglicht. So können Sie Backups zu einem beliebigen Zeitpunkt bis zu fünf Minuten oder weniger vor dem aktuellen Zeitpunkt wiederherstellen. Viele AWS-Services bieten die Möglichkeit, Backups in eine andere AWS-Region zu kopieren. AWS Backup ist ein Tool, mit dem Sie Datensicherungsprozesse über verschiedene AWS-Services hinweg zentralisieren und automatisieren können.

Amazon S3 kann als Backup-Ziel für selbstverwaltete und AWS-verwaltete Datenquellen verwendet werden. AWS-Services wie Amazon EBS, Amazon RDS und Amazon DynamoDB bieten integrierte

Möglichkeiten zur Backup-Erstellung. Sicherungssoftware von Drittanbietern kann ebenfalls eingesetzt werden.

On-Premises-Daten können in AWS Cloud unter Verwendung von AWS Storage Gateway oder AWS DataSyncgesichert werden. Amazon S3-Buckets können genutzt werden, um diese Daten in AWS zu speichern. Amazon S3 bietet mehrere Speicherebenen wie Amazon S3 Glacier oder S3 Glacier Deep Archive, um die Datenspeicherkosten zu senken.

Möglicherweise können Sie Ihre Datenwiederherstellungs-Anforderungen erfüllen, indem Sie Daten aus anderen Quellen reproduzieren. So könnten beispielsweise Amazon ElastiCache-Replikatknoten oder RDS-Lesereplikate für das Reproduzieren von Daten verwendet werden, wenn die primären Daten verloren gegangen sind. Wenn solche Quellen für das Erreichen Ihrer Recovery Time Objective (RTO) und Recovery Point Objective (RPO)genutzt werden können, benötigen Sie möglicherweise kein Backup. Weiteres Beispiel: Wenn Sie mit Amazon EMR arbeiten, ist es möglicherweise nicht erforderlich, Ihren HDFS-Datenspeicher zu sichern, solange Sie die Daten in EMR von S3 reproduzieren können.

Bei der Auswahl einer Backup-Strategie sollten Sie die für die Datenwiederherstellung benötigte Zeit berücksichtigen. Diese hängt von der Art des Backups (im Falle einer Backup-Strategie) oder von der Komplexität des Datenreproduktions-Mechanismus ab. Die benötigte Zeit sollte im RTO für die Workload enthalten sein.

Gewünschtes Ergebnis:

Datenquellen wurden basierend auf Kritikalität identifiziert und klassifiziert. Anschließend legen Sie eine auf dem RPO basierende Strategie für die Datenwiederherstellung fest. Diese Strategie involviert entweder die Sicherung dieser Datenquellen oder die Möglichkeit, Daten aus anderen Quellen zu reproduzieren. Im Falle eines Datenverlusts ermöglicht die implementierte Strategie die Wiederherstellung oder Reproduktion von Daten innerhalb der definierten RPO und RTO.

„Cloud-Reife“-Phase: Foundational

Gängige Antimuster:

- Nicht alle Datenquellen für die Workload und deren Kritikalität sind bekannt.
- Es erfolgen keine Backups kritischer Datenquellen.
- Es erfolgen nur Backups von manchen Datenquellen ohne die Verwendung von Kritikalität als Kriterium.
- Es wurde kein RPO definiert oder die Backup-Häufigkeit kann den RPO nicht erfüllen.

- Es erfolgt keine Bewertung, ob ein Backup erforderlich ist oder ob Daten aus anderen Quellen reproduziert werden können.

Vorteile der Einführung dieser bewährten Methode: Durch das Identifizieren der Orte, wo Backups notwendig sind und das Implementieren eines Mechanismus zur Erstellung von Backups bzw. die Möglichkeit, die Daten aus externen Quellen zu reproduzieren, ist es einfacher, Daten während eines Ausfalls wiederherzustellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Sie sollten die Backup-Funktionen der von der Workload genutzten AWS-Services und -Ressourcen verstehen und nutzen. Die meisten AWS-Services stellen Funktionen für Backups von Workloaddaten bereit.

Implementierungsschritte:

1. Identifizieren Sie alle Datenquellen für die Workload.. Daten können in einer Reihe von Ressourcen gespeichert werden, wie z. B. Datenbanken, Volumes, Dateisystemen, Protokollierungssystemenund Objektspeicher. Im Abschnitt „Ressourcen“ finden Sie Relevante Dokumente zu verschiedenen AWS-Services, in denen Daten gespeichert werden, sowie die Backup-Funktionen dieser Services.
2. Klassifizieren Sie Datenquellen basierend auf Kritikalität.. Unterschiedliche Datensätze haben unterschiedliche Kritikalitäts-Niveaus für eine Workload und damit auch verschiedene Anforderungen an die Ausfallsicherheit. So können beispielsweise bestimmte kritische Daten einen RPO erfordern, der gegen Null geht, während bei anderen, weniger kritischen Daten, ein höherer RPO und somit ein gewisser Datenverlust toleriert werden kann. Ebenso können unterschiedliche Datensätze auch unterschiedliche RTO-Anforderungen haben.
3. Nutzen Sie AWS- oder Drittanbieterservices, um Backups der Daten zu erstellen. AWS Backup ist ein verwalteter Service, mit dem Backups verschiedener Datenquellen auf AWS erstellt werden können. Die meisten dieser Services verfügen auch über native Funktionen für die Backup-Erstellung. Der AWS Marketplace umfasst zahlreiche Lösungen, die diese Funktionen ebenfalls bieten. Unter „Ressourcen“ unten finden Sie weitere Informationen dazu, wie man Backups von Daten aus verschiedenen AWS-Services erstellt.
4. Für Daten, die nicht gesichert werden, sollten Sie einen Datenreproduktions-Mechanismus festlegen. Es gibt verschiedene Gründe dafür, Daten, die aus anderen Quellen reproduziert werden können, nicht zu sichern. Möglicherweise ergibt sich die Situation, dass es günstiger

ist, Daten bei Bedarf aus Quellen zu reproduzieren als ein Backup zu erstellen, da mit der Speicherung von Backups gewisse Kosten verbunden sind. Ein weiterer Grund wäre, wenn das Wiederherstellen aus einem Backup länger dauert als die Reproduktion der Daten aus anderen Quellen, was zu einer Nichteinhaltung des RTO führen würde. In solchen Situationen sollten Sie sich einen Kompromiss überlegen und einen gut definierten Prozess festlegen, wie Daten aus diesen Quellen reproduziert werden können, wenn eine Datenwiederherstellung erforderlich ist. Wenn Sie beispielsweise Daten zur Analyse aus Amazon S3 in ein Data Warehouse (wie Amazon Redshift) oder einen MapReduce-Cluster (wie Amazon EMR) geladen haben, kann es sich dabei z. B. um Daten handeln, die aus anderen Quellen reproduziert werden können. Solange die Ergebnisse dieser Analysen gespeichert werden oder reproduzierbar sind, besteht kein Risiko eines Datenverlusts durch einen Ausfall im Data Warehouse oder MapReduce-Cluster. Andere Daten, die aus Quellen reproduziert werden können, sind Cache-Inhalte (z. B. Amazon ElastiCache) oder RDS Read Replicas.

5. Legen Sie eine Kadenz für die Sicherung von Daten fest. Das Erstellen von Datenquellen ist ein periodischer Prozess und die Häufigkeit sollte vom RPO abhängen.

Grad des Aufwands für den Implementierungsplan: Mittel

Ressourcen

Relevante bewährte Methoden:

REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten:

REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen

Relevante Dokumente:

- Was ist AWS Backup?
- Was ist AWS DataSync?
- Was ist Volume Gateway?
- APN-Partner: Partner, die Sie bei der Sicherung unterstützen können
- AWS Marketplace: Für die Sicherung geeignete Produkte
- Amazon EBS Snapshots
- Backups von Amazon EFS
- Backups von Amazon FSx für Windows File Server

- Backup und Wiederherstellung für ElastiCache for Redis
- Erstellen eines DB-Cluster-Snapshots in Neptune
- Erstellen eines DB-Snapshots
- Erstellen einer EventBridge-Regel, die nach einem Zeitplan ausgelöst wird
- Regionsübergreifende Replikation mit Amazon S3
- EFS-zu-EFS-Sicherung
- Exportieren von Protokolldaten zu Amazon S3
- Verwaltung des Objektlebenszyklus
- On-Demand-Sicherung und Wiederherstellung in DynamoDB
- Zeitpunktbezogene Wiederherstellung für DynamoDB
- Mit Amazon OpenSearch Service Index-Snapshots arbeiten

Relevante Videos:

- AWS re: Ivent 2021 - Backup, disaster recovery, and ransomware protection with AWS (AWS re:Invent 2021 - Backup, Notfallwiederherstellung und Ransomware-Schutz mit AWS)
- AWS Backup Demo: Cross-Account and Cross-Region Backup (AWS Backup Demo: Konto- und regionsübergreifendes Backup)
- AWS Backup re:Invent 2019: Ausführliche Beschreibung von AWS, mit Rackspace (STG341)

Ähnliche Beispiele:

- Well-Architected Lab: Implementieren einer bidirektionalen Cross-Region Replication (CRR, regionsübergreifende Replikation) für Amazon S3
- Well-Architected Lab: Testen von Backup und Wiederherstellung von Daten
- Well-Architected lab: Backup and Restore with Failback for Analytics Workload (Well-Architected Lab: Backups und Wiederherstellung mit Failback für Analytics-Workload)
- Well-Architected Lab: Notfallwiederherstellung – Backup und Wiederherstellung

REL09-BP02 Schützen und Verschlüsseln von Backups

Steuern und erkennen Sie den Zugriff auf Backups durch Authentifizierung und Autorisierung wie z. B. mit AWS IAM. Vermeiden und erkennen Sie mittels Verschlüsselung Beeinträchtigungen der Datenintegrität von Backups.

Amazon S3 unterstützt mehrere Verschlüsselungsmethoden für gespeicherte Daten. Mithilfe der serverseitigen Verschlüsselung akzeptiert Amazon S3 Ihre Objekte als unverschlüsselte Daten und sorgt für ihre Verschlüsselung bei der Speicherung. Bei der clientseitigen Verschlüsselung ist Ihre Workload-Anwendung für die Verschlüsselung der Daten verantwortlich, bevor sie an Amazon S3 gesendet werden. Beide Methoden ermöglichen Ihnen, zum Erstellen und Speichern des Datenschlüssels AWS Key Management Service (AWS KMS) zu verwenden oder einen eigenen Schlüssel bereitzustellen, für den Sie verantwortlich sind. Bei AWS KMS können Sie mithilfe von IAM festlegen, wer auf Ihre Datenschlüssel und entschlüsselten Daten zugreifen kann.

Wenn Sie bei Amazon RDS Ihre Datenbanken verschlüsseln, werden Ihre Sicherungsdaten ebenfalls verschlüsselt. DynamoDB-Sicherungen sind immer verschlüsselt.

Gängige Antimuster:

- Derselbe Zugriff auf die Sicherungen und die automatisierte Wiederherstellung wie auf die Daten.
- Keine Verschlüsselung der Sicherungen.

Vorteile der Einführung dieser bewährten Methode: Durch den Schutz der Datensicherungen wird eine Manipulation der Daten verhindert. Ihre Verschlüsselung verhindert den Zugriff auf die Daten, wenn sie versehentlich offengelegt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden Sie die Verschlüsselung für jeden Datenspeicher. Wenn Ihre Quelldaten verschlüsselt sind, wird die Sicherung ebenfalls verschlüsselt.
  - Aktivieren Sie die Verschlüsselung in RDS. Beim Erstellen einer RDS-Instance können Sie die Verschlüsselung im Ruhezustand mit AWS Key Management Service konfigurieren.
    - Verschlüsseln von Amazon RDS-Ressourcen
  - Aktivieren Sie die Verschlüsselung für EBS-Volumes. Während der Erstellung von Volumes können Sie eine Standardverschlüsselung konfigurieren oder einen eindeutigen Schlüssel angeben.
    - Amazon EBS-Verschlüsselung
  - Verwenden Sie die erforderliche Amazon DynamoDB-Verschlüsselung. DynamoDB verschlüsselt alle Daten im Ruhezustand. Sie können entweder einen AWS-eigenen AWS KMS-Schlüssel oder

einen AWS-verwalteten KMS-Schlüssel verwenden und dabei einen Schlüssel angeben, der in Ihrem Konto gespeichert wird.

  - DynamoDB-Verschlüsselung im Ruhezustand
  - Verwalten verschlüsselter Tabellen
- Verschlüsseln Sie Ihre in Amazon EFS gespeicherten Daten. Konfigurieren Sie die Verschlüsselung beim Erstellen des Dateisystems.
  - Verschlüsseln von Daten und Metadaten in EFS
- Konfigurieren Sie die Verschlüsselung in den Quell- und Zielregionen. Sie können die Verschlüsselung im Ruhezustand in Amazon S3 mit Schlüsseln konfigurieren, die in KMS gespeichert sind. Die Schlüssel sind jedoch regionsspezifisch. Sie können die Zielschlüssel angeben, während Sie die Replikation konfigurieren.
  - Zusätzliche CRR-Konfiguration: Replizieren von Objekten, die mit serverseitiger Verschlüsselung (SSE) unter Verwendung von Verschlüsselungsschlüsseln erstellt wurden, die in AWS KMS gespeichert wurden.
- Implementieren Sie Rechte mit geringsten Berechtigungen für den Zugriff auf Ihre Backups. Begrenzen Sie den Zugriff auf die Backups, Snapshots und Replikate anhand bewährter Methoden im Bereich Sicherheit.
  - Säule „Sicherheit“: AWS Well-Architected

Ressourcen

Relevante Dokumente:

- AWS Marketplace: Für die Sicherung geeignete Produkte
- Amazon EBS-Verschlüsselung
- Amazon S3: Daten durch Verschlüsselung schützen
- Zusätzliche CRR-Konfiguration: Replizieren von Objekten, die mit serverseitiger Verschlüsselung (SSE) unter Verwendung von Verschlüsselungsschlüsseln erstellt wurden, die in AWS KMS gespeichert wurden.
- DynamoDB-Verschlüsselung im Ruhezustand
- Verschlüsseln von Amazon RDS-Ressourcen
- Verschlüsseln von Daten und Metadaten in EFS
- Verschlüsselung für Backups in AWS.
- Verwalten verschlüsselter Tabellen

- Säule „Sicherheit“: AWS Well-Architected

Ähnliche Beispiele:

- Well-Architected Lab: Implementieren einer bidirektionalen Cross-Region Replication (CRR, regionsübergreifende Replikation) für Amazon S3

REL09-BP03 Automatische Daten-Backups

Sie können die Backups so konfigurieren, dass sie automatisch nach Zeitplan, der auf dem Recovery Point Objective (RPO) basiert, oder bei Änderungen am Datensatz durchgeführt werden. Kritische Datensätze, bei denen Datenverlust vermieden werden sollte, müssen regelmäßig automatisch gesichert werden, wohingegen weniger kritische Daten, bei denen ein gewisser Verlust akzeptabel ist, weniger häufig gesichert werden können.

AWS Backup kann zum Erstellen von automatisierten Daten-Backups verschiedener AWS-Datenquellen genutzt werden. Amazon RDS-Instances können fast kontinuierlich alle fünf Minuten gesichert werden und Amazon S3-Objekte können praktisch durchgehend alle 15 Minuten gesichert werden, was eine zeitpunktbezogene Wiederherstellung (PITR) an einem bestimmten Zeitpunkt im Backup-Verlauf ermöglicht. Andere AWS-Datenquellen wie Amazon EBS-Volumes, Amazon DynamoDB-Tabellen oder Amazon FSx-Dateisysteme kann AWS Backup stündlich ein automatisiertes Backup ausführen. Diese Service bieten auch native Backup-Funktionen. Zu den AWS-Services, die ein automatisiertes Backup mit zeitpunktbezogener Wiederherstellung bieten, gehören: Amazon DynamoDB, Amazon RDSund Amazon Keyspaces (für Apache Cassandra) – diese können an einem bestimmten Zeitpunkt im Backup-Verlauf wiederhergestellt werden. Die meisten anderen AWS-Datenspeicher-Services bieten die Möglichkeit, stündliche periodische Backups einzuplanen.

Amazon RDS und Amazon DynamoDB ermöglichen eine kontinuierliche Sicherung mit zeitpunktbezogener Wiederherstellung. Die Amazon S3-Versionsverwaltung erfolgt nach der Aktivierung automatisch. Amazon Data Lifecycle Manager kann genutzt werden, um das Erstellen, Kopieren und Löschen von Amazon EBS-Snapshots zu automatisieren. Außerdem kann damit das Erstellen, das Kopieren, die Deprekation und die Abmeldung von Amazon EBS-gestützten Amazon Machine Images (AMIs) und den zugrunde liegenden Amazon EBS-Snapshots automatisiert werden.

Für eine zentrale Ansicht Ihrer Sicherungsautomatisierung und des Verlaufs bietet AWS Backup eine vollständig verwaltete, richtlinienbasierte Sicherungslösung. Diese zentralisiert und automatisiert

die Sicherung von Daten in mehreren AWS-Services in der Cloud sowie vor Ort mithilfe des AWS Storage Gateway.

Zusätzlich zum Versioning bietet Amazon S3 eine Replikationsfunktion. Der gesamte S3-Bucket kann automatisch in einen anderen Bucket in einer anderen AWS-Region repliziert werden.

Gewünschtes Ergebnis:

Ein automatisierter Prozess, der Backups von Datenquellen in einer festgelegten Kadenz erstellt.

Gängige Antimuster:

- Sicherungen werden manuell durchgeführt.
- Es werden Ressourcen mit Sicherungsfunktionen verwendet, die Sicherung wird aber nicht in die Automatisierung einbezogen.

Vorteile der Einführung dieser bewährten Methode: Durch die Automatisierung von Backups wird sichergestellt, dass diese regelmäßig auf Grundlage Ihres RPO erstellt werden und Sie andernfalls benachrichtigt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

1. Identifizieren Sie Datenquellen, die aktuell manuell gesichert werden. Unter REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen finden Sie weitere Anweisungen hierzu.
2. Bestimmen Sie den RPO für die Workload. Unter REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten: finden Sie weitere Anweisungen hierzu.
3. Nutzen Sie eine automatisierte Backup-Lösung oder einen verwalteten Service. AWS Backup ist ein vollständig verwalteter Service, der das Zentralisieren und Automatisieren des Datenschutzes über verschiedene AWS-Services hinweg, in der Cloud sowie vor Ort erleichtert.. Backup-Pläne sind ein Feature von AWS Backup, mit dem Regeln erstellt werden können, die die zu sichernden Ressourcen sowie die Häufigkeit, mit der diese Backups erstellt werden, definieren. Diese Häufigkeit sollte auf dem in Schritt 2 festgelegten RPO basieren. Dieses WA Lab bietet eine praktische Anleitung zur Erstellung von automatisierten Backups mit AWS Backup. Native Backup-Funktionen werden von den meisten AWS-Services, die Daten speichern, angeboten. So kann

beispielsweise RDS für automatisierte Backups mit zeitpunktbezogener Wiederherstellung (PITR) genutzt werden.

4. Für Datenquellen, die nicht von einer automatisierten Backup-Lösung oder einem verwalteten Service unterstützt werden, wie etwa On-Premises-Datenquellen oder Nachrichten-Warteschlangen, können Sie für die Erstellung von automatisierten Backups eine Lösung von einem zuverlässigen Drittanbieter in Betracht ziehen. Als Alternative können Sie die Automatisierung für diesen Vorgang mit der AWS CLI oder mit SDKs erstellen. Sie können AWS Lambda-Funktionen oder AWS Step Functions nutzen, um die in die Erstellung eines Daten-Backups einbezogene Logik zu definieren und Amazon EventBridge verwenden, um es in einer auf Ihrem RPO (wie in Schritt 2 festgelegt) basierenden Häufigkeit auszuführen.

Grad des Aufwands für den Implementierungsplan: Niedrig

Ressourcen

Ähnliche Dokumente:

- APN-Partner: Partner, die Sie bei der Sicherung unterstützen können
- AWS Marketplace: Für die Sicherung geeignete Produkte
- Erstellen einer EventBridge-Regel, die nach einem Zeitplan ausgelöst wird
- Was ist AWS Backup?
- Was ist AWS Step Functions?

Ähnliche Videos:

- AWS re:Invent 2019: Ausführliche Beschreibung von AWS Backup, mit Rackspace (STG341)

Ähnliche Beispiele:

- Well-Architected Lab: Testen von Backup und Wiederherstellung von Daten

REL09-BP04 Verifizieren der Sicherungsintegrität und -verfahren durch regelmäßiges Wiederherstellen der Daten

Überprüfen Sie mit einem Wiederherstellungstest, ob sich mit Ihren Sicherungsverfahren das RTO und das RPO einhalten lassen.

Mit AWS können Sie eine Testumgebung einrichten und Ihre Sicherungen wiederherstellen, um RTO- und RPO-Funktionen zu bewerten und Tests für Dateninhalte und Integrität durchzuführen.

Darüber hinaus ermöglichen Amazon RDS und Amazon DynamoDB eine Point-in-Time-Wiederherstellung. Durch die kontinuierliche Sicherung können Sie Ihren Datensatz in den Zustand zurücksetzen, in dem er sich an einem bestimmten Datum und zu einer bestimmten Uhrzeit befand.

Gewünschtes Ergebnis: Daten aus Backups werden mittels gut definierter Mechanismen regelmäßig wiederhergestellt, um zu gewährleisten, dass die Wiederherstellung innerhalb des festgelegten Recovery Time Objective (RTO) für die Workload möglich ist. Überprüfen Sie, dass die Wiederherstellung aus einem Backup in eine Ressource erfolgt, die die Originaldaten enthält und dass keine dieser Daten korrupt oder nicht zugänglich sind, sowie dass sich der Datenverlust im Rahmen des Recovery Point Objective (RPO) bewegt.

Gängige Antimuster:

- Eine Sicherung wird wiederhergestellt, es werden aber keine Daten abgefragt oder abgerufen, um sicherzustellen, dass die Wiederherstellung nutzbar ist.
- Es wird angenommen, dass ein Backup existiert.
- Es wird angenommen, dass das Backup eines System voll funktionsfähig ist und Daten daraus wiederhergestellt werden können.
- Es wird angenommen, dass die Zeit für das Wiederherstellen von Daten aus einem Backup innerhalb des RTO für die Workload liegt.
- Es wird angenommen, dass die im Backup enthaltenen Daten in den RPO für die Workload fallen.
- Es erfolgt eine Ad-hoc-Wiederherstellung ohne die Nutzung eines Runbooks oder außerhalb eines festgelegten automatisierten Verfahrens.

Vorteile der Einführung dieser bewährten Methode: Durch das Testen der Wiederherstellung der Backups stellen Sie sicher, dass Daten bei Bedarf wiederhergestellt werden können (ohne dass Sie sich um fehlende oder korrupte Daten sorgen müssen), dass die Wiederherstellung innerhalb des RTO für die Workload möglich ist und dass sich mögliche Datenverluste im Rahmen des RPO für die Workload bewegen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Das Testen der Sicherungs- und Wiederherstellungsfunktionen stärkt das Vertrauen in die Fähigkeit zur Durchführung dieser Aktionen während eines Ausfalls. Stellen Sie regelmäßig Backups an einem neuen Speicherort wieder her und führen Sie Tests aus, um die Datenintegrität zu überprüfen. Zu den gängigen Tests, die ausgeführt werden sollten, gehören

das Überprüfen, ob die Daten verfügbar sind, nicht korrupt sind, zugänglich sind und ob ein möglicher Datenverlust innerhalb des RPO für die Workload liegt. Solche Tests können dabei helfen, zu ermitteln, ob die Wiederherstellungsmechanismen schnell genug sind, um dem RTO der Workload gerecht zu werden.

1. Identifizieren Sie Datenquellen, für die derzeit Backups erstellt werden, und prüfen Sie, wo diese Backups gespeichert werden. Unter REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen finden Sie eine Anleitung dazu, wie Sie dies umsetzen können.
2. Legen Sie Kriterien für die Datenvalidierung für jede Datenquelle fest. Verschieden Datentypen können unterschiedliche Eigenschaften aufweisen und somit auch unterschiedliche Validierungsmechanismen erfordern. Überlegen Sie, wie diese Daten validiert werden können, bevor Sie sie in der Produktion einsetzen. Häufig werden für die Datenvalidierung Daten- und Sicherungseigenschaften wie Datentyp, Format, Prüfsumme, Größe oder eine Kombination dieser Eigenschaften mit einer benutzerdefinierten Validierungslogik verwendet. Ein Beispiel hierfür wäre der Vergleich der Prüfsummenwerte zwischen der wiederhergestellten Ressource und der Datenquelle zum Zeitpunkt der Erstellung des Backups.
3. Bestimmen Sie RTO und RPO, um die Daten basierend auf der Datenkritikalität wiederherzustellen. Unter REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten: finden Sie eine Anleitung dazu, wie Sie dies umsetzen können.
4. Bewerten Sie die Funktion zur Datenwiederherstellung. Prüfen Sie Ihre Sicherungs- und Wiederherstellungsstrategie, um festzustellen, ob sie Ihre RTO und RPO erfüllen kann, und passen Sie die Strategie bei Bedarf an. Mit AWS Resilience Hubkönnen Sie eine Bewertung Ihrer Workload durchführen. Dabei wird Ihre Anwendungskonfiguration im Hinblick auf die Ausfallsicherheitsrichtlinien bewertet und Sie erfahren, ob Ihre RTO- und RPO-Ziele erfüllt werden können.
5. Führen Sie eine Test-Wiederherstellung mit den aktuell festgelegten Prozessen, die in der Produktion für die Datenwiederherstellung genutzt werden, durch. Diese Prozesse hängen davon ab, wie die ursprüngliche Datenquelle gesichert wurde sowie vom Format und der Speicherung des Backups selbst oder davon, ob die Daten aus anderen Quellen reproduziert werden. Wenn

Sie beispielsweise einen verwalteten Service wie AWS Backup verwenden, könnte der Prozess ganz einfach darin bestehen, das Backup in einer neuen Ressource wiederherzustellen.. Wenn Sie AWS Elastic Disaster Recovery verwendet haben, können Sie einen Wiederherstellung-Drill starten.

6. Validieren Sie die Datenwiederherstellung aus der wiederhergestellten Ressource (im vorangegangenen Schritt) basierend auf Kriterien, die Sie zuvor in Schritt 2 für die Datenvalidierung festgelegt haben. Enthalten diese wiederhergestellten Daten den neuesten Datensatz/das neueste Element zum Zeitpunkt des Backups? Fallen diese Daten in den RPO für die Workload?
7. Ermitteln Sie die für das Wiederherstellen benötigte Zeit und vergleichen Sie sie mit dem in Schritt 3 festgelegten RTO. Ist dieser Prozess Teil des RTO für die Workload? Vergleichen Sie beispielsweise den Zeitstempel des Starts des Wiederherstellungsprozesses und des Abschlusses der Wiederherstellungsbewertung, um zu ermitteln, wie lange dieser Prozess dauert. Alle AWS-API-Aufrufe haben einen Zeitstempel. Sie finden diese Informationen unter AWS CloudTrail. Während diese Informationen Details dazu liefern können, wann der Wiederherstellungsprozess gestartet wurde, sollte der End-Zeitstempel für den Abschluss der Validierung von der Validierungslogik aufgezeichnet werden. Wenn Sie einen automatisierten Prozess verwenden, können Services wie Amazon DynamoDB zum Speichern dieser Informationen genutzt werden. Darüber hinaus können viele AWS-Services ein Ereignisprotokoll bereitstellen, das mit einem Zeitstempel versehene Informationen dazu enthält, wann bestimmte Aktionen aufgetreten sind. Innerhalb von AWS Backup werden Sicherungs- und Wiederherstellungsaktionen als Jobsbezeichnet. Diese Jobs enthalten Zeitstempelinformationen als Teil ihrer Metadaten, die zum Ermitteln der für die Wiederherstellung benötigte Zeit verwendet werden können.
8. Benachrichtigen Sie die Beteiligten, wenn die Datenvalidierung fehlschlägt oder die für die Wiederherstellung benötigte Zeit den festgelegten RTO für die Workload überschreitet. Beim Implementieren der Automatisierung hierfür, wie in diesem Lab, können Services wie Amazon Simple Notification Service (Amazon SNS) genutzt werden, um Push-Benachrichtigungen wie E-Mails oder SMS an die Beteiligten zu senden. Diese Benachrichtigungen können auch in Nachrichtenanwendungen wie Amazon Chime, Slack oder Microsoft Teams veröffentlicht oder dazu verwendet werden, Aufgaben anhand von AWS Systems Manager OpsCenter als OpsItems zu erstellen.
9. Lassen Sie diesen Prozess regelmäßig automatisch ausführen. Sie können beispielsweise Services wie AWS Lambda oder einen Zustandsautomaten in AWS Step Functions nutzen, um die Wiederherstellungsprozesse zu automatisieren. Außerdem können Sie Amazon EventBridge verwenden, um diesen automatisierten Workflow regelmäßig auszulösen, wie im folgenden Architekturdiagramm abgebildet. Erfahren Sie, wie Sie die Validierung der Datenwiederherstellung

mit AWS Backup automatisieren. Darüber hinaus bietet dieses Well-Architected Lab eine praktische Schulung zum Automatisieren mehrerer der hier aufgeführten Schritte.

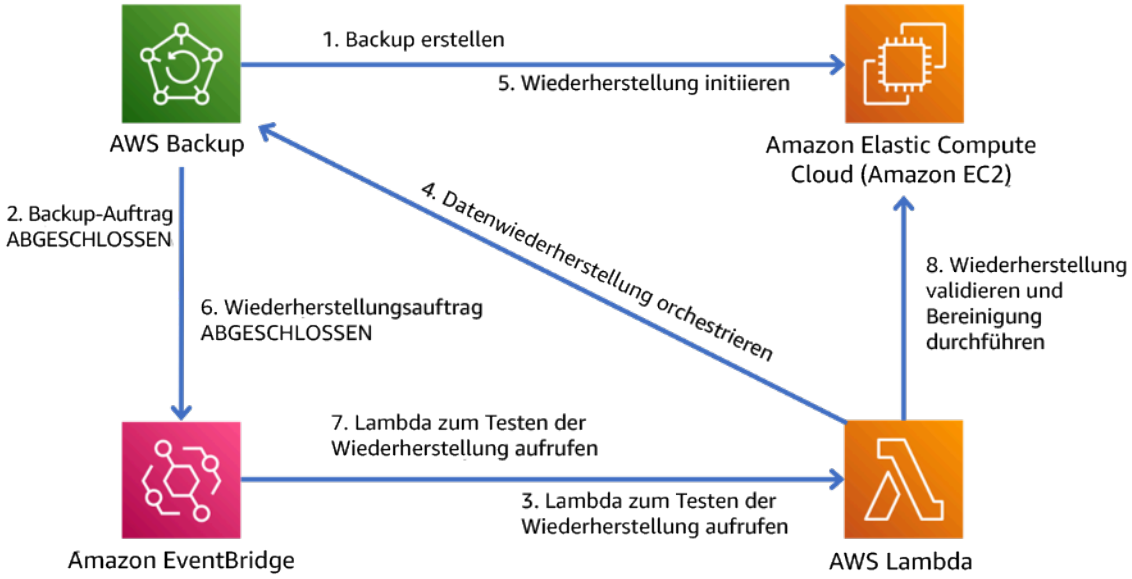


Abbildung 9. Ein automatisierter Sicherungs- und Wiederherstellungsprozess

Grad des Aufwands für den Implementierungsplan: Mittel bis hoch, je nach Komplexität des Validierungskriteriums.

Ressourcen

Ähnliche Dokumente:

- die Validierung der Datenwiederherstellung mit AWS Backup automatisieren
- APN-Partner: Partner, die Sie bei der Sicherung unterstützen können
- AWS Marketplace: Für die Sicherung geeignete Produkte
- Erstellen einer EventBridge-Regel, die nach einem Zeitplan ausgelöst wird
- On-Demand-Sicherung und Wiederherstellung in DynamoDB
- Was ist AWS Backup?
- Was ist AWS Step Functions?
- Was ist AWS Elastic Disaster Recovery
- AWS Elastic Disaster Recovery

Ähnliche Beispiele:

- Well-Architected Lab: Testen von Backup und Wiederherstellung von Daten

ZUV 10 Wie schützen Sie Ihren Workload mithilfe der Fehlerisolierung?

Fehlerisolierte Grenzen beschränken die Auswirkungen eines Ausfalls innerhalb eines Workloads auf eine begrenzte Anzahl von Komponenten. Komponenten außerhalb der Grenze sind vom Ausfall nicht betroffen. Wenn Sie mehrere fehlerisolierte Grenzen verwenden, können Sie die Auswirkungen auf Ihren Workload einschränken.

Bewährte Methoden

- REL10-BP01 Bereitstellen des Workloads an mehreren Standorten
- REL10-BP02 Auswählen der geeigneten Standorte für Ihre Multi-Standort-Bereitstellung
- REL10-BP03 Automatisierte Wiederherstellung für Komponenten, die auf einen einzelnen Standort beschränkt sind
- REL10-BP04 Verwenden von Bulkhead-Architekturen, um den Umfang von Beeinträchtigungen zu begrenzen

REL10-BP01 Bereitstellen des Workloads an mehreren Standorten

Verteilen Sie die Workload-Daten und -Ressourcen über mehrere Availability Zones oder ggf. über mehrere AWS-Regionen. Die Standorte können so vielfältig wie nötig sein.

Eins der grundlegenden Prinzipien für das Servicedesign in AWS ist die Vermeidung von Single Points of Failure in der zugrunde liegenden physischen Infrastruktur. Dies treibt uns an, Software und Systeme zu entwickeln, die mehrere Availability Zones verwenden und Schutz beim Ausfall einer einzelnen Region bieten. Außerdem sollen Systeme gegen den Ausfall einzelner Compute-Knoten, einzelner Speicher-Volumes oder einzelner Instances einer Datenbank geschützt sein. Bei der Entwicklung eines Systems, das auf redundanten Komponenten basiert, muss gewährleistet sein, dass die Komponenten unabhängig voneinander betrieben werden und im Falle von AWS-Regionen autonom sind. Die Vorteile theoretischer Verfügbarkeitsberechnungen mit redundanten Komponenten sind nur anwendbar, wenn diese Voraussetzung erfüllt ist.

Availability Zones (AZs)

AWS-Regionen bestehen aus mehreren voneinander unabhängigen Availability Zones. Die einzelnen Availability Zones sind durch eine signifikante physische Distanz voneinander getrennt, um korrelierte Fehlerszenarios aufgrund von Umweltgefahren wie Feuer, Überflutungen und Tornados zu vermeiden. Jede Availability Zone verfügt außerdem über eine unabhängige physische Infrastruktur:

eigene Verbindungen zur Stromversorgung, unabhängige Backup-Stromquellen, unabhängige mechanischen Services und unabhängige Netzwerkkonnektivität innerhalb der Availability Zone und darüber hinaus. Durch dieses Design bleiben Fehler in einem dieser Systeme auf die jeweils betroffene AZ beschränkt. Trotz ihrer geografischen Verteilung befinden sich Availability Zones in demselben regionalen Bereich, wodurch Netzwerke mit hohem Durchsatz und geringer Latenz ermöglicht werden. Die gesamte AWS-Region (über alle Availability Zones, die aus mehreren physisch unabhängigen Rechenzentren bestehen) kann wie ein logisches Bereitstellungsziel für Ihren Workload behandelt werden. Dies umfasst auch die Möglichkeit zum synchronen Replizieren von Daten (z. B. zwischen Datenbanken). So können Sie Availability Zones in einer Aktiv-Aktiv- oder einer Aktiv-Standby-Konfiguration nutzen.

Availability Zones sind voneinander unabhängig. Daher erhöht sich die Workload-Verfügbarkeit, wenn in der Architektur des Workloads mehrere Zonen verwendet werden. Einige AWS-Services (darunter auch die Amazon EC2-Instance-Datenebene) werden als strikte zonale Services bereitgestellt, die von denselben Fehlern betroffen sind wie die Availability Zone, in der sie sich befinden. Amazon EC2-Instances in den anderen AZs sind hingegen nicht betroffen und weiterhin funktionsfähig. Wenn entsprechend ein Fehler in einer Availability Zone zum Ausfall einer Amazon Aurora-Datenbank führt, kann eine Auslese-Replikat-Aurora-Instance in einer nicht betroffenen AZ automatisch zur primären Instance hochgestuft werden. Regionale AWS-Services wie Amazon DynamoDB wiederum verwenden intern mehrere Availability Zones in einer Aktiv-Aktiv-Konfiguration, um die Verfügbarkeitsdesignziele für den jeweiligen Service zu erfüllen, ohne dass Sie die AZ-Platzierung konfigurieren müssen.

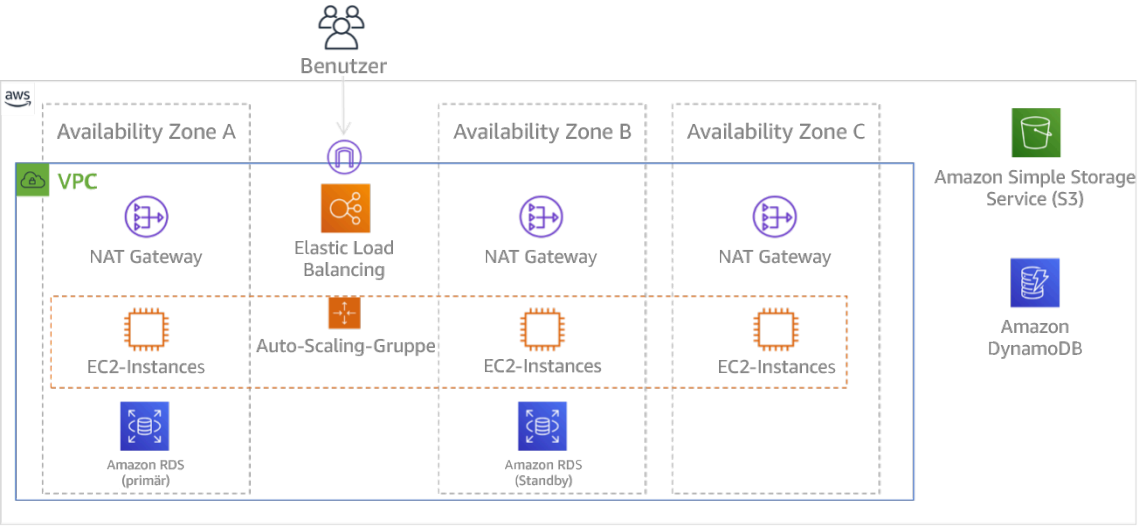


Abbildung 9: Mehrstufige Architektur, die in drei Availability Zones bereitgestellt wird. Amazon S3 und Amazon DynamoDB nutzen immer automatisch mehrere AZs. Auch der ELB wird in allen drei Zonen bereitgestellt.

Während Amazon EBS-Steuerebenen in der Regel die Möglichkeit bieten, Ressourcen innerhalb der gesamten Region (also in mehreren Availability Zones) zu verwalten, haben bestimmte Steuerebenen (wie AWS und Amazon EC2) die Fähigkeit, Ergebnisse in eine einzelne Availability Zone zu filtern. Wenn dies erledigt ist, wird die Anfrage nur in der angegebenen Availability Zone verarbeitet; dies reduziert die Wahrscheinlichkeit von Ausfällen in anderen Availability Zones. Dieses AWS CLI-Beispiel veranschaulicht das Abrufen von Amazon EC2-Instance-Informationen ausschließlich aus der Availability Zone „us-east-2c“:

```
AWS ec2 describe-instances --filters Name=availability-zone,Values=us-east-2c
```

AWS Local Zones

AWS Local Zones verhalten sich ähnlich wie Availability Zones innerhalb ihrer jeweiligen AWS-Region. Sie können als Platzierungsstandort für zonale AWS-Ressourcen wie Subnetze und EC2-Instances ausgewählt werden. Das Besondere daran ist, dass sie sich nicht in der zugehörigen AWS-Region befinden, sondern in der Nähe großer Ballungsräume, Industrie- und IT-Zentren, in denen derzeit keine AWS-Region vorhanden ist. Sie sorgen dennoch für eine sichere Verbindung mit hoher Bandbreite zwischen lokalen Workloads in der lokalen Zone und Workloads in der AWS-Region.
Sie sollten AWS Local Zones verwenden, um Workloads mit Anforderungen an eine geringe Latenz näher bei Ihren Benutzern bereitzustellen.

Amazon Global Edge Network

Amazon Global Edge Network besteht aus Edge-Standorten in Städten auf der ganzen Welt. Amazon CloudFront nutzt dieses Netzwerk, um Inhalte mit geringerer Latenz für Endbenutzer bereitzustellen. Mit AWS Global Accelerator können Sie Ihre Workload-Endpunkte an diesen Edge-Standorten erstellen, um ein Onboarding in das globale AWS-Netzwerk in der Nähe Ihrer Benutzer zu ermöglichen. Amazon API Gateway können Sie Edge-optimierte API-Endpunkte mithilfe einer CloudFront-Verteilung verwenden, um den Client-Zugriff über den nächstgelegenen Edge-Standort zu erleichtern.

AWS-Regionen

AWS-Regionen sind autonom konzipiert. Daher können Sie dedizierte Kopien von Services für jede Region bereitstellen, um einen multiregionalen Ansatz zu verwenden.

Ein multiregionaler Ansatz wird häufig für Strategien der Notfallwiederherstellung eingesetzt, um Wiederherstellungsziele zu erfüllen, falls einmalige Ereignisse mit großer Reichweite auftreten.

Siehe Planung der Notfallwiederherstellung für weitere Informationen zu diesen Strategien. Hier liegt der Schwerpunkt allerdings auf der Verfügbarkeit, wobei versucht wird, ein mittleres Betriebszeitziel über einen längeren Zeitraum zu erreichen. Wenn eine hohe Verfügbarkeit angestrebt wird, ist eine multiregionale Architektur normalerweise Aktiv-Aktiv konzipiert. Dabei sind die einzelnen Servicekopien (in den jeweiligen Regionen) aktiv (und bearbeiten Anfragen).

> **Empfehlung**
>
> Die Verfügbarkeitsziele für die meisten Workloads können mithilfe einer Multi-AZ-Strategie innerhalb einer einzelnen AWS-Region erfüllt werden. Ziehen Sie multiregionale Architekturen nur in Betracht, wenn für Workloads extreme Verfügbarkeitsanforderungen gelten oder andere Unternehmensziele eine solche Architektur erforderlich machen.

AWS bietet Ihnen die Möglichkeit, Services regionsübergreifend zu betreiben. AWS stellt beispielsweise eine fortlaufende asynchrone Datenreplikation mit Amazon S3-Replikation (Amazon Simple Storage Service), Amazon RDS-Lesereplikaten (u. a. Aurora-Lesereplikaten) und globalen Amazon DynamoDB-Tabellen bereit. Bei der fortlaufenden Replikation sind Versionen Ihrer Daten für die fast sofortige Nutzung in jeder aktiven Region verfügbar.

Mit AWS CloudFormation können Sie Ihre Infrastruktur definieren und einheitlich in AWS-Konten und AWS-Regionen bereitstellen. AWS CloudFormation StackSets erweitern diese Funktionen, indem Sie AWS CloudFormation-Stacks mit nur einem Vorgang in verschiedenen Konten und Regionen erstellen, aktualisieren oder löschen können. Bei Amazon EC2-Instance-Bereitstellungen wird ein AMI (Amazon Machine Image) verwendet, um Informationen wie die Hardwarekonfiguration und installierte Software bereitzustellen. Sie können eine Amazon EC2 Image Builder-Pipeline implementieren, die die benötigten AMIs erstellt, und diese in Ihre aktiven Regionen kopieren. Diese goldenen AMIs enthalten alles, was Sie zum Bereitstellen und Skalieren von Workloads in neuen Regionen benötigen.

Zum Weiterleiten von Datenverkehr ermöglichen sowohl Amazon Route 53 als auch AWS Global Accelerator das Definieren von Richtlinien, die angeben, welche Benutzer zu welchem aktiven regionalen Endpunkt geleitet werden. Mit Global Accelerator legen Sie für den Datenverkehr einen Prozentwert fest, der an die einzelnen Anwendungsendpunkte geleitet wird. Route 53 unterstützt diesen Ansatz mit Prozentwerten sowie eine Vielzahl weiterer Richtlinien, u. a. auf Grundlage der geografischen Nähe oder der Latenz. Global Accelerator nutzt automatisch das umfassende Netzwerk von AWS-Edge-Servern, um Datenverkehr an den Backbone des AWS-Netzwerks zu senden, sobald dies möglich ist. Dies führt zu einer geringeren Latenz bei Abfragen.

Alle diese Funktionen sind so konzipiert, dass die Autonomie der einzelnen Regionen erhalten wird. Es gibt nur sehr wenige Ausnahmen von diesem Ansatz, darunter unsere Services für eine weltweite Edge-Lieferung (z. B. Amazon CloudFront und Amazon Route 53) und die Steuerebene für den AWS Identity and Access Management-Service (IAM). Die meisten Services werden vollständig innerhalb einer einzigen Region betrieben.

On-Premises-Rechenzentrum

Für Workloads, die in einem On-Premises-Rechenzentrum ausgeführt werden, sollten Sie nach Möglichkeit eine hybride Umgebung erstellen. AWS Direct Connect bietet eine dedizierte Netzwerkverbindung zwischen Ihrem Standort und AWS, sodass eine Ausführung in beiden Umgebungen möglich ist.

Außerdem haben Sie die Möglichkeit, AWS-Infrastruktur und -Services mit AWS Outposts lokal auszuführen. AWS Outposts ist ein vollständig verwalteter Service, der die AWS-Infrastruktur, AWS-Services, APIs und Tools auf Ihr Rechenzentrum erweitert. Die gleiche Hardwareinfrastruktur, die in der AWS Cloud verwendet wird, wird dafür in Ihrem Rechenzentrum installiert. AWS Outposts werden dann mit der nächstgelegenen AWS-Region verbunden. Anschließend können Sie AWS Outposts verwenden, um Workloads mit geringer Latenz oder lokalen Datenverarbeitungsanforderungen zu unterstützen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Verwenden Sie mehrere Availability Zones und AWS-Regionen. Verteilen Sie die Workload-Daten und -Ressourcen über mehrere Availability Zones oder ggf. über mehrere AWS-Regionen. Die Standorte können so vielfältig wie nötig sein.
  - Regionale Services werden von Haus aus in Availability Zones bereitgestellt.
    - Dazu gehören Amazon S3, Amazon DynamoDB und AWS Lambda (wenn keine VPC-Verbindung vorhanden ist).
  - Stellen Sie Ihre Container-, Instance- und funktionsbasierten Workloads in mehreren Availability Zones bereit. Verwenden Sie Multi-AZ-Datenspeicher, einschließlich Cache. Nutzen Sie EC2 Auto Scaling, die ECS-Aufgabenplatzierung, ElastiCache-Cluster sowie bei Ausführung in Ihrer VPC AWS Lambda-Funktionen.
    - Verwenden Sie für die Bereitstellung von Auto-Scaling-Gruppen Subnetze in getrennten Availability Zones.
      - Beispiel: Verteilen von Instances in Availability Zones

    - Strategien zur Aufgabenplatzierung mit Amazon ECS
    - Konfigurieren einer AWS Lambda-Funktion für den Zugriff auf Ressourcen in einer Amazon VPC
    - Auswählen von Regionen und Availability Zones
  - Verwenden Sie für die Bereitstellung von Auto-Scaling-Gruppen Subnetze in getrennten Availability Zones.
    - Beispiel: Verteilen von Instances in Availability Zones
  - Verwenden Sie ECS-Parameter für die Platzierung von Aufgaben unter Angabe von DB-Subnetzgruppen.
    - Strategien zur Aufgabenplatzierung mit Amazon ECS
  - Nutzen Sie Subnetze in mehreren Availability Zones, wenn Sie eine in Ihrem VPC auszuführende Funktion konfigurieren.
    - Konfigurieren einer AWS Lambda-Funktion für den Zugriff auf Ressourcen in einer Amazon VPC
  - Verwenden Sie mehrere Availability Zones mit ElastiCache-Clustern.
    - Auswählen von Regionen und Availability Zones
- Wenn Ihr Workload für mehrere Regionen bereitgestellt werden muss, sollten Sie sich für eine Strategie mit mehreren Regionen entscheiden. Die meisten Zuverlässigkeitsanforderungen können mithilfe einer Multi-Availability-Zone-Strategie innerhalb einer einzelnen AWS-Region erfüllt werden. Verwenden Sie eine Multi-Regionen-Strategie, wenn notwendig, um Ihre Geschäftsanforderungen zu erfüllen.
  - AWS re:Invent 2018: Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen (ARC209-R2)
    - Ein Backup in einer anderen AWS-Region kann zusätzliche Gewissheit bieten, dass Daten verfügbar sind, wenn sie benötigt werden.
    - Für einige Workloads gibt es gesetzliche Anforderungen, die eine Multi-Region-Strategie erfordern.
- Evaluieren Sie AWS Outposts für Ihren Workload. Wenn Ihre Workload eine niedrige Latenz für Ihr Rechenzentrum vor Ort erfordert oder lokale Datenverarbeitungsanforderungen hat. Führen Sie anschließend AWS-Infrastruktur und -Services On-Premises mit AWS Outposts aus.
  - Was ist AWS Outposts?
- Ermitteln Sie, ob AWS Local Zones Sie bei der Bereitstellung von Services für Ihre Benutzer unterstützt. Wenn Sie Anforderungen an eine geringe Latenz haben, prüfen Sie, ob sich AWS Local

Zones in der Nähe Ihrer Benutzer befindet. Wenn dies der Fall ist, stellen Sie damit Workloads näher an diesen Benutzern bereit.

- AWS Local Zones – häufig gestellte Fragen

Ressourcen

Ähnliche Dokumente:

- Globale AWS-Infrastruktur
- AWS Local Zones – häufig gestellte Fragen
- Strategien zur Aufgabenplatzierung mit Amazon ECS
- Auswählen von Regionen und Availability Zones
- Beispiel: Verteilen von Instances in Availability Zones
- Globale Tabellen: Multiregionale Replikation mit DynamoDB
- Verwenden von Amazon Aurora Global Databases
- Blog-Reihe: Creating a Multi-Region Application with AWS Services (Erstellen einer Multi-Region-Anwendung mit AWS-Services)
- Was ist AWS Outposts?

Relevante Videos:

- AWS re:Invent 2018: Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen (ARC209-R2)
- AWS re:Invent 2019: Innovation und Betrieb der globalen Netzwerkinfrastruktur von AWS (NET339)

REL10-BP02 Auswählen der geeigneten Standorte für Ihre Multi-Standort-Bereitstellung

Gewünschtes Ergebnis

Für eine hohe Verfügbarkeit stellen Sie Ihre Workload-Komponenten (falls möglich) immer in mehreren Availability Zone (AZ) bereit, wie in Abbildung 10 dargestellt. Überdenken Sie bei Workloads mit extremen Anforderungen an die Ausfallsicherheit die Optionen für eine Multi-Region-Architektur genau.

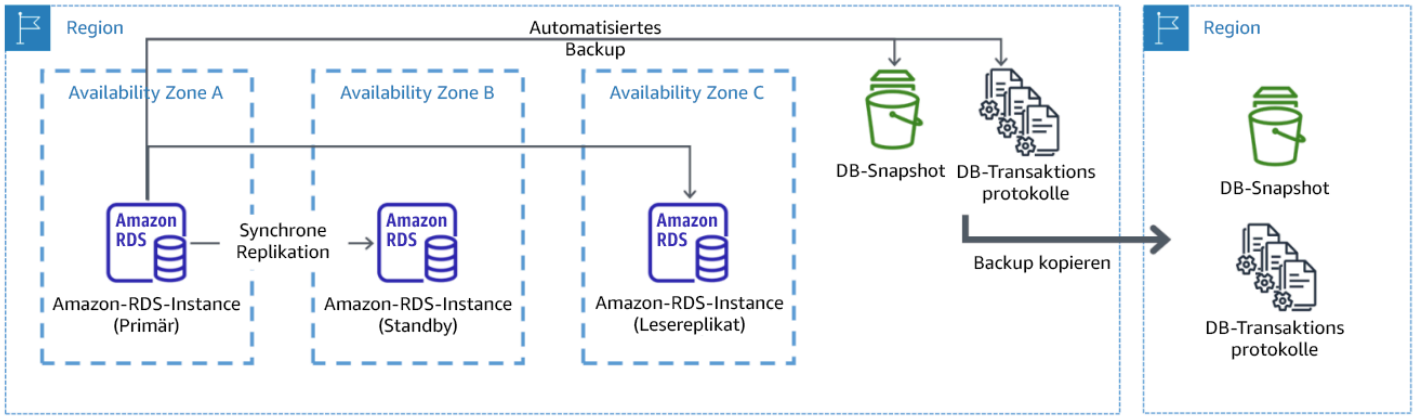


Abbildung 10: Resiliente Multi-AZ-Datenbankbereitstellung mit Backup in einer anderen AWS-Region

Gängige Antimuster

- Entscheidung für das Design einer Multi-Region-Architektur, wenn eine Multi-AZ-Architektur für die Anforderungen ausreichend wäre.
- Fehlende Berücksichtigung der Abhängigkeiten zwischen Anwendungskomponenten, wenn diese Komponenten unterschiedliche Anforderungen im Bezug auf Ausfallsicherheit und mehrere Standorte aufweisen.

Vorteile der Einführung dieser bewährten Methode:

Für die Ausfallsicherheit sollten Sie einen Ansatz wählen, bei dem verschiedene Verteidigungsebenen aufgebaut werden. Eine Ebene schützt vor kleineren, häufiger auftretenden Unterbrechungen, indem eine hochverfügbare Architektur mit mehreren AZs erstellt wird. Eine weitere Verteidigungsebene schützt vor seltenen Ereignissen wie Naturkatastrophen mit großer Reichweite und Unterbrechungen auf Regionsebene. Für diese zweite Ebene muss die Architektur Ihrer Anwendung mehrere AWS-Regionen umfassen.

- Der Unterschied zwischen einer Verfügbarkeit von 99,5 % und 99,99 % beträgt über 3,5 Stunden pro Monat. Die erwartete Verfügbarkeit eines Workloads kann nur „four nines“ (d. h. 99,99 %) erreichen, wenn er sich in mehreren AZs befindet.
- Indem Sie einen Workload in mehreren AZs ausführen, können Sie Fehler bei der Stromversorgung, Kühlung, im Netzwerk sowie die meisten Naturkatastrophen wie Feuer und Überflutung isolieren.
- Wenn Sie eine Multi-Region-Strategie für Ihren Workload implementieren, ist er vor weitreichenden Naturkatastrophen, die einen großen geografischen Bereich in einem Land betreffen, oder

technischen Fehlern in einer ganzen Region geschützt. Beachten Sie dabei, dass das Implementieren einer Multi-Region-Architektur äußerst komplex sein kann und bei den meisten Workloads nicht erforderlich ist.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Bei einer Unterbrechung oder dem teilweisen Ausfall einer Availability Zone hilft die Implementierung eines hoch verfügbaren Workloads in mehreren Availability Zones innerhalb einer einzelnen AWS-Region, die Folgen von Naturkatastrophen oder technischen Problemen zu begrenzen. Jede AWS-Region besteht aus mehreren Availability Zones, die von Fehlern in den jeweils anderen Zonen isoliert sind und die eine deutliche Distanz aufweisen. In Bezug auf Notfallereignisse, bei denen das Risiko des Ausfalls mehrerer, voneinander weit entfernter Availability-Zone-Komponenten besteht, sollten Sie Optionen für die Notfallwiederherstellung implementieren. So können Sie Fehler eingrenzen, die sich auf eine ganze Region auswirken. Bei Workloads, für die eine extreme Ausfallsicherheit erforderlich ist (kritische Infrastruktur, gesundheitsbezogene Anwendungen, Infrastruktur von Finanzsystemen usw.) wird möglicherweise eine Multi-Region-Strategie benötigt.

Implementierungsschritte

1. Analysieren Sie Ihren Workload und bestimmen Sie, ob die Anforderungen an die Ausfallsicherheit mit einem Multi-AZ-Ansatz erfüllt werden (eine AWS-Region) oder ob ein Multi-Region-Ansatz erforderlich ist. Das Implementieren einer Multi-Region-Architektur, um diese Anforderungen zu erfüllen, führt zu einer höheren Komplexität. Betrachten Sie daher Ihren Anwendungsfall und wägen Sie die Anforderungen sorgfältig ab. Die Anforderungen an die Ausfallsicherheit können fast immer auch mit einer AWS-Region erfüllt werden. Berücksichtigen Sie bei der Entscheidung, ob Sie mehrere Regionen verwenden möchten, die folgenden möglichen Anforderungen:
    a. Notfallwiederherstellung (Disaster Recovery, DR): Bei einer Unterbrechung oder dem teilweisen Ausfall einer Availability Zone hilft die Implementierung eines hoch verfügbaren Workloads in mehreren Availability Zones innerhalb einer einzelnen AWS-Region, die Folgen von Naturkatastrophen oder technischen Problemen zu begrenzen. In Bezug auf Notfallereignisse, bei denen das Risiko des Ausfalls mehrerer, voneinander weit entfernter Availability Zone-Komponenten besteht, sollten Sie eine Notfallwiederherstellung in mehreren Regionen implementieren. So können Sie die Risiken durch Naturkatastrophen oder technische Fehler eingrenzen, die sich auf eine ganze Region auswirken.

b. Hohe Verfügbarkeit (High Availability, HA): Mit einer Multi-Region-Architektur (mit mehreren AZs in jeder Region) kann eine höhere Verfügbarkeit als „four 9's" (> 99,99 %) erreicht werden.

c. Stack-Lokalisierung: Beim Bereitstellen eines Workloads für Benutzer weltweit können Sie lokalisierte Stacks in verschiedenen AWS-Regionen bereitstellen, um die Benutzer in diesen Regionen zu versorgen. Die Lokalisierung kann Sprache, Währung und die gespeicherten Datentypen umfassen.

d. Nähe zu den Benutzern: Wenn Sie einen Workload für Benutzer weltweit bereitstellen, können Sie die Latenz reduzieren, indem Sie Stacks in AWS-Regionen in der Nähe der Endbenutzer bereitstellen.

e. Datenresidenz: Für einige Workloads gelten Anforderungen an die Datenresidenz, d. h. die Daten von bestimmten Nutzern müssen innerhalb der Grenzen eines bestimmten Landes gespeichert werden. Abhängig von der jeweiligen Regelung können Sie einen ganzen Stack oder nur die Daten in der AWS-Region innerhalb dieser Landesgrenzen bereitstellen.

2. Im Folgenden finden Sie einige Bespiele für Multi-AZ-Funktionen, die von AWS-Services bereitgestellt werden:

   a. Um Workloads mit EC2 oder ECS zu schützen, stellen Sie einen Elastic Load Balancer vor den Datenverarbeitungsressourcen bereit. Elastic Load Balancing bietet so die Lösung, um Instances in fehlerhaften Zonen zu erkennen und den Datenverkehr zu fehlerfreien Zonen zu leiten.

      i. Erste Schritte mit Application Load Balancers

      ii. Erste Schritte mit Network Load Balancers

   b. Bei EC2-Instances, auf denen kommerzielle Standardsoftware ohne Unterstützung für Load Balancing ausgeführt wird, können Sie eine gewisse Fehlertoleranz durch die Implementierung einer Methodologie für die Multi-AZ-Notfallwiederherstellung erreichen.

      i. the section called "REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen"

   c. Stellen Sie für Amazon ECS-Aufgaben den Service gleichmäßig auf drei AZs verteilt bereit, um eine ausgeglichene Verteilung von Verfügbarkeit und Kosten zu erreichen.

      i. Bewährte Methoden für die Amazon ECS-Verfügbarkeit | Container

   d. Wenn Sie nicht mit Aurora Amazon RDS arbeiten, können Sie Multi-AZ als Konfigurationsoption auswählen. Beim Ausfall der primären Datenbank-Instance stuft Amazon RDS automatisch eine Standby-Datenbank hoch, sodass sie Datenverkehr in einer anderen Availability Zone empfangen kann. Außerdem können Multi-Region-Lesereplikate erstellt werden, um die Ausfallsicherheit zu steigern.

   i. Amazon RDS-Multi-AZ-Bereitstellungen

   ii. Erstellen eines Lesereplikats in einer anderen AWS-Region

3. Im Folgenden finden Sie einige Bespiele für Multi-Region-Funktionen, die von AWS-Services bereitgestellt werden:

   a. Für Amazon S3-Workloads, bei denen Multi-AZ-Verfügbarkeit automatisch vom Service bereitgestellt wird, erwägen Sie Multi-Region-Zugriffspunkte, wenn eine Multi-Region-Bereitstellung benötigt wird.

      i. Multi-Region-Zugriffspunkte in Amazon S3

   b. Wenn bei DynamoDB-Tabellen Multi-AZ-Verfügbarkeit automatisch vom Service bereitgestellt wird, können Sie vorhandene Tabellen problemlos in globale Tabellen konvertieren, um mehrere Regionen nutzen zu können.

      i. Konvertieren von Amazon DynamoDB-Tabellen für eine Region in globale Tabellen

   c. Wenn Ihr Workload hinter Application Load Balancers oder Network Load Balancers liegt, verwenden Sie AWS Global Accelerator, um die Verfügbarkeit Ihrer Anwendung zu verbessern, indem Sie Datenverkehr zu mehreren Regionen mit fehlerfreien Endpunkten leiten.

      i. Endpunkte für Standard-Accelerators in AWS Global Accelerator – AWS Global Accelerator (amazon.com)

   d. Erwägen Sie bei Anwendungen, die AWS EventBridge nutzen, die Verwendung von regionsübergreifenden Buses, um Ereignisse an ausgewählte Regionen weiterzuleiten.

      i. Senden und Empfangen von Amazon EventBridge-Ereignissen zwischen AWS-Regionen

   e. Ziehen Sie bei Amazon Aurora-Datenbanken globale Aurora-Datenbanken in Erwägungen, die mehrere AWS-Regionen umfassen können. Vorhandene Cluster können ebenfalls geändert werden, um neue Regionen hinzuzufügen.

      i. Erste Schritte mit globalen Amazon Aurora-Datenbanken

   f. Wenn Ihr Workload AWS Key Management Service-Verschlüsselungsschlüssel (AWS KMS) umfasst, überlegen Sie, ob Multi-Region-Schlüssel für Ihre Anwendung geeignet sind.

      i. Multi-Region-Schlüssel in AWS KMS

   g. Weitere Funktionen von AWS-Services finden Sie in dieser Blog-Reihe zum Erstellen einer Multi-Region-Anwendung mit AWS-Services

Grad des Aufwands für den Implementierungsplan: Mittel bis hoch

Ressourcen

Ähnliche Dokumente:

- Erstellen einer Multi-Region-Anwendung mit AWS-Services
- Disaster Recovery (DR) Architecture on AWS, Part IV: Multi-site Active/Active (Architektur für die Notfallwiederherstellung (Disaster Recovery, DR) in AWS, Teil IV: Multi-Site Aktiv-Aktiv)
- Globale AWS-Infrastruktur
- AWS Local Zones – häufig gestellte Fragen
- Architektur für die Notfallwiederherstellung in AWS, Teil I: Strategien für die Wiederherstellung in der Cloud
- Die Notfallwiederherstellung in der Cloud unterscheidet sich
- Globale Tabellen: Multiregionale Replikation mit DynamoDB

Relevante Videos:

- AWS re:Invent 2018: Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen (ARC209-R2)
- Auth0: multiregionale Architektur mit hoher Verfügbarkeit, die auf mehr als 1,5 Milliarden Anmeldungen pro Monat mit automatisiertem Failover skaliert werden kann.

Ähnliche Beispiele:

- Architektur für die Notfallwiederherstellung in AWS, Teil I: Strategien für die Wiederherstellung in der Cloud
- DTCC erzielt Resilienz weit über das hinaus, was On-Premises möglich wäre
- Expedia Group nutzt eine Architektur mit mehreren Regionen und Availability Zones und einem proprietären DNS-Service, um den Anwendungen Resilienz hinzuzufügen.
- Uber: Notfallwiederherstellung für multiregionales Kafka
- Netflix: Aktiv-Aktiv für multiregionale Resilienz
- Entwicklung von Data Residency für Atlassian Cloud
- Intuit TurboTax wird über zwei Regionen ausgeführt

REL10-BP03 Automatisierte Wiederherstellung für Komponenten, die auf einen einzelnen Standort beschränkt sind

Wenn Komponenten des Workloads nur in einer einzelnen Availability Zone oder einem On-Premises-Rechenzentrum ausgeführt werden können, müssen Sie die Funktion implementieren, um eine vollständige Neuerstellung des Workloads innerhalb festgelegter Wiederherstellungsziele durchzuführen.

Wenn die bewährte Methode zur Bereitstellung des Workloads an mehreren Standorten aufgrund technologischer Einschränkungen nicht möglich ist, müssen Sie einen alternativen Pfad zur Ausfallsicherheit implementieren. Sie müssen die Möglichkeit automatisieren, die erforderliche Infrastruktur neu zu erstellen, Anwendungen neu bereitzustellen und die erforderlichen Daten für diese Fälle neu zu erstellen.

Amazon EMR startet beispielsweise alle Knoten für einen bestimmten Cluster in derselben Availability Zone, da die Ausführung eines Clusters in derselben Zone eine höhere Datenzugriffsrate bietet und dadurch eine höhere Leistung für die Aufgabenbearbeitung bereitstellt. Wenn diese Komponente für die Ausfallsicherheit von Workloads erforderlich ist, müssen Sie die Möglichkeit haben, den Cluster und seine Daten erneut bereitzustellen. Für Amazon EMR sollten Sie nicht nur Multi-AZs verwenden, um für Redundanz zu sorgen. Sie können mehrere Knoten bereitstellen. Mit EMR File System (EMRFS)können Daten in EMR in Amazon S3 gespeichert und dann über mehrere Availability Zones oder AWS-Regionen repliziert werden.

Ähnlich wie bei Amazon Redshift wird Ihr Cluster standardmäßig in einer zufällig ausgewählten Availability Zone innerhalb der ausgewählten AWS-Region bereitgestellt. Alle Cluster-Knoten werden in derselben Zone bereitgestellt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Implementieren Sie Selbstreparatur. Stellen Sie Ihre Instances oder Container nach Möglichkeit mit automatischer Skalierung bereit. Wenn dies nicht möglich ist, nutzen Sie für EC2-Instances die automatische Wiederherstellung oder implementieren Sie eine automatische Selbstreparatur basierend auf Amazon EC2- oder ECS-Container-Lebenszyklusereignissen.
  - Verwenden Sie Auto-Scaling-Gruppen für Instances und Container-Workloads, die keine IP-Adresse für eine einzelne Instance, keine private IP-Adresse, keine elastische IP-Adresse und keine Instance-Metadaten benötigen.
    - Was ist EC2 Auto Scaling?

    - Automatische Skalierung von Services
        - Die Benutzerdaten der Startkonfiguration können für die Automatisierung der Selbstreparatur der meisten Workloads verwendet werden.
- Verwenden Sie die automatische Wiederherstellung von EC2-Instances für Workloads, die eine IP-Adresse für eine einzelne Instance, eine private IP-Adresse, eine elastische IP-Adresse und Instance-Metadaten benötigen.
    - Stellen Sie Ihre Instance wieder her.
        - Automatic Recovery sendet Benachrichtigungen zum Wiederherstellungsstatus an ein SNS-Thema, wenn der Instance-Fehler erkannt wird.
- Verwenden Sie EC2-Instance-Lebenszyklusereignisse bzw. ECS-Ereignisse für die Automatisierung der Selbstreparatur, wenn die automatische Skalierung oder EC2-Wiederherstellung nicht verwendet werden kann.
    - Lebenszyklus-Hooks für Amazon EC2 Auto Scaling
    - Amazon ECS-Events
        - Verwenden Sie die Ereignisse, um die Automatisierung der Reparatur der Komponente entsprechend der erforderlichen Prozesslogik aufzurufen.

Ressourcen

Ähnliche Dokumente:

- Amazon ECS-Events
- Lebenszyklus-Hooks für Amazon EC2 Auto Scaling
- Stellen Sie Ihre Instance wieder her.
- Automatische Skalierung von Services
- Was ist EC2 Auto Scaling?

REL10-BP04 Verwenden von Bulkhead-Architekturen, um den Umfang von Beeinträchtigungen zu begrenzen

Wie Schotten auf einem Schiff stellt dieses Bulkhead-Muster sicher, dass ein Fehler auf eine kleine Teilmenge von Anfragen oder Clients eingeschränkt bleibt. So wird die Anzahl der beeinträchtigten Anfragen begrenzt und die meisten Anfragen können fehlerfrei ausgeführt werden. Bulkheads

für Daten werden häufig als Partitionen bezeichnet, während Bulkheads für Services als Zellen bezeichnet werden.

In einer zellenbasierten Architekturist jede Zelle eine vollständige, unabhängige Instance des Service und hat eine feste maximale Größe. Mit zunehmender Last wachsen die Workloads, indem weitere Zellen hinzugefügt werden. Bei eingehendem Datenverkehr wird mit einem Partitionsschlüssel ermittelt, welche Zelle die Anfrage verarbeitet. Jeder Fehler beschränkt sich auf die Zelle, in der er auftritt, sodass die Anzahl der beeinträchtigten Anfragen begrenzt ist, da andere Zellen weiterhin fehlerfrei funktionieren. Es ist wichtig, den richtigen Partitionsschlüssel zu identifizieren, um zellenübergreifende Interaktionen zu minimieren und zu verhindern, dass bei jeder Anfrage komplexe Zuordnungsservices berücksichtigt werden müssen. Services, die komplexe Zuordnungen erfordern, führen nur zu einer Verlagerung des Problems auf die Zuordnungsservices, während Services, für die zellenübergreifende Interaktionen erforderlich sind, Abhängigkeiten zwischen den Zellen schaffen (und damit die angenommenen Verfügbarkeitsverbesserungen reduzieren).

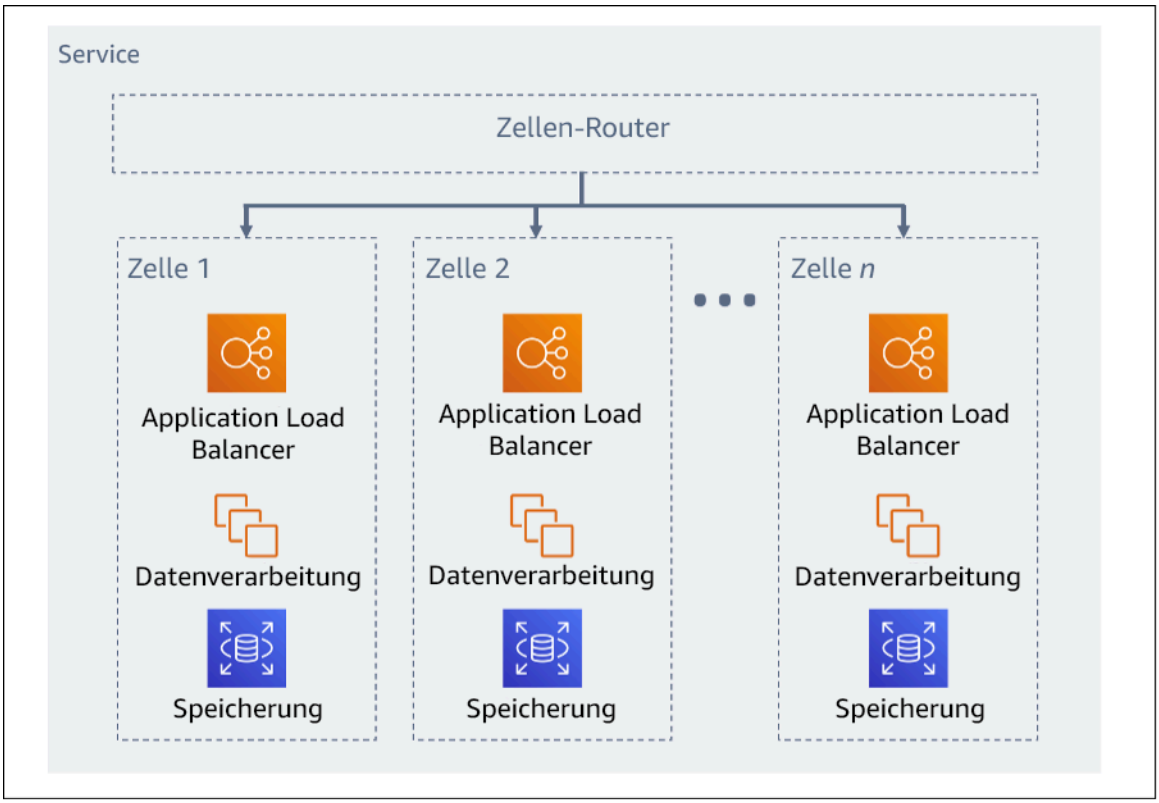


Abbildung 11: Zellenbasierte Architektur

Colm MacCarthaigh erläutert in seinem AWS-Blogbeitrag, wie Amazon Route 53 das Konzept des Shuffle Sharding nutzt, um Kundenanfragen in Shards zu isolieren. Ein Shard besteht in diesem Fall aus mindestens zwei Zellen. Auf der Basis des Partitionsschlüssels wird der Datenverkehr von einem Kunden (oder von Ressourcen, je nachdem, was Sie isolieren möchten) an den zugewiesenen

Shard weitergeleitet. Bei acht Zellen mit zwei Zellen pro Shard und Kunden, die auf die vier Shards aufgeteilt sind, sind im Falle eines Problems 25 % der Kunden betroffen.

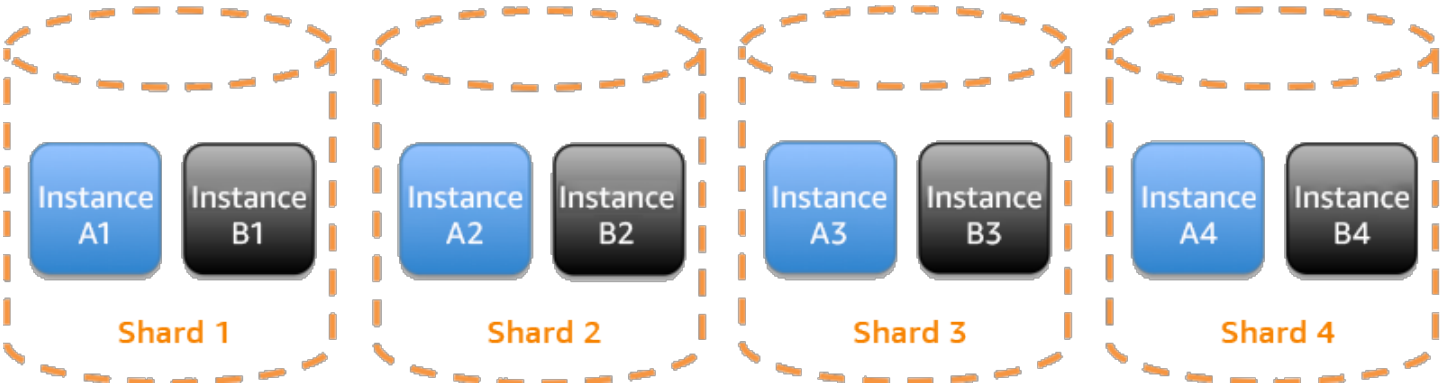


Abbildung 12: Service, der in vier herkömmliche Shards mit je zwei Zellen aufgeteilt ist

Mit Shuffle Sharding erstellen Sie virtuelle Shards mit jeweils zwei Zellen und weisen Ihre Kunden einem dieser virtuellen Shards zu. Wenn ein Problem auftritt, können Sie zwar trotzdem ein Viertel des gesamten Service verlieren, aber die Art der Kunden- oder Ressourcenzuweisung sorgt dafür, dass der Umfang der Auswirkungen durch Shuffle Sharding deutlich kleiner ausfällt als 25 %. Bei acht Zellen gibt es 28 eindeutige Kombinationen von zwei Zellen, was bedeutet, dass es 28 mögliche Shuffle Shards (virtuelle Shards) gibt. Wenn Sie Hunderte oder Tausende von Kunden haben und jeden Kunden einem Shuffle Shard zuweisen, beträgt der Umfang der Auswirkungen aufgrund eines Problems nur 1/28. Das ist siebenmal besser als beim regulären Sharding.

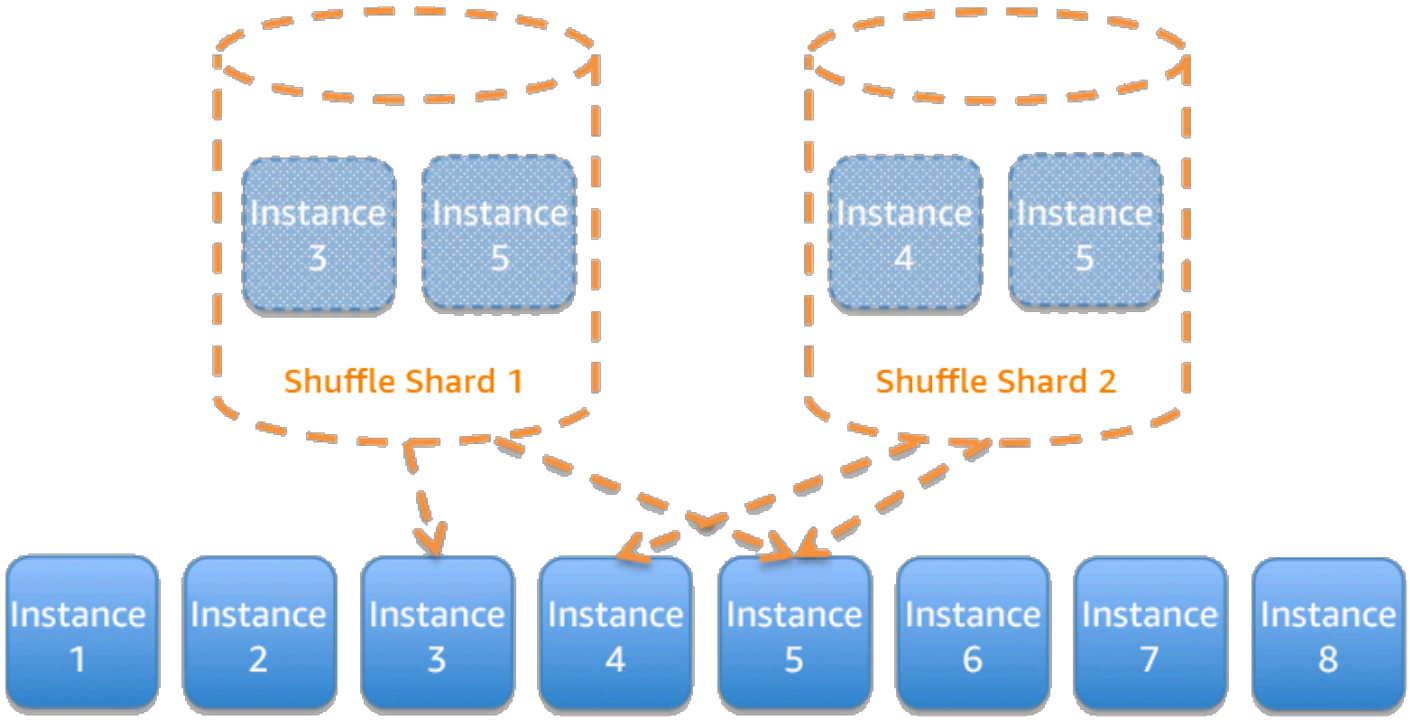


Abbildung 13: Service unterteilt in 28 Shuffle Shards (virtuelle Shards) mit jeweils zwei Zellen (nur zwei Shuffle Shards der 28 möglichen Shards werden gezeigt)

Ein Shard kann zusätzlich zu den Zellen für Server, Warteschlangen oder andere Ressourcen verwendet werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Verwenden Sie Bulkhead-Architekturen. Wie Schotten auf einem Schiff stellt dieses Bulkhead-Muster sicher, dass ein Fehler auf eine kleine Teilmenge von Anfragen oder Benutzern eingeschränkt bleibt. So wird die Anzahl der beeinträchtigten Anfragen begrenzt und die meisten Anfragen können fehlerfrei ausgeführt werden. Bulkheads für Daten werden häufig als Partitionen bezeichnet, während Bulkheads für Services als Zellen bezeichnet werden.
  - Well-Architected Lab: Fehlerisolierung mit Shuffle Sharding
  - Shuffle Sharding: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)
  - AWS re:Invent 2018: So minimiert AWS den Wirkungsradius von Fehlern (ARC338)
- Evaluieren Sie eine zellenbasierte Architektur für Ihren Workload. In einer zellenbasierten Architektur ist jede Zelle eine vollständige, unabhängige Instance des Service und hat eine feste maximale Größe. Mit zunehmender Last wachsen die Workloads, indem weitere Zellen hinzugefügt werden. Bei eingehendem Datenverkehr wird mit einem Partitionsschlüssel ermittelt, welche Zelle die Anfrage verarbeitet. Jeder Fehler beschränkt sich auf die Zelle, in der er auftritt, sodass die Anzahl der beeinträchtigten Anfragen begrenzt ist, da andere Zellen weiterhin fehlerfrei funktionieren. Es ist wichtig, den richtigen Partitionsschlüssel zu identifizieren, um zellenübergreifende Interaktionen zu minimieren und zu verhindern, dass bei jeder Anfrage komplexe Zuordnungsservices berücksichtigt werden müssen. Services, die komplexe Zuordnungen erfordern, führen nur zu einer Verlagerung des Problems auf die Zuordnungsservices, während Services, für die zellenübergreifende Interaktionen erforderlich sind, die Autonomie von Zellen (und damit die angenommenen Verfügbarkeitsverbesserungen) reduzieren.
  - Colm MacCarthaigh beschreibt in seinem Beitrag zum AWS-Blog, wie Amazon Route 53 das Konzept des Shuffle Sharding nutzt, um Kundenanfragen in Shards zu isolieren.
    - Shuffle Sharding: massive und magische Fehlerisolierung

Ressourcen

Ähnliche Dokumente:

- Shuffle Sharding: massive und magische Fehlerisolierung
- Die Amazon Builders' Library: Workload-Isolation mit Shuffle Sharding

Relevante Videos:

- AWS re:Invent 2018: So minimiert AWS den Wirkungsradius von Fehlern (ARC338)
- Shuffle Sharding: AWS re:Invent 2019: Einführung in die Amazon Builders’ Library (DOP328)

Ähnliche Beispiele:

- Well-Architected Lab: Fehlerisolierung mit Shuffle Sharding

ZUV 11 Wie lassen sich Workloads so gestalten, dass sie Komponentenausfälle verkraften?

Workloads, die eine hohe Verfügbarkeit und eine niedrige mittlere Wiederherstellungszeit (Mean Time To Recovery, MTTR) benötigen, müssen auf Resilienz ausgelegt sein.

Bewährte Methoden

- REL11-BP01 Überwachen aller Komponenten der Workload auf Fehler
- REL11-BP02 Failover zu fehlerfreien Ressourcen
- REL11-BP03 Automatisieren der Reparatur auf allen Ebenen
- REL11-BP04 Nutzen der Datenebene und nicht der Steuerebene während der Wiederherstellung
- REL11-BP05 Verhindern von bimodalem Verhalten mithilfe statischer Stabilität
- REL11-BP06 Senden von Benachrichtigungen, wenn sich Ereignisse auf die Verfügbarkeit auswirken

REL11-BP01 Überwachen aller Komponenten der Workload auf Fehler

Überwachen Sie den Zustand Ihrer Workload kontinuierlich, damit Sie und die automatisierten Systeme eine Verschlechterung oder einen Ausfall umgehend bemerken. Überwachen Sie Key

Performance Indicators (KPIs, wichtige Leistungskennzahlen) auf Grundlage des geschäftlichen Wertes.

Alle Wiederherstellungs- und Reparaturmechanismen müssen auf eine schnelle Erkennung von Problemen ausgelegt sein. Technische Fehler sollten zuerst erkannt werden, damit sie behoben werden können. Die Verfügbarkeit basiert jedoch auf der Fähigkeit Ihrer Workload, einen Unternehmenswert zu liefern. Daher müssen wichtige Leistungskennzahlen (KPIs), die dies messen, in Ihre Erkennungs- und Behebungsstrategie integriert sein.

Gängige Antimuster:

- Es sind keine Alarme konfiguriert, sodass Ausfälle ohne Benachrichtigung auftreten.
- Alarme sind vorhanden, aber mit Schwellenwerten, die keine ausreichende Zeit für die Reaktion bieten.
- Metriken werden nicht häufig genug erfasst, um das Recovery Time Objective (RTO, Wiederherstellungsdauer) zu erreichen.
- Nur die kundenseitige Ebene der Workload wird aktiv überwacht.
- Es werden nur technische Metriken erfasst, keine Metriken für Geschäftsfunktionen.
- Es gibt keine Metriken, die die Benutzererfahrung der Workload messen.

Vorteile der Einführung dieser bewährten Methode: Wenn alle Ebenen entsprechend überwacht werden, können Sie die Wiederherstellungszeit durch eine schnellere Fehlererkennung verkürzen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Bestimmen Sie das Erfassungsintervall für die Komponenten auf Grundlage Ihrer Wiederherstellungsziele.
  - Das Überwachungsintervall hängt davon ab, wie schnell Wiederherstellungen durchgeführt werden müssen. Die Wiederherstellungszeit hängt davon ab, wie viel Zeit für eine Wiederherstellung benötigt wird. Daher müssen Sie die Häufigkeit der Erfassung bestimmen, indem Sie diese Zeit und das RTO einkalkulieren.
- Konfigurieren Sie eine detaillierte Überwachung für die Komponenten.
  - Legen Sie fest, ob eine detaillierte Überwachung für EC2-Instances und Auto Scaling erforderlich ist. Die detaillierte Überwachung stellt Metriken in 1-Minuten-Intervallen bereit; die Standardüberwachung stellt Metriken in 5-Minuten-Intervallen bereit.

    - Aktivieren oder deaktivieren Sie die detaillierte Überwachung für Ihre Instance
    - Überwachen Ihrer Auto-Scaling-Gruppen und Instances mit Amazon CloudWatch.
  - Legen Sie fest, ob eine erweiterte Überwachung für RDS notwendig ist. Die erweiterte Überwachung verwendet einen Agenten in den RDS-Instances, um nützliche Informationen zu verschiedenen Prozessen oder Threads in einer RDS-Instance abzurufen.
    - Enhanced Monitoring
- Erstellen Sie benutzerdefinierte Metriken, um Leistungskennzahlen (KPIs) zu messen. Mit Workloads werden wichtige Geschäftsfunktionen implementiert. Diese Funktionen sollten als KPIs verwendet werden, um die Identifizierung indirekter Probleme zu unterstützen.
  - Veröffentlichen benutzerdefinierter Metriken
- Überwachen Sie das Benutzererlebnis auf Fehler mithilfe von Benutzer-Canaries. Synthetische Transaktionstests (auch bekannt als „Canary-Tests“, die aber nicht mit Canary-Bereitstellungen zu verwechseln sind), mit denen das Kundenverhalten simuliert werden kann, gehören zu den wichtigsten Testprozessen. Führen Sie diese Tests für Ihre Workload-Endpunkte konstant von verschiedenen Remote-Standorten aus.
  - Amazon CloudWatch Synthetics unterstützt Sie bei der Erstellung von Benutzer-Canaries.
- Erstellen Sie benutzerdefinierte Metriken zur Nachverfolgung des Benutzererlebnisses. Wenn Sie das Kundenerlebnis instrumentieren können, können Sie die Verschlechterung des Kundenerlebnisses feststellen.
  - Veröffentlichen benutzerdefinierter Metriken
- Richten Sie Alarme ein, um zu erkennen, wenn ein Teil Ihrer Workload nicht ordnungsgemäß funktioniert, und um anzugeben, wann Ressourcen automatisch skaliert werden müssen. Alarme können visuell in Dashboards angezeigt werden, Warnungen per Amazon SNS oder E-Mail senden und mit Auto Scaling die Ressourcen für eine Workload auf- oder abzuskalieren.
  - Verwenden von Amazon CloudWatch-Alarmen
- Erstellen Sie Dashboards, um Ihre Metriken zu visualisieren. Dashboards können verwendet werden, um Trends, Ausreißer und andere Indikatoren für potenzielle Probleme zu visualisieren, und auf Probleme hinweisen, die Sie untersuchen sollten.
  - Verwenden von CloudWatch-Dashboards

Ressourcen

Relevante Dokumente:

- Amazon CloudWatch Synthetics unterstützt Sie bei der Erstellung von Benutzer-Canaries.
- Aktivieren oder deaktivieren Sie die detaillierte Überwachung für Ihre Instance
- Enhanced Monitoring
- Überwachen Ihrer Auto-Scaling-Gruppen und Instances mit Amazon CloudWatch.
- Veröffentlichen benutzerdefinierter Metriken
- Verwenden von Amazon CloudWatch-Alarmen
- Verwenden von CloudWatch-Dashboards

Ähnliche Beispiele:

- Well-Architected Lab: Level 300: Implementieren von Zustandsprüfungen und Verwalten von Abhängigkeiten zur Verbesserung der Zuverlässigkeit

REL11-BP02 Failover zu fehlerfreien Ressourcen

Stellen Sie sicher, dass fehlerfreie Ressourcen weiterhin Anforderungen erfüllen können, wenn ein Ressourcenausfall auftritt. Stellen Sie bei Standortausfällen (z. B. einer Availability Zone oder AWS-Region) sicher, dass Sie Failover zu fehlerfreien Ressourcen an nicht beeinträchtigten Standorten eingerichtet haben.

AWS-Services wie Elastic Load Balancing und AWS Auto Scaling helfen dabei, Lasten über verschiedene Ressourcen und Availability Zones hinweg zu verteilen. Daher können der Ausfall einer einzelnen Ressource (wie etwa einer EC2-Instance) oder die Beeinträchtigung einer Availability Zone gemindert werden, indem Datenverkehr verlagert wird, um Ressourcen fehlerfrei zu halten. Bei Workloads mit mehreren Regionen ist dies komplizierter. Regionsübergreifende Lesereplikate ermöglichen Ihnen beispielsweise die Bereitstellung Ihrer Daten in mehreren AWS-Regionen. Sie müssen die Lesereplikate jedoch als primär hochstufen und Ihren Datenverkehr bei einem Failover darauf verweisen. Amazon Route 53 und AWS Global Accelerator können dabei helfen, Datenverkehr über AWS-Regionen zu leiten.

Wenn in Ihrer Workload AWS-Services wie Amazon S3 oder Amazon DynamoDB verwendet werden, werden diese automatisch in mehreren Availability Zones bereitgestellt. Bei einem Ausfall leitet die AWS-Steuerebene den Datenverkehr automatisch an fehlerfreie Standorte weiter. Die Daten werden redundant in mehreren Availability Zones gespeichert und bleiben verfügbar. Für Amazon RDS müssen Sie Multi-AZ als Konfigurationsoption auswählen. Bei einem Ausfall leitet AWS den Datenverkehr dann automatisch an die fehlerfreie Instance weiter. Für Amazon EC2-Instances,

Amazon ECS-Aufgaben oder Amazon EKS-Pods wählen Sie aus, in welchen Availability Zones die Bereitstellung erfolgen soll. Elastic Load Balancing bietet dann die Lösung, um Instances in fehlerhaften Zonen zu erkennen und den Datenverkehr an die fehlerfreien Zonen weiterzuleiten. Elastic Load Balancing kann den Datenverkehr sogar an Komponenten in Ihrem On-Premises-Rechenzentrum weiterleiten.

Für multiregionale Ansätze (zu denen auch On-Premises-Rechenzentren gehören können) bietet Amazon Route 53 eine Möglichkeit, Internetdomänen zu definieren und Routing-Richtlinien zuzuweisen, die Zustandsprüfungen enthalten können. So wird sichergestellt, dass der Datenverkehr an fehlerfreie Regionen weitergeleitet wird. Alternativ stellt AWS Global Accelerator statische IP-Adressen bereit, die als fester Einstiegspunkt in Ihre Anwendung dienen, und sorgt für eine Weiterleitung an Endpunkte in AWS-Regionen Ihrer Wahl. Dabei wird anstelle des Internets das globale AWS-Netzwerk verwendet, das mehr Leistung und Zuverlässigkeit bietet.

Beim Design der Services berücksichtigt AWS immer die Wiederherstellung nach einem Fehler. Wir konzipieren Services mit dem Ziel, die Wiederherstellungszeit nach Ausfällen und die Auswirkungen auf Daten zu minimieren. Unsere Services verwenden primär Datenspeicher, die Anfragen erst akzeptieren, nachdem sie dauerhaft auf mehreren Replikaten in einer Region gespeichert wurden. Zu diesen Services und Ressourcen gehören Amazon Aurora, Amazon Relational Database Service (Amazon RDS) Multi-AZ-DB-Instances, Amazon S3, Amazon DynamoDB, Amazon Simple Queue Service (Amazon SQS) und Amazon Elastic File System (Amazon EFS). Sie sind so aufgebaut, dass sie eine zellenbasierte Isolation und die Fehlerisolierung von Availability Zones nutzen. In unseren betrieblichen Abläufen setzen wir sehr stark auf Automatisierung. Außerdem optimieren wir unsere Funktionalität für Ersetzungsvorgänge und Neustarts, um nach Unterbrechungen eine schnelle Wiederherstellung zu ermöglichen.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Failover zu fehlerfreien Ressourcen. Stellen Sie sicher, dass fehlerfreie Ressourcen weiterhin Anforderungen erfüllen können, wenn ein Ressourcenausfall auftritt. Stellen Sie bei Standortausfällen (z. B. einer Availability Zone oder AWS-Region) sicher, dass Sie Failover zu fehlerfreien Ressourcen an nicht beeinträchtigten Standorten eingerichtet haben.
  - Wenn in Ihrer Workload AWS-Services wie Amazon S3 oder Amazon DynamoDB verwendet werden, werden diese automatisch in mehreren Availability Zones bereitgestellt. Bei einem Ausfall leitet die AWS-Steuerebene den Datenverkehr automatisch an fehlerfreie Standorte weiter.

- Für Amazon RDS müssen Sie Multi-AZ als Konfigurationsoption auswählen. Bei einem Ausfall leitet AWS den Datenverkehr dann automatisch an die fehlerfreie Instance weiter.
  - Hochverfügbarkeit (Multi-AZ) für Amazon RDS
- Für Amazon EC2-Instances oder Amazon ECS-Aufgaben wählen Sie aus, in welchen Availability Zones die Bereitstellung erfolgen soll. Elastic Load Balancing bietet dann die Lösung, um Instances in fehlerhaften Zonen zu erkennen und den Datenverkehr an die fehlerfreien Zonen weiterzuleiten. Elastic Load Balancing kann den Datenverkehr sogar an Komponenten in Ihrem On-Premise-Rechenzentrum weiterleiten.
- Bei multiregionalen Ansätzen (die auch On-Premises-Rechenzentren einschließen können) sollten Sie sicherstellen, dass Daten und Ressourcen an fehlerfreien Standorten weiterhin Anforderungen erfüllen können.
  - Regionsübergreifende Lesereplikate ermöglichen Ihnen beispielsweise die Bereitstellung Ihrer Daten in mehreren AWS-Regionen. Sie müssen die Lesereplikate jedoch hochstufen, um den Datenverkehr zu steuern und weiterzuleiten, wenn der primäre Standort ausfüllt.
    - Arbeiten mit Lesereplikaten
  - Amazon Route 53 ermöglicht die Definition von Internetdomänen und die Zuweisung von Routing-Richtlinien, die Zustandsprüfungen enthalten können. So wird sichergestellt, dass der Datenverkehr an fehlerfreie Regionen weitergeleitet wird. Alternativ stellt AWS Global Accelerator statische IP-Adressen bereit, die als fester Einstiegspunkt in Ihre Anwendung dienen, und sorgt für eine Weiterleitung an Endpunkte in AWS-Regionen Ihrer Wahl. Dabei wird anstelle des öffentlichen Internets das globale AWS-Netzwerk verwendet, das mehr Leistung und Zuverlässigkeit bietet.
    - Amazon Route 53: Auswählen einer Routing-Richtlinie
    - Was ist AWS Global Accelerator?

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Automatisierung der Fehlertoleranz unterstützen können
- AWS Marketplace: Zur Erzielung von Fehlertoleranz geeignete Produkte
- AWS OpsWorks: Verwenden von Auto Healing zum Austausch fehlgeschlagener Instances
- Amazon Route 53: Auswählen einer Routing-Richtlinie
- Hochverfügbarkeit (Multi-AZ) für Amazon RDS

- Arbeiten mit Lesereplikaten
- Strategien für die Platzierung von Aufgaben in Amazon ECS
- Creating Kubernetes Auto Scaling Groups for Multiple Availability Zones (Erstellen von Kubernetes-Auto-Scaling-Gruppen für mehrere Availability Zones)
- Was ist AWS Global Accelerator?

Ähnliche Beispiele:

- Well-Architected Lab: Level 300: Implementieren von Zustandsprüfungen und Verwalten von Abhängigkeiten zur Verbesserung der Zuverlässigkeit

REL11-BP03 Automatisieren der Reparatur auf allen Ebenen

Verwenden Sie bei Erkennung eines Fehlers automatisierte Funktionen, um Maßnahmen zur Behebung durchzuführen.

Die Möglichkeit zum Neustart ist ein wichtiges Tool zur Behebung von Fehlern. Wie zuvor für verteilte Systeme beschrieben, besteht eine bewährte Methode darin, Services nach Möglichkeit zustandslos zu machen. Dadurch wird der Verlust von Daten oder Verfügbarkeit beim Neustart verhindert. In der Cloud können (und sollten) Sie die gesamte Ressource (z. B. eine EC2-Instance oder Lambda-Funktion) im Rahmen des Neustarts ersetzen. Der Neustart selbst ist eine einfache und zuverlässige Methode zur Wiederherstellung nach einem Ausfall. Bei Workloads treten viele verschiedene Arten von Fehlern auf. Fehler können sich auf Hardware, Software, Kommunikation und den Betrieb beziehen. Statt neue Mechanismen zu entwickeln, um die verschiedenen Fehlertypen zu erfassen, zu identifizieren und zu korrigieren, sollten Sie viele verschiedene Fehlerkategorien derselben Wiederherstellungsstrategie zuordnen. Instances können aufgrund von Hardware- oder Betriebssystemfehlern, aufgrund von unzureichendem Speicher oder aus anderen Gründen ausfallen. Anstatt eine benutzerdefinierte Fehlerbehebung für jede Situation zu entwickeln, sollten Sie alle Szenarios als Instance-Ausfälle behandeln. Beenden Sie die Instance und lassen Sie sie durch AWS Auto Scaling ersetzen. Die ausgefallene Ressource können Sie genauer untersuchen, nachdem sie außer Betrieb genommen wurde.

Ein weiteres Beispiel ist die Möglichkeit, eine Netzwerkanfrage neu zu starten. Nutzen Sie denselben Wiederherstellungsansatz für eine Netzwerk-Zeitüberschreitung und einen Abhängigkeitsfehler, bei dem die Abhängigkeit einen Fehler ausgibt. Beide Ereignisse wirken sich in ähnlicher Weise auf das System aus. Statt also zu versuchen, aus den einzelnen Ereignissen einen "Sonderfall" zu

konstruieren, sollten Sie eine ähnliche Strategie anwenden und versuchen, einen exponentiellen Backoff mit Jitter durchzuführen.

Die Möglichkeit zum Neustart ist ein Wiederherstellungsmechanismus, der in Recovery Oriented Computing und Cluster-Architekturen mit hoher Verfügbarkeit verwendet wird.

Mit Amazon EventBridge lassen sich Ereignisse wie CloudWatch-Alarme oder Statusänderungen in anderen AWS-Services überwachen und filtern. Anhand der Ereignisinformationen kann der Service anschließend AWS Lambda, AWS Systems Manager-Automation oder andere Ziele auslösen, um für Ihre Workload eine benutzerdefinierte Korrekturlogik auszuführen.

Amazon EC2 Auto Scaling kann dafür konfiguriert werden, den Zustand der EC2-Instance zu prüfen. Wenn sich die Instance nicht im ausgeführten Status befindet oder der Systemstatus beeinträchtigt ist, betrachtet Amazon EC2 Auto Scaling die Instance als fehlerhaft und startet eine Ersatz-Instance. Wenn Sie AWS OpsWorks verwenden, können Sie Auto Healing für EC2-Instances auf der OpsWorks-Layer-Ebene konfigurieren.

Für umfangreiche Ersetzungen (z. B. beim Verlust einer gesamten Availability Zone) ist statische Stabilität die bevorzugte Methode, um für hohe Verfügbarkeit zu sorgen, statt mehrere neue Ressourcen gleichzeitig abzurufen.

Gängige Antimuster:

- Einzelne Bereitstellung von Anwendungen in Instances oder Containern.
- Bereitstellen von Anwendungen, die nicht ohne automatische Wiederherstellung an mehreren Standorten bereitgestellt werden können.
- Manuelle Reparatur von Anwendungen, die sich mit Auto Scaling und einer automatischen Wiederherstellung nicht reparieren lassen.

Vorteile der Einführung dieser bewährten Methode: Selbst wenn die Workload jeweils nur an einem Standort bereitgestellt werden kann, verkürzt die automatisierte Reparatur die durchschnittliche Zeit bis zur Wiederherstellung und stellt die Verfügbarkeit der Workload sicher.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Stellen Sie die Ebenen in einer Workload mithilfe von Auto-Scaling-Gruppen bereit. Die automatische Skalierung kann Selbstreparaturen für zustandslose Anwendungen ausführen sowie Kapazitäten hinzufügen oder entfernen.

    - Funktionsweise von Skalierungsplänen
- Implementieren Sie eine automatische Wiederherstellung für EC2-Instances, in denen Anwendungen bereitgestellt werden, die nicht an mehreren Standorten bereitgestellt werden können, und die einen Neustart nach Ausfällen tolerieren. Mithilfe der automatischen Wiederherstellung kann ausgefallene Hardware ersetzt und die Instance neu gestartet werden, wenn die Anwendung sich nicht an mehreren Standorten bereitstellen lässt. Die Metadaten der Instance und die zugehörigen IP-Adressen werden beibehalten, ebenso wie die Amazon EBS-Volumes und Bindungspunkte für Elastic File Systems oder Dateisysteme für Lustre und Windows.
    - Automatische Wiederherstellung in Amazon EC2
    - Amazon Elastic Block Store (Amazon EBS)
    - Amazon Elastic File System (Amazon EFS)
    - Was ist Amazon FSx für Lustre?
    - Was ist Amazon FSx für Windows File Server?
        - Wenn Sie AWS OpsWorks verwenden, können Sie Auto Healing für EC2-Instances auf Layer-Ebene konfigurieren.
            - AWS OpsWorks: Verwenden von Auto Healing zum Austausch fehlgeschlagener Instances
- Implementieren Sie die automatisierte Wiederherstellung mit AWS Step Functions und AWS Lambda, wenn keine automatische Skalierung oder Wiederherstellung möglich ist oder die automatische Wiederherstellung fehlschlägt. Wenn Sie keine automatische Skalierung verwenden können und die automatische Wiederherstellung entweder nicht genutzt werden kann oder fehlschlägt, können Sie die Reparatur mithilfe von AWS Step Functions und AWS Lambda automatisieren.
    - Was ist AWS Step Functions?
    - Was ist AWS Lambda?
        - Mit Amazon EventBridge lassen sich Ereignisse wie CloudWatch-Alarme oder Statusänderungen in anderen AWS-Services überwachen und filtern. Anhand der Ereignisinformationen kann der Service anschließend AWS Lambda (oder andere Ziele) auslösen, um eine benutzerdefinierte Korrekturlogik für die Workload auszuführen.
            - Was ist Amazon EventBridge?
            - Verwenden von Amazon CloudWatch-Alarmen

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Automatisierung der Fehlertoleranz unterstützen können
- AWS Marketplace: Zur Erzielung von Fehlertoleranz geeignete Produkte
- AWS OpsWorks: Verwenden von Auto Healing zum Austausch fehlgeschlagener Instances
- Automatische Wiederherstellung in Amazon EC2
- Amazon Elastic Block Store (Amazon EBS)
- Amazon Elastic File System (Amazon EFS)
- Funktionsweise von Skalierungsplänen
- Verwenden von Amazon CloudWatch-Alarmen
- Was ist Amazon EventBridge?
- Was ist AWS Lambda?
- AWS Systems Manager Automation
- Was ist AWS Step Functions?
- Was ist Amazon FSx für Lustre?
- Was ist Amazon FSx für Windows File Server?

Relevante Videos:

- Statische Stabilität in AWS: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)

Ähnliche Beispiele:

- Well-Architected Lab: Level 300: Implementieren von Zustandsprüfungen und Verwalten von Abhängigkeiten zur Verbesserung der Zuverlässigkeit

REL11-BP04 Nutzen der Datenebene und nicht der Steuerebene während der Wiederherstellung

Die Steuerebene wird für die Konfigurierung von Ressourcen verwendet. Die Datenebene stellt Services bereit. Datenebenen besitzen in der Regel höhere Ziele in Bezug auf das Verfügbarkeitsdesign als Steuerebenen und sind in der Regel weniger komplex. Bei der Implementierung von Wiederherstellungs- oder Eingrenzungsantworten auf Ereignisse, die sich potenziell auf die Resilienz auswirken könnten, kann durch die Verwendung von Operationen auf Steuerebene die Gesamtresilienz Ihrer Architektur reduziert werden. Sie können beispielsweise die

Amazon Route 53-Datenebene nutzen, um DNS-Abfragen auf der Basis von Zustandsprüfungen zuverlässig weiterzuleiten. Da bei der Aktualisierung von Route 53-Routing-Richtlinien jedoch die Steuerebene verwendet wird, sollten Sie diese nicht für Wiederherstellungen verwenden.

Die Route 53-Datenebenen beantworten DNS-Abfragen, führen Zustandsprüfungen durch und bewerten diese. Sie werden global für ein 100%-iges Service Level Agreement (SLA) verteilt und entworfen. Die Route 53-Management-APIs und -Konsolen, in denen Sie Route 53-Ressourcen erstellen, aktualisieren und löschen können, werden auf Steuerebenen ausgeführt. Diese Ebenen sind darauf ausgelegt, die starke Konsistenz und Stabilität zu priorisieren, die Sie bei der Verwaltung von DNS benötigen. Zu diesem Zweck befinden sich die Steuerebenen in einer einzelnen Region, US East (N. Virginia). Beide Systeme sind zwar äußerst zuverlässig, aber die Steuerebenen sind nicht in der SLA enthalten. In seltenen Fällen kann es vorkommen, dass das ausfallsichere Design der Datenebene es ermöglicht, die Verfügbarkeit aufrechtzuerhalten, während die Steuerebene dies nicht tut. Verwenden Sie für die Notfallwiederherstellung und Failover-Mechanismen Datenebenen-Funktionen, um die bestmögliche Zuverlässigkeit bereitzustellen.

Weitere Informationen über Datenebenen, Steuerebenen und wie AWS Services aufbaut, um Hochverfügbarkeitsziele zu erfüllen, finden Sie im Dokument Statische Stabilität mithilfe von Availability Zones und in der Amazon Builders' Library.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Nutzen Sie die Datenebene statt der Steuerebene, wenn Sie Amazon Route 53 für die Notfallwiederherstellung verwenden. Route 53 Application Recovery Controller hilft Ihnen, anhand von Bereitschaftsprüfungen und Routing-Steuerung Failover-Vorgänge zu verwalten und zu koordinieren. Diese Funktionen überwachen kontinuierlich die Fähigkeit Ihrer Anwendung, nach Fehlern wiederhergestellt zu werden, und ermöglichen Ihnen die Steuerung der Anwendungswiederherstellung über mehrere AWS-Regionen, Availability Zones und On-Premises.
  - Was ist Route 53 Application Recovery Controller?
  - Erstellen von Mechanismen für die Notfallwiederherstellung mit Amazon Route 53
  - Entwickeln hoch resilienter Anwendungen mit Amazon Route 53 Application Recovery Controller, Teil 1: Stack für eine einzelne Region
  - Entwickeln hoch resilienter Anwendungen mit Amazon Route 53 Application Recovery Controller, Teil 2: Stack für eine mehrere Regionen
- Erfahren Sie, welche Operationen auf der Datenebene und welche Operationen auf der Steuerebene ausgeführt werden.

- Amazon Builders' Library: Vermeiden von Überlastungen verteilter Systeme durch Übernahme der Steuerung durch den kleineren Service
- Amazon DynamoDB API (Steuerebene und Datenebene)
- AWS Lambda-Ausführungen (in Steuerebene und Datenebene unterteilt)
- AWS Lambda-Ausführungen (in Steuerebene und Datenebene unterteilt)

Ressourcen

Relevante Dokumente:

- APN-Partner: Partner, die Sie bei der Automatisierung der Fehlertoleranz unterstützen können
- AWS Marketplace: Zur Erzielung von Fehlertoleranz geeignete Produkte
- Amazon Builders' Library: Vermeiden von Überlastungen verteilter Systeme durch Übernahme der Steuerung durch den kleineren Service
- Amazon DynamoDB API (Steuerebene und Datenebene)
- AWS Lambda-Ausführungen (in Steuerebene und Datenebene unterteilt)
- AWS-Elemental-MediaStore-Datenebene
- Entwickeln hoch resilienter Anwendungen mit Amazon Route 53 Application Recovery Controller, Teil 1: Stack für eine einzelne Region
- Entwickeln hoch resilienter Anwendungen mit Amazon Route 53 Application Recovery Controller, Teil 2: Stack für eine mehrere Regionen
- Erstellen von Mechanismen für die Notfallwiederherstellung mit Amazon Route 53
- Was ist Route 53 Application Recovery Controller?

Ähnliche Beispiele:

- What is Amazon Route 53 Application Recovery Controller? (Was ist Amazon Route 53 Application Recovery Controller?)

REL11-BP05 Verhindern von bimodalem Verhalten mithilfe statischer Stabilität

Bimodales Verhalten bedeutet, dass eine Workload im normalen Modus und im Fehlermodus unterschiedliche Verhaltensweisen zeigt, indem sie z. B. bei Ausfall einer Availability Zone neue Instances startet. Stattdessen sollten Sie Workloads erstellen, die statisch stabil sind und nur in

einem Modus betrieben werden. In diesem Fall sollten Sie genügend Instances in jeder Availability Zone bereitstellen, damit die Verarbeitung der Workload auch beim Entfernen einer Availability Zone gewährleistet ist. Anschließend sollten Sie die beeinträchtigten Instances mithilfe von Elastic Load Balancing oder Amazon Route 53-Zustandsprüfungen entlasten.

Statische Stabilität für die Bereitstellung von Rechenleistung (z. B. EC2-Instances oder -Container) führt zu höchster Zuverlässigkeit. Dabei müssen Sie das Kosten-Nutzen-Verhältnis abwägen. Es ist kostengünstiger, weniger Rechenkapazität bereitzustellen und sich bei einem Ausfall auf das Starten neuer Instances zu verlassen. Bei großen Ausfällen (z. B. einem Ausfall einer Availability Zone) ist dieser Ansatz jedoch weniger effektiv, da er sich darauf stützt, auf bereits eingetretene Beeinträchtigungen zu reagieren, statt auf diese Beeinträchtigungen vorbereitet zu sein, bevor sie auftreten. Ihre Lösung sollte die Zuverlässigkeits- und Kostenanforderungen für Ihre Workload berücksichtigen. Wenn Sie eine größere Anzahl von Availability Zones verwenden, verringert sich die Menge der zusätzlichen Rechenleistung, die Sie für die statische Stabilität benötigen.

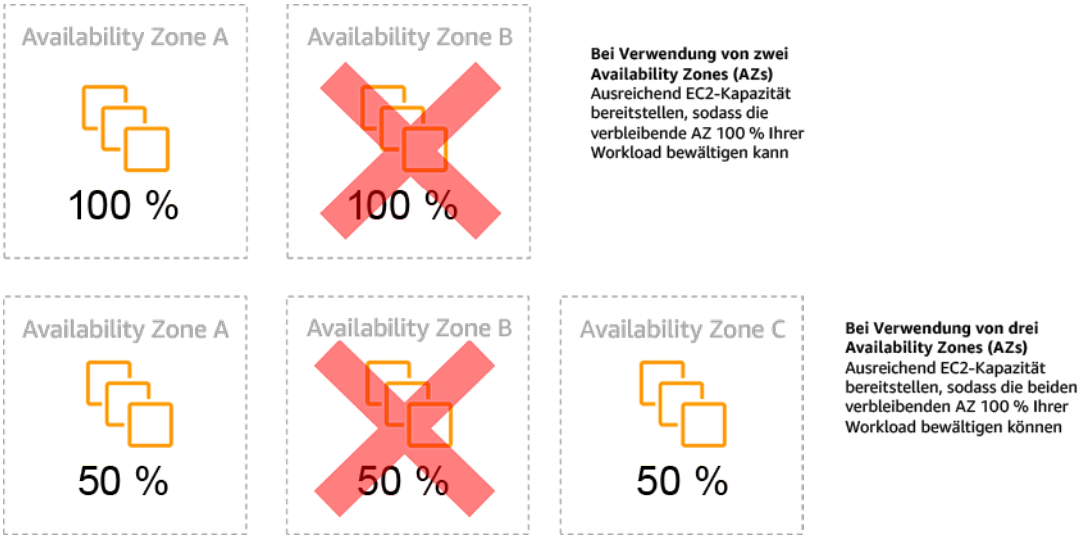


Abbildung 14: Statische Stabilität von EC2-Instances in Availability Zones

Nachdem der Datenverkehr verlagert wurde, können Sie AWS Auto Scaling verwenden, um Instances in der ausgefallenen Zone asynchron zu ersetzen und sie in den fehlerfreien Zonen zu starten.

Ein weiteres Beispiel für bimodales Verhalten ist eine Netzwerk-Zeitüberschreitung, die dazu führen kann, dass ein System versucht, den Konfigurationsstatus des gesamten Systems zu aktualisieren. Dies kann zu einer unerwarteten Belastung einer anderen Komponente führen, die daraufhin ausfallen könnte und möglicherweise weitere unerwartete Konsequenzen nach sich zieht. Diese negative Feedback-Schleife wirkt sich auf die Verfügbarkeit Ihrer Workload aus. Deshalb sollten Sie Systeme erstellen, die statisch stabil sind und nur in einem Modus betrieben

werden. Ein statisch stabiles Design besteht aus konstanter Arbeit und einer regelmäßigen Aktualisierung des Konfigurationsstatus. Wenn ein Aufruf fehlschlägt, verwendet die Workload den zuvor zwischengespeicherten Wert und löst einen Alarm aus.

Ein weiteres Beispiel für bimodales Verhalten: Sie lassen zu, dass Clients im Fehlerfall den Workload-Cache umgehen. Dies scheint eine Lösung zu sein, die Clientanforderungen erfüllt, sollte aber nicht zugelassen werden, da sie die Belastung Ihrer Workload erheblich ändert und wahrscheinlich zu Fehlern führt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Nutzen Sie statische Stabilität, um bimodales Verhalten zu verhindern. Bimodales Verhalten bedeutet, dass eine Workload im normalen Modus und im Fehlermodus unterschiedliche Verhaltensweisen zeigt, indem sie z. B. bei Ausfall einer Availability Zone neue Instances startet.
  - Minimierung der Abhängigkeiten bei der Planung der Notfallwiederherstellung
  - Die Amazon Builders' Library: Statische Stabilität durch Availability Zones
  - Statische Stabilität in AWS: AWS re:Invent 2019: Einführung in die Amazon Builders’ Library (DOP328)
    - Sie sollten stattdessen Systeme erstellen, die statisch stabil sind und nur in einem einzigen Modus ausgeführt werden. In diesem Fall sollten Sie genügend Instances in jeder Zone bereitstellen, damit die Verarbeitung der Workload auch beim Entfernen einer AZ gewährleistet ist, und verwenden Sie anschließend Elastic Load Balancing oder Amazon Route 53-Zustandsprüfungen, um die Last von den beeinträchtigten Instances wegzuverlagern.
    - Ein weiteres Beispiel für bimodales Verhalten: Sie lassen zu, dass Clients im Fehlerfall den Workload-Cache umgehen. Dies mag zwar wie eine praktikable Lösung zur Erfüllung der Clientanforderungen aussehen, sollte aber vermieden werden, da sie die Ansprüche an die Workload erheblich verändert und wahrscheinlich zu Fehlern führt.

Ressourcen

Relevante Dokumente:

- Minimierung der Abhängigkeiten bei der Planung der Notfallwiederherstellung
- Die Amazon Builders' Library: Statische Stabilität durch Availability Zones

Relevante Videos:

- Statische Stabilität in AWS: AWS re:Invent 2019: Einführung in die Amazon Builders' Library (DOP328)

REL11-BP06 Senden von Benachrichtigungen, wenn sich Ereignisse auf die Verfügbarkeit auswirken

Benachrichtigungen werden nach Erkennung wichtiger Ereignisse gesendet, auch wenn das durch das Ereignis verursachte Problem automatisch behoben wurde.

Auto Healing sorgt dafür, dass Ihre Workload zuverlässig ist. Allerdings können dadurch auch zugrunde liegende Probleme verschleiert werden, die behoben werden müssen. Implementieren Sie geeignete Überwachungsfunktionen und Ereignisse, damit Sie Problemmuster erkennen können, einschließlich solcher, die durch Auto Healing behoben werden. Auf diese Weise können Sie die Fehlerursachen beheben. Amazon CloudWatch-Alarme können basierend auf auftretenden Fehlern ausgelöst werden. Sie können auch basierend auf Aktionen der automatischen Fehlerbehebung ausgelöst werden. CloudWatch-Alarme können so konfiguriert werden, dass E-Mails gesendet oder Vorfälle mithilfe der Amazon SNS-Integration in Drittanbietersystemen zur Nachverfolgung von Vorfällen protokolliert werden.

Gängige Antimuster:

- Senden von Alarmen, auf die niemand reagiert.
- Durchführen automatischer Reparaturen ohne die Benachrichtigung, dass eine Reparatur erforderlich war.

Vorteile der Einführung dieser bewährten Methode: Benachrichtigungen zu Wiederherstellungen sorgen dafür, dass Sie selten auftretende Probleme nicht ignorieren.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Alarme für wichtige geschäftliche Leistungskennzahlen, wenn diese eine niedrige Schwelle überschreiten. Wenn Sie eine niedrige Alarmschwelle für Ihre geschäftlichen KPIs ansetzen, können Sie besser erkennen, wann Ihre Workload nicht verfügbar ist oder nicht funktioniert.
  - Erstellen eines CloudWatch-Alarms auf der Basis eines statischen Schwellenwerts
- Alarme für Ereignisse, die eine automatisierte Reparatur auslösen. Sie können eine SNS-API direkt aufrufen, um bei von Ihnen erstellten Automatisierungen Benachrichtigungen zu senden.

- Was ist Amazon Simple Notification Service?

Ressourcen

Relevante Dokumente:

- Erstellen eines CloudWatch-Alarms auf der Basis eines statischen Schwellenwerts
- Was ist Amazon EventBridge?
- Was ist Amazon Simple Notification Service?

ZUV 12 Wie lässt sich die Zuverlässigkeit testen?

Nachdem Sie Ihre Workload so konzipiert haben, dass sie den Belastungen der Produktion standhält, sind Tests die einzige Möglichkeit, sie auf die erwartete Funktionalität und Ausfallsicherheit hin zu testen.

Bewährte Methoden

- REL12-BP01 Untersuchen von Fehlern mit Playbooks:
- REL12-BP02 Durchführen von Analysen nach Vorfällen
- REL12-BP03 Testen funktionaler Anforderungen
- REL12-BP04 Testen von Skalierungs- und Leistungsanforderungen
- REL12-BP05 Testen der Ausfallsicherheit mit Chaos-Engineering
- REL12-BP06 Regelmäßiges Abhalten von Gamedays

REL12-BP01 Untersuchen von Fehlern mit Playbooks:

Ermöglichen Sie konsistente und schnelle Antworten auf noch unbekannte Fehlerszenarien, indem Sie den Untersuchungsprozess in Playbooks dokumentieren. Playbooks sind vordefinierte Abläufe zum Identifizieren der Faktoren, die zu einem Fehlerszenario beitragen. Die Ergebnisse aus jedem Prozessschritt sind die Grundlage für die nächsten Schritte. Nach diesem Muster wird vorgegangen, bis das Problem identifiziert oder eskaliert wird.

Das Playbook ist eine proaktive Planung, die für effektive Reaktionen erforderlich ist. Wenn nicht vom Playbook abgedeckte Fehlerszenarien in der Produktion auftreten, beheben Sie zunächst das Problem. Analysieren Sie danach die unternommenen Schritte und verwenden Sie diese, um einen neuen Eintrag im Playbook hinzuzufügen.

Beachten Sie, dass Playbooks als Reaktion auf bestimmte Vorfälle verwendet werden, während Runbooks verwendet werden, um bestimmte Ergebnisse zu erzielen. Häufig werden Runbooks für Routineaktivitäten verwendet, Playbooks hingegen, um auf außergewöhnliche Ereignisse zu reagieren.

Gängige Antimuster:

- Planen der Bereitstellung eines Workloads, ohne die Prozesse für die Diagnose von Problemen oder die Reaktion auf Vorfälle zu kennen.
- Ungeplante Entscheidungen darüber, in welchen Systemen bei der Untersuchung von Ereignissen Protokolle und Metriken erfasst werden sollen.
- Metriken und Ereignisse werden nicht lange genug aufbewahrt, um die Daten abrufen zu können.

Vorteile der Einführung dieser bewährten Methode: Durch das Erfassen von Playbooks wird sichergestellt, dass Prozesse konsistent befolgt werden können. Ihre Playbooks werden als Code festgehalten, um die Entstehung von Fehlern durch manuelle Aktivitäten zu reduzieren. Durch die Automatisierung von Playbooks kann schneller auf Ereignisse reagiert werden, weil Teammitglieder nicht eingreifen müssen oder ihnen vor dem Eingreifen zusätzliche Informationen zur Verfügung gestellt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Ermitteln von Probleme mit Playbooks. Playbooks sind dokumentierte Prozesse für die Untersuchung von Problemen. Durch die Dokumentation der Prozesse in Playbooks schaffen Sie die Voraussetzung für eine einheitliche und schnelle Reaktion auf Fehlerszenarien. Playbooks müssen die Informationen und Anleitungen enthalten, die eine entsprechend qualifizierte Person zum Zusammentragen sachdienlicher Informationen, zum Identifizieren möglicher Fehlerursachen, zum Isolieren von Fehlern und zum Bestimmen beitragender Faktoren (zum Analysieren nach einem Vorfall) benötigt.
  - Implementieren von Playbooks als Code. Führen Sie Ihre Operationen als Code aus, indem Sie Skripts für Ihre Playbooks erstellen, um Konsistenz sicherzustellen und Fehler zu reduzieren, die durch manuelle Prozesse verursacht werden. Playbooks können aus mehreren Skripts bestehen, die die verschiedenen Schritte darstellen, die erforderlich sein können, um die zu einem Problem beitragenden Faktoren zu identifizieren. Runbook-Aktivitäten können ausgelöst oder im Rahmen von Playbook-Aktivitäten ausgeführt werden. Sie können auch als Antwort auf identifizierte Ereignisse die Ausführung eines Playbooks auslösen.

- Automatisieren Sie Ihre operativen Playbooks mit AWS Systems Manager
- AWS Systems Manager Befehl ausführen
- AWS Systems Manager Automation
- Was ist AWS Lambda?
- Was ist Amazon EventBridge?
- Verwenden von Amazon CloudWatch Alarmen

Ressourcen

Zugehörige Dokumente:

- AWS Systems Manager Automation
- AWS Systems Manager Befehl ausführen
- Automatisieren Sie Ihre operativen Playbooks mit AWS Systems Manager
- Verwenden von Amazon CloudWatch Alarmen
- Verwenden von Canaries (Amazon CloudWatch Synthetics)
- Was ist Amazon EventBridge?
- Was ist AWS Lambda?

Ähnliche Beispiele:

- Automatisieren von Vorgängen mit Playbooks und Runbooks

REL12-BP02 Durchführen von Analysen nach Vorfällen

Überprüfen Sie die Ereignisse mit Auswirkungen auf Kunden und bestimmen Sie die beitragenden Faktoren und Präventivmaßnahmen. Entwickeln Sie anhand dieser Informationen Abhilfemaßnahmen, um ein wiederholtes Auftreten nach Möglichkeit zu verhindern. Entwickeln Sie Verfahren für schnelle und effektive Reaktionen. Informieren Sie nach Bedarf auf zielgruppengerechte Weise über beitragende Faktoren und Korrekturmaßnahmen. Legen Sie eine Kommunikationsmethode fest, um andere bei Bedarf über die Ursachen zu informieren.

Bewerten Sie, warum bestehende Tests das Problem nicht gefunden haben. Fügen Sie Tests für diesen Fall hinzu, wenn noch keine Tests vorhanden sind.

Gängige Antimuster:

- Beitragende Faktoren werden ermittelt, es wird jedoch nicht weiter nach anderen potenziellen Problemen und Lösungsansätzen gesucht.
- Es werden nur menschliche Fehlerursachen ermittelt, es wird aber keine Schulung oder Automatisierung bereitgestellt, die menschliche Fehler verhindern könnte.

Vorteile der Einführung dieser bewährten Methode: Durch Analysen von Vorfällen und das Teilen von Ergebnissen können die Risiken für andere Workloads mit den gleichen beitragenden Faktoren die Risiken verringert werden. Außerdem können Abhilfemaßnahmen oder automatisierte Wiederherstellungen implementiert werden, bevor es zu einem Vorfall kommt.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Festlegen eines Standards für Analysen nach Vorfällen. Durch gute Analysen nach Vorfällen lassen sich allgemeine Lösungen für Probleme mit Architekturmustern ermitteln, die Sie bereits an anderer Stelle in den Systemen anwenden.
  - Sorgen Sie dafür, dass die beitragenden Faktoren auf ehrliche Weise und ohne Schuldzuweisungen aufgeführt werden.
  - Wenn Sie Probleme nicht dokumentieren, können Sie sie auch nicht korrigieren.
    - Verzichten Sie bei Analysen nach Vorfällen auf Schuldzuweisungen, damit Sie die Korrekturmaßnahmen unparteiisch vorschlagen können. Fördern Sie zudem in Ihren Anwendungsteams eine ehrliche Selbstbewertung und Zusammenarbeit.
- Verwenden eines Prozesses zur Ermittlung beitragender Faktoren. Erarbeiten Sie ein Verfahren, um die beitragenden Faktoren eines Ereignisses zu ermitteln und zu dokumentieren. Damit können Sie Abhilfemaßnahmen entwickeln, um ein erneutes Auftreten einzudämmen oder gänzlich zu verhindern, und Verfahren für eine rasche und wirksame Reaktion erstellen. Kommunizieren Sie beitragende Faktoren wie zutreffend, jeweils auf die Zielgruppen ausgerichtet.
  - Was ist Protokollanalytik?

Ressourcen

Zugehörige Dokumente:

- Was ist Protokollanalytik?
- Darum sollten Sie eine Fehlerkorrektur (COE) entwickeln

REL12-BP03 Testen funktionaler Anforderungen

Verwenden Sie Techniken wie Komponenten- und Integrationstests, mit denen die erforderliche Funktionalität validiert wird.

Im Idealfall sollten diese Tests automatisch als Teil von Build- und Bereitstellungsaktionen ausgeführt werden. Mit AWS CodePipeline übergeben Entwickler beispielsweise Änderungen an ein Quell-Repository, in dem CodePipeline die Änderungen automatisch erkennt. Diese Änderungen werden vorgenommen und Tests werden ausgeführt. Nachdem die Tests abgeschlossen sind, wird der erstellte Code für Tests auf Staging-Servern bereitgestellt. Auf dem Staging-Server führt CodePipeline weitere Tests aus, z. B. Integrations- oder Belastungstests. Nach dem erfolgreichen Abschluss dieser Tests stellt CodePipeline den getesteten und genehmigten Code für Produktions-Instances bereit.

Außerdem zeigen frühere Erfahrungen, dass synthetische Transaktionstests (auch bekannt als Canary-Tests, aber nicht zu verwechseln mit Canary-Bereitstellungen), die ausgeführt werden können und das Kundenverhalten simulieren, zu den wichtigsten Testprozessen gehören. Führen Sie diese Tests für Ihre Workload-Endpunkte konstant von verschiedenen Remote-Standorten aus. Mit Amazon CloudWatch Synthetics können Sie Canaries erstellen, um Ihre Endpunkte und APIs zu überwachen.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Testen funktionaler Anforderungen: Dazu gehören Komponenten- und Integrationstests, mit denen die erforderliche Funktionalität validiert wird.
  - Verwenden von AWS CodeBuild mit CodePipeline zum Testen von Code und zum Ausführen von Builds
  - AWS CodePipeline Adds Support for Unit and Custom Integration Testing with AWS CodeBuild (AWS CodePipeline fügt Unterstützung für Komponententests und angepasste Integrationstests mit AWS CodeBuild hinzu)
  - Kontinuierliche Bereitstellung und kontinuierliche Integration
  - Using synthetic monitoring (Amazon CloudWatch Synthetics) (Verwenden von synthetischer Überwachung (Amazon CloudWatch Synthetics))
  - Automatisierung von Softwaretests

Ressourcen

Zugehörige Dokumente:

- APN-Partner: Partner, die Sie bei der Implementierung einer Continuous Integration-Pipeline unterstützen können
- AWS CodePipeline Adds Support for Unit and Custom Integration Testing with AWS CodeBuild (AWS CodePipeline fügt Unterstützung für Komponententests und angepasste Integrationstests mit AWS CodeBuild hinzu)
- AWS Marketplace: Für die kontinuierliche Integration geeignete Produkte
- Kontinuierliche Bereitstellung und kontinuierliche Integration
- Automatisierung von Softwaretests
- Verwenden von AWS CodeBuild mit CodePipeline zum Testen von Code und zum Ausführen von Builds
- Using synthetic monitoring (Amazon CloudWatch Synthetics) (Verwenden von synthetischer Überwachung (Amazon CloudWatch Synthetics))

REL12-BP04 Testen von Skalierungs- und Leistungsanforderungen

Verwenden Sie Techniken wie Lasttests, um zu überprüfen, ob die Workload die Skalierungs- und Leistungsanforderungen erfüllt.

In der Cloud können Sie bei Bedarf eine Testumgebung für Ihren Workload in Produktionsumgebungen erstellen. Wenn Sie diese Tests auf einer herunterskalierten Infrastruktur ausführen, müssen Sie die Ergebnisse auf den Maßstab der Produktionsumgebung hochrechnen. Last- und Leistungstests können auch in der Produktion durchgeführt werden. Achten Sie dabei darauf, Benutzer nicht zu beeinträchtigen und Ihre Testdaten mit Tags zu versehen, sodass sie nicht mit Benutzerdaten vermischt werden und Nutzungsstatistiken oder Produktionsberichte verfälschen.

Stellen Sie mit Tests sicher, dass Ihre Basisressourcen, Skalierungseinstellungen, Servicekontingente und die Ausfallsicherheit unter Auslastung wie erwartet funktionieren.

Risikostufe, wenn diese Best Practice nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Testen Sie Skalierungs- und Leistungsanforderungen. Führen Sie Lasttests durch, um zu prüfen, ob der Workload die Skalierungs- und Leistungsanforderungen erfüllt.

- Verteilte Lasttests auf AWS: Simulation Tausender verbundener Benutzer
- Apache JMeter
  - Stellen Sie Ihre Anwendung in einer Umgebung bereit, die mit Ihrer Produktionsumgebung identisch ist, und führen Sie einen Lasttest durch.
    - Erstellen Sie auf Grundlage von "Infrastructure as Code"-Konzepten eine Umgebung, die Ihrer Produktionsumgebung möglichst ähnlich ist.

Ressourcen

Zugehörige Dokumente:

- Verteilte Lasttests auf AWS: Simulation Tausender verbundener Benutzer
- Apache JMeter

REL12-BP05 Testen der Ausfallsicherheit mit Chaos-Engineering

Führen Sie regelmäßig Chaos-Experimente in oder nahe an Produktionsumgebungen aus, um zu verstehen, wie Ihr System auf ungünstige Bedingungen reagiert.

Gewünschtes Ergebnis:

Die Ausfallsicherheit der Workload wird regelmäßig durch die Anwendung von Chaos-Engineering in Form von Fehlerinjektionsexperimenten oder einer Injektion unerwarteter Last überprüft. Dazu kommen Tests der Ausfallsicherheit, um das bekannte erwartete Verhalten der Workload während eines Ereignisses zu validieren. Kombinieren Sie Chaos-Engineering mit Tests der Ausfallsicherheit, um sicher zu sein, dass Ihre Workload Komponentenausfällen standhalten und sich von unerwarteten Unterbrechungen erholen kann – mit minimalen oder gar keinen Auswirkungen.

Typische Anti-Muster:

- Auslegung der Systeme auf Ausfallsicherheit, aber keine Überprüfung, wie die Workload als Ganzes funktioniert, wenn Fehler auftreten.
- Keine Experimente unter echten Bedingungen und der erwarteten Last.
- Keine Behandlung der Experimente als Code und fehlendes Aufrechterhalten während des Entwicklungszyklus.
- Keine Durchführung von Chaosexperimenten als Teil Ihrer CI/CD-Pipeline und außerhalb von Bereitstellungen.

- Keine Nutzung früherer Analysen nach Vorfällen bei der Entscheidung über die Fehler, mit denen experimentiert werden soll.

Vorteile der Nutzung dieser bewährten Methode: Durch die Injektion von Fehlern zur Überprüfung der Resilienz Ihres Workloads gewinnen Sie die nötige Zuversicht, dass die Wiederherstellungsverfahren Ihres resilienten Entwurfs im Fall eines realen Fehlers funktionieren.

Risikostufe bei fehlender Befolgung dieser Best Practice: Mittel

Implementierungsleitfaden

Das Chaos-Engineering bietet Ihren Teams die nötigen Chancen, um auf kontrollierte Weise kontinuierlich reale Störungen (Simulationen) auf Serviceanbieter-, Infrastruktur-, Workload- und Komponentenebene zu injizieren – mit nur minimalen oder gar keinen Auswirkungen auf Ihre Kunden. Ihre Teams können so aus Fehlern lernen und die Resilienz Ihrer Workloads beobachten, messen und verbessern. Darüber hinaus können sie überprüfen, ob Warnungen ausgelöst werden und die Teams über Ereignisse benachrichtigt werden.

Bei kontinuierlicher Ausführung kann das Chaos-Engineering Mängel in Ihren Workloads aufzeigen, die sich negativ auf Verfügbarkeit und Ausführung auswirken könnten, wenn sie nicht behoben werden.

> **Note**
>
> Beim Chaos-Engineering geht es um das Experimentieren mit einem System, um sich davon zu überzeugen, dass das System in der Produktion auch außergewöhnlichen Bedingungen standhalten kann. – Grundlagen des Chaos-Engineering

Wenn ein System diesen Disruptionen standhalten kann, sollte das Chaos-Experiment weiter als automatisierter Regressionstest ausgeführt werden. In dieser Form sollten Chaos-Experimente als Teil Ihres Systementwicklungszyklus (Systems Development Lifecycle, SDLC) und Ihrer CI/CD-Pipeline ausgeführt werden.

Um sicherzustellen, dass Ihr Workload resilient gegenüber dem Ausfall von Komponenten ist, sollten Sie im Rahmen Ihrer Experimente Ereignisse aus der Praxis injizieren. Sie könnten beispielsweise mit dem Verlust von Amazon EC2-Instances oder einem Failover der primären Amazon RDS-Datenbank-Instance experimentieren und so verifizieren, dass Ihr Workload nicht beeinträchtigt wird

(oder nur minimal beeinträchtigt wird). Mit einer Kombination von Komponentenfehlern könnten Sie Ereignisse simulieren, die von einer Disruption in einer Availability Zone verursacht werden könnten.

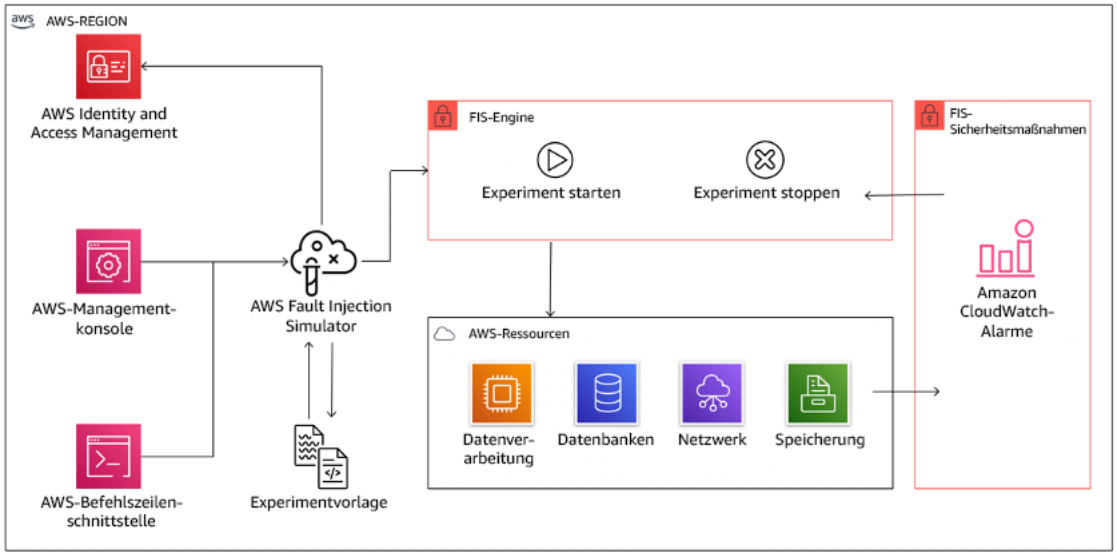
Hinsichtlich Fehlern auf Anwendungsebene (z. B. Abstürzen) könnten Sie mit Stressfaktoren wie Speicher- und CPU-Auslastung beginnen.

Zur Validierung von Fallback- oder Failover-Mechanismen für externe Abhängigkeiten, die bei zeitweisen Netzwerkdisruptionen ausgelöst werden, sollten Ihre Komponenten diese Ereignisse durch das Blockieren des Zugriffs auf externe Anbieter über einen bestimmten Zeitraum simulieren, der von wenigen Sekunden bis zu mehreren Stunden dauern kann.

Andere Degradierungsmodi führen möglicherweise zu einer reduzierten Funktionalität und zu verzögerten Reaktionen, was eine Disruption Ihrer Services verursachen kann. Bekannte Quellen für diese Degradierung sind eine erhöhte Latenz bei kritischen Services und eine unzuverlässige Netzwerkkommunikation (Verlust von Paketen). Experimente mit diesen Fehlern, darunter Netzwerkeffekten wie Latenz, Nachrichtenverlust und DNS-Ausfällen, könnten die fehlende Fähigkeit zur Auflösung eines Namens, zum Erreichen des DNS-Service oder zur Herstellung von Verbindungen zu abhängigen Services umfassen.

Chaos-Engineering-Tools:

AWS Fault Injection Service (AWS FIS) ist ein vollständig verwalteter Service für die Injektion von Fehlern, den Sie innerhalb oder außerhalb Ihrer CD-Pipeline verwenden können, um mit diesen Fehlern zu experimentieren. AWS FIS ist eine gute Wahl für Gamedays, die dem Chaos-Engineering gewidmet sind. Der Service unterstützt die gleichzeitige Injektion von Fehlern in verschiedene Arten von Ressourcen, darunter Amazon EC2, Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS) und Amazon RDS. Zu diesen Fehlern gehören die Beendigung von Ressourcen, die Erzwingung von Failovern, die Auslastung von CPU oder Arbeitsspeicher, Drosselung, Latenz und Paketverluste. Da dieser Service in Amazon CloudWatch Alarms integriert ist, können Sie Stoppbedingungen als Integritätsschutz einrichten, um Experimente rückgängig zu machen, wenn sie unerwartete Auswirkungen haben.



Diagramm, das die Integration von AWS Fault Injection Service in AWS-Ressourcen zeigt, um Ihnen die Ausführung von Fehlerinjektionsexperimenten für Ihre Workloads zu ermöglichen.

Es gibt auch verschiedene Drittanbieteroptionen für Fehlerinjektionsexperimente. Dazu gehören Open-Source-Tools wie Chaos Toolkit, Chaos Meshund Litmus Chaossowie kommerzielle Optionen wie Gremlin. Zur Erweiterung der Art der Fehler, die in AWS injiziert werden können, kann AWS FIS in Chaos Mesh und Litmus Chaos integriert werden.So können Sie Fehlerinjektions-Workflows über verschiedene Tools hinweg koordinieren. Sie können beispielsweise einen Stresstest für die CPU eines Pods mit Chaos-Mesh- oder Litmus-Fehlern ausführen und gleichzeitig einen zufällig ausgewählten Prozentsatz von Cluster-Knoten mit AWS FIS-Fehleraktionen beenden.

Implementierungsschritte

- Ermitteln Sie die Fehler, mit denen experimentiert werden soll.

  Bewerten Sie das Design Ihres Workloads in Bezug auf die Resilienz. Diese Designs (anhand der Best Practices des Well-Architected Frameworkerstellt) berücksichtigen Risiken im Zusammenhang mit kritischen Abhängigkeiten, früheren Ereignissen, bekannten Problemen und Compliance-Anforderungen. Listen Sie die einzelnen Elemente des Designs auf, die Resilienz zeigen sollen, und die Fehler, denen es standhalten soll. Weitere Informationen zur Erstellung dieser Listen finden Sie im Whitepaper zur Überprüfung der betrieblichen Bereitschaft. Dieses Whitepaper führt Sie durch die Entwicklung eines Prozesses zur Verhinderung der Wiederholung früherer Vorfälle. Der Prozess für die Analyse von Fehlerarten und ihren Auswirkungen (Failure Modes and Effects Analysis, FMEA) stellt Ihnen ein Framework für Fehleranalysen auf Komponentenebene und die Analyse der Auswirkungen dieser Fehler auf Ihren Workload bereit. FMEA wird von Adrian

Cockcroft in Failure Modes and Continuous Resilience(Fehlerarten und kontinuierliche Resilienz) detaillierter beschrieben.

- Weisen Sie jedem Fehler eine Priorität zu.

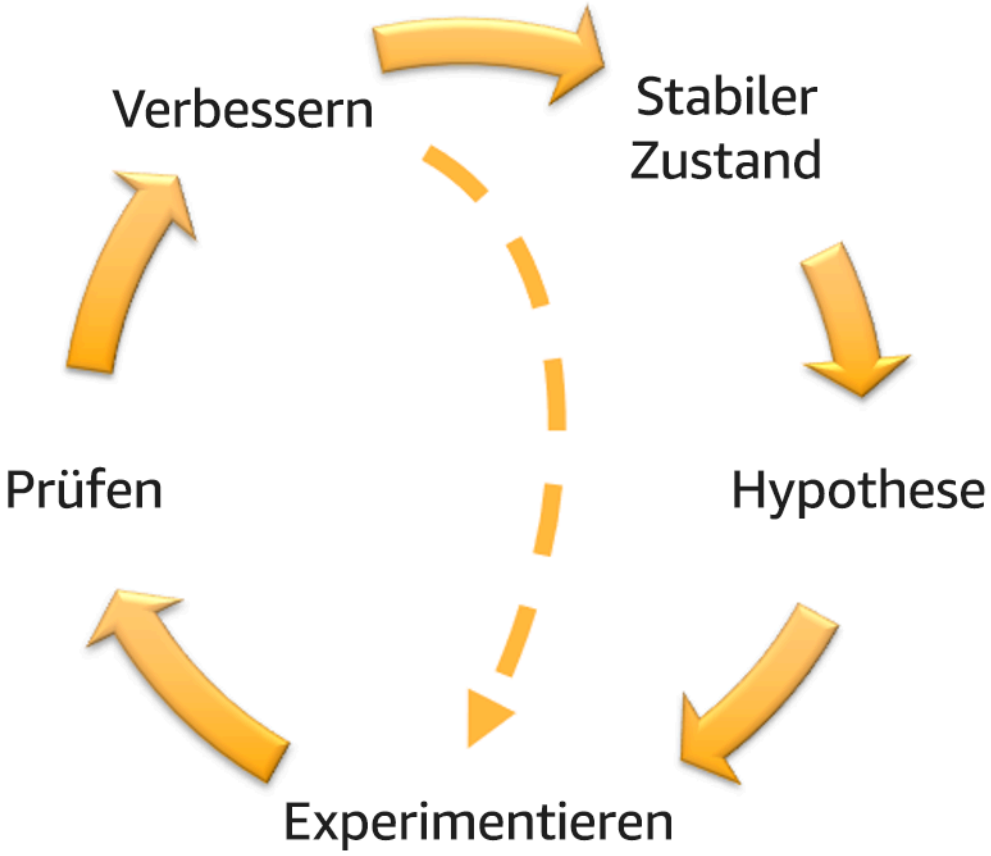
  Beginnen Sie mit einer groben Kategorisierung wie hoch, mittel oder niedrig. Berücksichtigen Sie bei der Festlegung der Priorität die Häufigkeit des Fehlers und die Auswirkungen des Fehlers auf den Workload insgesamt.

  Analysieren Sie hinsichtlich der Häufigkeit eines bestimmten Fehlers frühere Daten für den betreffenden Workload, wenn verfügbar. Wenn keine Daten verfügbar sind, verwenden Sie Daten zu anderen Workloads, die in einer ähnlichen Umgebung ausgeführt werden.

  Bei der Betrachtung der Auswirkungen eines bestimmten Fehlers gilt, dass die Auswirkungen im Allgemeinen umso größer sind, je größer der vom Fehler betroffene Bereich ist. Sie sollten auch das Design und den Zweck des Workloads berücksichtigen. Beispielsweise ist für einen Workload, der Daten transformiert und analysiert, der Zugriff auf die Quelldatenspeicher von kritischer Bedeutung. In diesem Fall würden Sie Experimente im Zusammenhang mit Zugriffsfehlern, Zugriffsdrosselungen und Latenzen priorisieren.

  Nach Vorfällen durchgeführte Analysen stellen eine gute Datenquelle dar, um Häufigkeit und Auswirkungen von Fehlerarten besser zu verstehen.

  Legen Sie anhand der zugewiesenen Priorität die Fehler fest, mit denen zuerst experimentiert werden soll, und die Reihenfolge, in der neue Fehlerinjektionsexperimente entwickelt werden sollen.
- Für jedes von Ihnen ausgeführte Experiment sollten Sie sich am Schwungrad für Chaos-Engineering und kontinuierliche Resilienz orientieren.



Schwungrad für Chaos-Engineering und kontinuierliche Resilienz unter Verwendung der wissenschaftlichen Methode von Adrian Hornsby.

- Definieren Sie den Steady-State als die messbare Ausgabe eines Workloads, der ein normales Verhalten zeigt.

  Ihr Workload befindet sich im Steady-State, wenn er zuverlässig und wie erwartet ausgeführt wird. Daher sollten Sie die Integrität Ihres Workloads überprüfen, bevor Sie den Steady-State definieren. Steady-State bedeutet nicht notwendigerweise, dass sich ein Fehler nicht auf den Workload auswirkt, da ein bestimmter Prozentsatz an Fehlern innerhalb akzeptabler Grenzen liegen könnte. Der Steady-State ist die Basislinie, die Sie während des Experiments beobachten. Diese wird Anomalien aufweisen, wenn Ihre Hypothese, die Sie im nächsten Schritt definieren, nicht die erwarteten Ergebnisse zeigt.

  Der Steady-State eines Zahlungssystems kann beispielsweise als die Verarbeitung von 300 TPS mit einer Erfolgsrate von 99 % und einer Roundtrip-Zeit von 500 ms definiert sein.

- Formulieren Sie eine Hypothese dazu, wie der Workload auf den Fehler reagieren wird.

  Eine gute Hypothese basiert darauf, wie der Workload den Fehler voraussichtlich bewältigt, um den Steady-State zu wahren. Die Hypothese besagt, dass bei einem Fehler eines spezifischen Typs das System oder der Workload weiter im Steady-State bleiben, da der Workload mit bestimmten Resilienzmerkmalen entworfen wurde. Der spezifische Fehlertyp und die Fehlerbewältigung sollten in der Hypothese angegeben werden.

  Sie können für die Hypothese die folgende Vorlage verwenden (andere Formulierungen sind jedoch auch akzeptabel):

  > **Note**
  >
  > Wenn *`(spezifischer Fehler)`* auftritt, wird der *`Workload`* (Name des Workloads) *`(Maßnahmen zur Bewältigung beschreiben),`* um die Auswirkungen auf *`geschäftliche oder technische Metriken einzudämmen`*.

  Beispiel:

  - Wenn 20 % der Knoten in der Amazon EKS-Knotengruppe ausfallen, wird die Transaction Create API das 99. Perzentil der Anforderungen weiter in weniger als 100 ms erfüllen (Steady-State). Die Amazon EKS-Knoten werden innerhalb von fünf Minuten wiederhergestellt und die Pods werden geplant und verarbeiten Traffic innerhalb von acht Minuten nach der Einleitung des Experiments. Warnungen werden innerhalb von drei Minuten ausgelöst.
  - Wenn eine einzelne Amazon EC2-Instance ausfällt, veranlasst die Elastic Load Balancing-Zustandsprüfung des Bestellsystems Elastic Load Balancing, Anforderungen ausschließlich an die noch intakten Instances zu senden, während Amazon EC2 Auto Scaling die ausgefallene Instance ersetzt. Dabei kommt es zu einer Steigerung der serverseitigen Fehler (5xx) um weniger als 0,01 % (Steady-State).
  - Wenn die primäre Amazon RDS-Datenbank-Instance ausfällt, führt der Workload für die Erfassung von Lieferkettendaten einen Failover aus und stellt eine Verbindung zur Amazon RDS-Standby-Datenbank-Instance her, sodass es für weniger als 1 Minute zu Lese- oder Schreibfehlern für die Datenbank kommt (Steady-State).
- Führen Sie das Experiment aus, indem Sie den Fehler injizieren.

  Ein Experiment sollte grundsätzlich nicht zu einem Ausfall führen und vom Workload toleriert werden. Wenn Sie wissen, dass der Workload ausfallen wird, sollten Sie das Experiment

nicht durchführen. Das Chaos-Engineering sollte verwendet werden, um bekannt-unbekannte oder unbekannt-unbekannte Ereignisse zu untersuchen. Bekannt-unbekannte Ereignisse sind Ereignisse, die Ihnen bekannt sind, die Sie jedoch nicht vollständig verstehen. Unbekannt-unbekannte Ereignisse sind Ereignisse, die Sie weder kennen noch vollständig verstehen. Wenn Sie Experimente für einen Workload ausführen, von dem Sie wissen, dass er fehlerhaft ist, werden Sie keine neuen Erkenntnisse gewinnen. Ihr Experiment sollte sorgfältig geplant sein, einen klaren Wirkungsumfang besitzen und einen Rollback-Mechanismus besitzen, der bei unerwarteten Störungen angewendet werden kann. Wenn eine sorgfältige Überprüfung zeigt, dass Ihr Workload das Experiment überstehen sollte, können Sie das Experiment starten. Für die Injektion von Fehlern gibt es verschiedene Optionen. Für AWS-Workloads stellt AWS FIS zahlreiche vordefinierte Fehlersimulationen bereit, die als Aktionenbezeichnet werden. Sie können auch angepasste Aktionen für AWS FIS definieren, die mithilfe von AWS Systems Manager-Dokumenten ausgeführt werden.

Wir raten davon ab, angepasste Skripts für Chaos-Experimente zu verwenden, es sei denn, die Skripts können den aktuellen Zustand des Workloads erkennen, können Protokolle ausgeben und stellen Rollback-Mechanismen und Stoppbedingungen bereit, soweit möglich.

Ein effektives Framework oder Toolset, das Chaos-Engineering unterstützt, sollte den aktuellen Status des Experiments nachverfolgen, Protokolle ausgeben und Rollback-Mechanismen bereitstellen, um eine kontrollierte Ausführung zu unterstützen. Beginnen Sie mit einem verbreitet verwendeten Service wie AWS FIS, der Ihnen die Ausführung von Experimenten mit einem klar definierten Umfang ermöglicht und Sicherheitsmechanismen bereitstellt, um ein Experiment rückgängig machen zu können, wenn es zu unerwarteten Störungen führt. Weitere Informationen zu Experimenten unter Verwendung von AWS FIS finden Sie im Resilient and Well-Architected Apps with Chaos Engineering Lab. Darüber hinaus analysiert AWS Resilience Hub Ihren Workload und erstellt Experimente, die Sie in AWS FIS implementieren und ausführen können.

> **Note**
>
> Sie sollten den Umfang und die Auswirkungen jedes Experiments genau verstehen. Wir empfehlen, Fehler zunächst in einer Nichtproduktionsumgebung zu simulieren, bevor sie in der Produktion ausgeführt werden.

Experimente sollten in der Produktion unter realen Bedingungen ausgeführt werden. Dabei sollten nach Möglichkeit Canary-Bereitstellungen verwendet werden, die sowohl ein

Kontrollsystem als auch ein Experimentsystem bereitstellen. Die Ausführung von Experimenten außerhalb von Spitzenzeiten stellt ein empfehlenswertes Verfahren dar, um potenzielle Auswirkungen zu reduzieren, wenn ein Experiment zum ersten Mal in der Produktion durchgeführt wird. Wenn die Verwendung von tatsächlichem Kunden-Traffic ein zu großes Risiko darstellt, können Sie unter Verwendung der Kontroll- und Experimentbereitstellungen Experimente mit synthetischem Traffic in der Produktionsinfrastruktur durchführen. Wenn ein Experiment nicht in der Produktion ausgeführt werden kann, führen Sie es in einer Präproduktionsumgebung aus, die der Produktionsumgebung so nahe wie möglich ist.

Sie müssen einen Integritätsschutz einrichten und überwachen, um sicherzustellen, dass sich das Experiment nicht jenseits akzeptabler Grenzen auf den Produktions-Traffic oder andere Systeme auswirkt. Richten Sie Stoppbedingungen ein, um ein Experiment anhalten zu können, wenn es in einer Integritätsschutz-Metrik einen von Ihnen definierten Schwellenwert erreicht. Diese Metriken sollten die Metrik für den Steady-State des Workloads und die Metrik für die Komponenten einschließen, in die Sie den Fehler injizieren. Die synthetische Überwachung (auch als Benutzer-Canary bezeichnet) gehört zu den Metriken, die Sie in der Regel als Benutzer-Proxy einschließen sollten. Stoppbedingungen für AWS FIS werden als Teil der Experimentvorlage unterstützt. Es sind bis zu fünf Stoppbedingungen pro Vorlage möglich.

Zu den Grundsätzen des Chaos-Engineering gehört die Minimierung von Umfang und Auswirkungen des Experiments:

Auch wenn einige kurzfristige negative Auswirkungen zulässig sein sollten, ist der Chaos-Engineer dafür verantwortlich, die Auswirkungen der Experimente zu minimieren und einzudämmen.

Eine Methode für die Überprüfung des Umfangs und der möglichen Auswirkungen besteht darin, das Experiment statt in der Produktionsumgebung zunächst in einer Nichtproduktionsumgebung durchzuführen. Dabei wird überprüft, ob die Schwellenwerte für Stoppbedingungen während des Experiments wie vorgesehen aktiviert werden und ob das Experiment beobachtet werden kann, um Ausnahmen abzufangen.

Wenn Sie Fehlerinjektionsexperimente durchführen, müssen alle verantwortlichen Beteiligten gut informiert sein. Teilen Sie den betroffenen Teams mit, wann die Experimente durchgeführt werden und was zu erwarten ist. Dies können Operations-Teams, die für die Servicezuverlässigkeit verantwortlichen Teams und der Kundensupport sein. Stellen Sie diesen Teams Kommunikationstools bereit, damit sie das Team, das das Experiment durchführt, über nachteilige Auswirkungen informieren können.

Sie müssen nach dem Experiment den Workload und die zugrunde liegenden Systeme wieder in den ursprünglichen, gut funktionierenden Zustand zurückversetzen. Häufig führt das resiliente Design des betreffenden Workloads eine Selbstreparatur durch. Einige Fehlerdesigns oder fehlgeschlagenen Experimente können Ihren Workload jedoch in einem nicht erwarteten Fehlerzustand zurücklassen. Nach dem Ende des Experiments müssen Sie dies erkennen und den Workload und die Systeme wiederherstellen können. Mit AWS FIS können Sie eine Rollback-Konfiguration innerhalb der Aktionsparameter einrichten (auch als „Post-Aktion“ bezeichnet). Eine Post-Aktion führt das Ziel in den Zustand zurück, in dem es sich vor Ausführung der Aktion befunden hat. Ob automatisiert (bei Verwendung von AWS FIS) oder manuell – diese Post-Aktionen sollten Teil eines Playbooks sein, das die Erkennung und Behandlung von Fehlern und Ausfällen beschreibt.

- Prüfen Sie die Hypothese.

  Grundlagen des Chaos-Engineering stellt die folgende Anleitung für die Verifizierung des Steady-State Ihres Workloads bereit:

  Konzentrieren Sie sich auf die messbare Ausgabe des Systems und nicht auf die internen Attribute des Systems. Messungen dieser Ausgabe über einen kurzen Zeitraum stellen einen Proxy für den Steady-State des Systems dar. Der Gesamtdurchsatz, die Fehlerraten und die Latenz-Perzentile des Systems könnten Metriken sein, die das Steady-State-Verhalten beschreiben. Durch die Konzentration auf die Verhaltensmuster des Systems während Experimenten überprüft das Chaos-Engineering, ob das System funktioniert, statt zu versuchen, die Art der Funktion zu validieren.

  In unseren beiden Beispielen oben verwenden wir die Steady-State-Metrik einer Erhöhung von weniger als 0,01 % bei serverseitigen Fehlern (5xx) und von weniger als einer Minute, in der Datenbankschreib- und Lesefehler auftreten.

  Die 5xx-Fehler stellen eine gute Metrik dar, da sie die Folge des Fehlermodus sind, dem ein Client des Workloads direkt unterliegen wird. Die Messung der Datenbankfehler ist als direkte Folge des Fehlers gut als Metrik geeignet, sollte jedoch durch eine Messung der Client-Auswirkungen ergänzt werden, beispielsweise in Form von fehlgeschlagenen Kundenanfragen oder Fehlern im Client. Zusätzlich sollten Sie für alle APIs oder URIs, auf die der Client Ihres Workloads direkt zugreift, eine synthetische Überwachung einrichten (auch als Benutzer-Canary bezeichnet).

- Verbessern Sie das Workload-Design hinsichtlich der Resilienz.

Wenn der Steady-State nicht bewahrt wurde, untersuchen Sie, wie das Workload-Design verbessert werden könnte, um den Fehler zu bewältigen. Wenden Sie dabei die Best Practices der AWS Well-Architected-Säule „Zuverlässigkeit“ an. Zusätzliche Anleitungen und Ressourcen finden Sie in der AWS Builder's Library.Diese Bibliothek enthält Artikel zur Verbesserung von Zustandsprüfungen oder zur Nutzung von Wiederholungen mit Backoff im Anwendungscodeund mehr.

Führen Sie das Experiment nach der Implementierung dieser Änderungen erneut durch (angezeigt durch die gepunktete Linie im Flywheel für das Chaos-Engineering), um ihre Effektivität zu ermitteln. Wenn der Verifizierungsschritt zeigt, dass die Hypothese zutrifft, befindet sich der Workload im Steady-State und der Zyklus wird fortgesetzt.

- Führen Sie regelmäßig Experimente durch.

  Ein Chaos-Experiment ist ein Zyklus. Daher sollten Experimente regelmäßig als Teil des Chaos-Engineering durchgeführt werden. Wenn die Hypothese eines Experiments auf einen Workload zutrifft, sollte das Experiment automatisiert werden, um innerhalb Ihrer CI/CD-Pipeline kontinuierlich als Regression ausgeführt zu werden. Informationen hierzu finden Sie in diesem Blog, der die Ausführung von AWS FIS-Experimenten mit AWS CodePipelinebeschreibt. Dieses Lab für wiederholte AWS FIS-Experimente in einer CI/CD-Pipeline ermöglicht Ihnen de Sammlung praktischer Erfahrungen.

  Fehlerinjektionsexperimente sind auch Bestandteil von Gamedays (siehe REL12-BP06 Regelmäßiges Abhalten von Gamedays). Bei Gamedays wird ein Fehler oder Ereignis simuliert, um Systeme, Prozesse und die Reaktionen von Teams zu testen. Dabei sollen die auszuführenden Aktionen vom Team wie im Fall eines außergewöhnlichen Ereignisses tatsächlich ausgeführt werden.

- Erfassen und speichern Sie die Ergebnisse der Experimente.

  Die Ergebnisse von Fehlerinjektionsexperimenten müssen erfasst und gespeichert werden. Erfassen Sie dabei alle notwendigen Daten (wie Zeit, Workload und Bedingungen), um die Ergebnisse und Trends von Experimenten später analysieren zu können. Beispiele für erfasste Ergebnisse können Screenshots von Dashboards, CSV-Versionen der Metrikdatenbank oder manuell eingegebene Aufzeichnungen von Ereignissen und Beobachtungen während des Experiments sein. Die Protokollierung von Experimenten mit AWS FIS kann Bestandteil dieser Datenerfassung sein.

Ressourcen

Zugehörige bewährte Methoden:

- REL08-BP03 Integrieren von Ausfallsicherheitstests in die Bereitstellung
- REL13-BP03 Testen der Implementierung der Notfallwiederherstellung zur Validierung:

Zugehörige Dokumente:

- Was ist AWS Fault Injection Service?
- Was ist AWS Resilience Hub?
- Grundlagen des Chaos-Engineering
- Chaos-Engineering: Planung Ihres ersten Experiments
- Resilience Engineering: Aus Fehlern lernen
- Chaos-Engineering-Geschichten
- Vermeiden von Fallback in verteilten Systemen
- Canary-Bereitstellung für Chaos-Experimente

Zugehörige Videos:

- AWS re:Invent 2020: Testing resiliency using chaos engineering (ARC316)
- AWS re:Invent 2019: Improving resiliency with chaos engineering (DOP309-R1)
- AWS re:Invent 2019: Performing chaos engineering in a serverless world (CMY301)

Zugehörige Beispiele:

- Well-Architected Lab: Level 300: Testen auf Resilienz von Amazon EC2, Amazon RDS und Amazon S3
- Chaos Engineering in AWS (Lab)
- Resilient and Well-Architected Apps with Chaos Engineering Lab
- Serverless-Chaos (Lab)
- Messen und Verbessern der Resilienz Ihrer Anwendung mit AWS Resilience Hub (Lab)

Zugehörige Tools:

- AWS Fault Injection Service
- AWS Marketplace: Gremlin Chaos Engineering Platform
- Chaos Toolkit
- Chaos Mesh
- Litmus

REL12-BP06 Regelmäßiges Abhalten von Gamedays

Nutzen Sie Gamedays, um Ihre Verfahren für Reaktionen auf Ereignisse und Fehler unter möglichst produktionsnahen Bedingungen (einschließlich Produktionsumgebungen) regelmäßig mit den Personen zu testen, die auch in tatsächlichen Fehlerszenarien beteiligt sind. Bei Gamedays werden Vorkehrungen getroffen, die sicherstellen, das sich Produktionsereignisse nicht auf Benutzer auswirken.

Bei Gamedays wird ein Fehler oder Ereignis simuliert, um Systeme, Prozesse und die Reaktion von Teams zu testen. Dabei sollen die auszuführenden Aktionen vom Team wie im Fall eines außergewöhnlichen Ereignisses tatsächlich ausgeführt werden. So können Sie nachvollziehen, wo nachgebessert werden kann. Zudem üben Sie dabei ein, wie Ihre Organisation mit Ereignissen umgeht. Gamedays sollten regelmäßig ausgeführt werden, damit die Reaktion für Ihr Team zu einem Reflex wird.

Nachdem Sie Ihre Maßnahmen für Ausfallsicherheit implementiert und in Umgebungen abseits der Produktion getestet haben, können Sie an einem Gameday feststellen, ob in der Produktion alles wie geplant funktioniert. An einem Gameday, insbesondere am ersten, werden alle Entwickler und Betriebsteams miteinbezogen und über Zeitpunkt sowie Ablauf des Tests informiert. Die Runbooks müssen vorhanden sein. Simulierte Ereignisse, auch potenzielle Ausfallereignisse, werden wie vorgeschrieben in den Produktionssystemen ausgeführt und deren Auswirkungen werden bewertet. Wenn alle Systeme wie vorgesehen funktionieren, erfolgen Erkennung und Selbstreparatur mit minimalen oder gar keinen Auswirkungen. Wenn jedoch negative Auswirkungen festgestellt werden, wird ein Rollback des Tests durchgeführt und die Workload-Probleme werden bei Bedarf manuell behoben (gemäß Runbook). Da Gamedays oft in der Produktion stattfinden, sollten alle Vorkehrungen getroffen werden, um Kunden vor Beeinträchtigungen der Verfügbarkeit zu schützen.

Gängige Antimuster:

- Die eigenen Verfahren werden dokumentiert, jedoch nie trainiert.
- Entscheidungsträger werden bei den Tests außen vorgelassen.

Vorteile der Einführung dieser Best Practice: Die regelmäßige Durchführung von Gamedays sorgt dafür, dass bei einem tatsächlichen Vorfall alle Mitarbeiter die Richtlinien und Verfahren befolgen. Außerdem wird überprüft, ob diese Richtlinien und Verfahren geeignet sind.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Planen Sie Gamedays, um Ihre Runbooks und Playbooks regelmäßig zu trainieren. An Gamedays sollten alle Mitarbeiter beteiligt werden, die von Produktionsunterbrechungen betroffen sein können: Geschäftsinhaber, Entwickler, Produktionsmitarbeiter und die Teams, die auf Vorfälle reagieren.
  - Führen Sie Ihre Last- oder Leistungstests durch und schleusen Sie anschließend Fehler ein.
  - Prüfen Sie die Runbooks auf Anomalien und suchen Sie nach Möglichkeiten zur Ausführung der Playbooks.
    - Optimieren Sie bei Abweichungen die Runbooks oder ändern Sie das Verhalten. Ermitteln Sie bei Ausführung eines Playbooks das Runbook, das hätte verwendet werden sollen, oder erstellen Sie ein neues.

Ressourcen

Zugehörige Dokumente:

- Was ist AWS GameDay?

Zugehörige Videos:

- AWS re:Invent 2019: Verbesserung der Ausfallsicherheit mit Chaos-Engineering (DOP309-R1)

  Zugehörige Beispiele:

- AWS Well-Architected Labs: Testen der Ausfallsicherheit

ZUV 13 Was ist bei der Planung der Notfallwiederherstellung zu beachten?

Backups und redundante Workload-Komponenten sind der Ausgangspunkt Ihrer Strategie für die Notfallwiederherstellung. RTO und RPO sind Ihre Ziele für die Wiederherstellung Ihrer Workload. Legen Sie diese Ziele entsprechend den geschäftlichen Anforderungen fest. Implementieren Sie

eine Strategie, um diese Ziele zu erreichen. Berücksichtigen Sie dabei Standorte und Funktionen von Workload-Ressourcen und -Daten. Die Wahrscheinlichkeit von Disruptionen und die Kosten von Wiederherstellungen sind ebenfalls wichtige Faktoren bei der Ermittlung des Unternehmenswerts, den Notfallwiederherstellungen von Workloads bieten.

Bewährte Methoden

- REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten:
- REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen
- REL13-BP03 Testen der Implementierung der Notfallwiederherstellung zur Validierung:
- REL13-BP04 Verwalten der Konfigurationsabweichungen am Standort oder in der Region der Notfallwiederherstellung:
- REL13-BP05: Automatisieren der Wiederherstellung

REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten:

Für die Workload gelten ein Recovery Time Objective (RTO, Wiederherstellungsdauer) und ein Recovery Point Objective (RPO, Wiederherstellungszeitpunkt).

Die Wiederherstellungsdauer ist die maximal akzeptable Verzögerung zwischen der Unterbrechung und der Wiederherstellung des Service. Damit wird festgelegt, was als akzeptables Zeitfenster gilt, wenn der Service nicht verfügbar ist.

Der Wiederherstellungszeitpunkt ist die maximal zulässige Zeitspanne seit dem letzten Wiederherstellungspunkt. Damit wird festgelegt, was als akzeptabler Datenverlust zwischen dem letzten Wiederherstellungspunkt und der Service-Unterbrechung gilt.

RTO- und RPO-Werte sind wichtige Überlegungen bei der Auswahl einer geeigneten Notfallwiederherstellungsstrategie (Disaster Recovery, DR) für Ihre Workload. Diese Ziele werden vom Unternehmen festgelegt und dann von den technischen Teams zur Auswahl und Umsetzung einer DR-Strategie verwendet.

Gewünschtes Ergebnis:

Jeder Workload sind ein RTO und ein RPO zugewiesen, die auf der Grundlage der geschäftlichen Auswirkungen definiert werden. Die Workload wird einer vordefinierten Stufe zugewiesen, die die Serviceverfügbarkeit und den akzeptablen Datenverlust mit einem entsprechenden RTO und

RPO definiert. Wenn eine solche Einstufung nicht möglich ist, kann die Zuweisung individuell pro Workload erfolgen, mit der Absicht, zu einem späteren Zeitpunkt Stufen zu erstellen. RTO und RPO werden als eine der Hauptüberlegungen für die Auswahl einer Notfallwiederherstellungsstrategie für die Workload verwendet. Weitere Überlegungen bei der Auswahl einer DR-Strategie sind Kostenbeschränkungen, Abhängigkeiten von der Workload und betriebliche Anforderungen.

Bei der RTO sind die Auswirkungen anhand der Dauer eines Ausfalls zu verstehen. Ist sie linear oder gibt es nichtlineare Auswirkungen? (Beispiel: Nach vier Stunden wird eine Fertigungsstraße bis zum Beginn der nächsten Schicht stillgelegt.)

Eine Matrix der Notfallwiederherstellung wie die folgende kann Ihnen helfen zu verstehen, wie die Kritikalität der Workload mit den Wiederherstellungszielen zusammenhängt. (Beachten Sie, dass die tatsächlichen Werte für die X- und Y-Achsen an die Bedürfnisse Ihres Unternehmens angepasst werden sollten.)

Matrix der Notfallwiederherstellung						
		Wiederherstellungszeitpunkt				
		< 1 Minute	< 1 Stunde	< 6 Stunden	< 1 Tag	+ 1 Tag
Wiederher-stellungsdauer	< 10 Minuten	Kritisch	Kritisch	Hoch	Mittel	Mittel
	< 2 Stunden	Kritisch	Hoch	Mittel	Mittel	Niedrig
	< 8 Stunden	Hoch	Mittel	Mittel	Niedrig	Niedrig
	< 24 Stunden	Mittel	Mittel	Niedrig	Niedrig	Niedrig
	24 + Stunden	Mittel	Niedrig	Niedrig	Niedrig	Niedrig

Abbildung 16: Matrix der Notfallwiederherstellung

Gängige Antimuster:

- Keine definierten Wiederherstellungsziele.
- Auswählen beliebiger Wiederherstellungsziele.
- Auswählen von Wiederherstellungszielen, die zu lasch sind und die Geschäftsziele nicht erfüllen.
- Kein Verständnis des Auswirkung von Ausfallzeiten und Datenverlust.
- Auswahl unrealistischer Wiederherstellungsziele, wie z. B. Null-Zeit bis zur Wiederherstellung und Null-Datenverlust, die für Ihre Workload-Konfiguration möglicherweise nicht erreicht werden können.

- Auswählen von Wiederherstellungszielen, die strikter sind als die tatsächlichen Geschäftsziele. Dies erzwingt Implementierungen für die Notfallwiederherstellung, die kostspieliger und komplizierter sind als die Anforderungen der Workload.
- Auswahl von Wiederherstellungszielen, die mit denen einer abhängigen Workloads unvereinbar sind.
- Ihre Wiederherstellungsziele berücksichtigen nicht die Einhaltung gesetzlicher Vorschriften.
- RTO und RPO sind für eine Workload definiert, aber nie getestet.

Vorteile der Einführung dieser bewährten Methode: Die Wiederherstellungsziele für Dauer und Datenverlust sind als Orientierungshilfe für die Implementierung der Notfallwiederherstellung erforderlich.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Bei der gegebenen Workload müssen Sie die Auswirkungen von Ausfallzeiten und Datenverlusten auf Ihr Unternehmen verstehen. Die Auswirkungen werden in der Regel mit zunehmender Ausfallzeit oder Datenverlust größer, aber die Form dieses Anstiegs kann je nach Art der Workload unterschiedlich sein. So können Sie z. B. Ausfallzeiten bis zu einer Stunde ohne größere Beeinträchtigung tolerieren, danach steigen die Auswirkungen jedoch schnell an. Die Auswirkungen auf das Unternehmen zeigen sich in vielen Formen, darunter monetäre Kosten (z. B. entgangene Einnahmen), Kundenvertrauen (und Auswirkungen auf den Ruf), betriebliche Probleme (z. B. fehlende Gehaltsabrechnungen oder verringerte Produktivität) und gesetzliche Risiken. Führen Sie die folgenden Schritte aus, um diese Auswirkungen zu verstehen und RTO und RPO für Ihre Workload festzulegen.

Implementierungsschritte

1. Bestimmen Sie die Interessengruppen Ihres Unternehmens für diese Workload und arbeiten Sie mit ihnen zusammen, um diese Schritte umzusetzen. Die Wiederherstellungsziele für eine Workload sind eine geschäftliche Entscheidung. Die technischen Teams arbeiten dann mit den Business-Stakeholdern zusammen, um anhand dieser Ziele eine DR-Strategie auszuwählen.

> **Note**
>
> Für die Schritte 2 und 3 können Sie Folgendes verwenden: the section called “Implementierungsarbeitsblatt”.

2. Sammeln Sie die notwendigen Informationen, um eine Entscheidung zu treffen, indem Sie die folgenden Fragen beantworten.
3. Gibt es in Ihrem Unternehmen Kategorien oder Stufen der Kritikalität für die Auswirkungen von Workloads?
   a. Falls zutreffend, ordnen Sie diese Workload einer Kategorie zu.
   b. Falls nicht zutreffend, richten Sie diese Kategorien ein. Legen Sie fünf oder weniger Kategorien fest und verfeinern Sie die Spanne der angestrebten Wiederherstellungszeit für jede Kategorie. Zu den Beispielkategorien gehören: kritisch, hoch, mittel, niedrig. Um zu verstehen, wie sich Workloads den Kategorien zuordnen lassen, sollten Sie prüfen, ob die Workload unternehmenskritisch, geschäftswichtig oder nicht geschäftsrelevant ist.
   c. Legen Sie RTO und RPO für die Workload je nach Kategorie fest. Wählen Sie immer eine Kategorie, die strikter ist (niedrigere RTO- und RPO-Werte) als die bei der Eingabe dieses Schritts berechneten Rohwerte. Wenn dies zu einer unangemessen großen Veränderung des Wertes führt, sollten Sie eine neue Kategorie anlegen.
4. Weisen Sie auf der Grundlage dieser Antworten der Workload RTO- und RPO-Werte zu. Dies kann direkt geschehen oder durch Zuweisung der Workload zu einer vordefinierten Serviceebene.
5. Dokumentieren Sie den Notfallwiederherstellungsplan (Disaster Recovery Plan, DRP) für diese Workload, der Teil der Unternehmensstrategie ist. Betriebskontinuitätsplan (BCP)an einem Ort, der für das Workload-Team und die Stakeholder zugänglich ist
   a. Halten Sie die RTO- und RPO-Werte sowie die zur Ermittlung dieser Werte verwendeten Informationen fest. Geben Sie eine Strategie zur Bewertung der Auswirkungen der Workload auf das Unternehmen an.
   b. Erfassen Sie neben RTO und RPO auch andere Metriken, die Sie für Notfallwiederherstellungsziele verfolgen oder zu verfolgen planen
   c. Sie fügen diesem Plan Details zu Ihrer DR-Strategie und Ihrem Runbook hinzu, wenn Sie diese erstellen.
6. Indem Sie die Kritikalität der Workload in einer Matrix wie der in Abbildung 15 nachschlagen, können Sie damit beginnen, vordefinierte Serviceebenen für Ihr Unternehmen festzulegen.

7. Nachdem Sie eine DR-Strategie (oder einen Machbarkeitsnachweis für eine DR-Strategie) gemäß implementiert haben,the section called “REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen”testen Sie diese Strategie, um die tatsächliche RTC (Recovery Time Capability) und RPC (Recovery Point Capability) der Workload zu bestimmen. Wenn diese nicht den angestrebten Wiederherstellungszielen entsprechen, arbeiten Sie entweder mit Ihren Stakeholdern zusammen, um diese Ziele anzupassen, oder nehmen Sie Änderungen an der DR-Strategie vor, um die Zielvorgaben zu erreichen.

Primäre Fragen

1. Wie lange kann die Workload maximal ausfallen, bevor es zu schwerwiegenden Auswirkungen auf das Unternehmen kommt?
   a. Bestimmen Sie die monetären Kosten (direkte finanzielle Auswirkungen) für das Unternehmen pro Minute, wenn die Workload unterbrochen wird.
   b. Bedenken Sie, dass die Auswirkungen nicht immer linear sind. Die Auswirkungen können zunächst begrenzt sein und dann ab einem kritischen Zeitpunkt rasch zunehmen.
2. Wie groß ist die maximale Datenmenge, die verloren gehen kann, bevor es zu schwerwiegenden Auswirkungen auf das Unternehmen kommt?
   a. Berücksichtigen Sie diesen Wert für Ihren wichtigsten Datenspeicher. Identifizieren Sie die jeweilige Kritikalität für andere Datenspeicher.
   b. Können Workload-Daten bei Verlust wiederhergestellt werden? Wenn dies aus betrieblicher Sicht einfacher ist als Backup und Wiederherstellung, dann wählen Sie das RPO auf der Grundlage der Kritikalität der Ursprungsdaten, die zur Wiederherstellung der Workload-Daten verwendet werden.
3. Wie lauten die Wiederherstellungsziele und Verfügbarkeitserwartungen von Workloads, von denen dieser abhängt (Downstream), oder von Workloads, die von diesem abhängen (Upstream)?
   a. Wählen Sie Wiederherstellungsziele, die es dieser Workload ermöglichen, die Anforderungen der vorgelagerten Abhängigkeiten zu erfüllen
   b. Wählen Sie Wiederherstellungsziele, die angesichts der Wiederherstellungsmöglichkeiten der nachgelagerten Abhängigkeiten erreichbar sind. Unkritische nachgelagerte Abhängigkeiten (die Sie „umgehen“ können) können ausgeschlossen werden. Oder arbeiten Sie mit kritischen, nachgelagerten Abhängigkeiten zusammen, um deren Wiederherstellungsmöglichkeiten zu verbessern.

Weitere Fragen

Überlegen Sie sich, wie diese Fragen auf diese Workload zutreffen könnten:

4. Haben Sie unterschiedliche RTO und RPO je nach Art des Ausfalls (Region vs. Region)? AZ, etc.)?
5. Gibt es einen bestimmten Zeitpunkt (Saisonabhängigkeit, Verkaufsveranstaltungen, Produkteinführungen), zu dem sich Ihr RTO/RPO ändern kann? Wenn ja, was ist die unterschiedliche Messung und die zeitliche Begrenzung?
6. Wie viele Kunden sind von einer Unterbrechung der Workload betroffen?
7. Welche Auswirkungen hat es auf den Ruf, wenn die Workload unterbrochen wird?
8. Welche anderen betrieblichen Auswirkungen können auftreten, wenn die Workload unterbrochen wird? Zum Beispiel Auswirkungen auf die Produktivität der Mitarbeiter, wenn die E-Mail-Systeme nicht verfügbar sind oder wenn die Lohnbuchhaltungssysteme keine Transaktionen übermitteln können.
9. Wie stimmen RTO und RPO der Workload mit der DR-Strategie der Geschäftsbereiche und des Unternehmens überein?
10. Gibt es interne vertragliche Verpflichtungen für die Erbringung einer Dienstleistung? Gibt es Strafen für die Nichteinhaltung dieser Vorgaben?
11. Welche rechtlichen oder Compliance-Bedingungen gelten für die Daten?

Implementierungsarbeitsblatt

Sie können dieses Arbeitsblatt für die Implementierungsschritte 2 und 3 verwenden. Sie können dieses Arbeitsblatt an Ihre speziellen Bedürfnisse anpassen, indem Sie beispielsweise zusätzliche Fragen hinzufügen.

Schritt 2: primäre Fragen	Gilt für Workload?	Workload-RTO	Workload-RPO	RTO anpassen	RPO anpassen	Anleitungen
[1] Maximale Zeit, in der der Workload ausfallen kann						Gemessen Zeit seit Beginn des Ausfalls bis zur Wiederherstellung
[2] Maximale Datenmenge, die verloren gehen kann						Gemessen in Zeit seit dem letzten bekannten gut wiederherstellbaren Datensatz
[3a] Vorgelagerte Abhängigkeiten						Strengste nachgelagerte Wiederherstellungsziele eingeben
[3b] Nachgelagerte Abhängigkeiten						Am wenigsten strenge nachgelagerte Wiederherstellungsziele eingeben
[3a] Abgeglichene vorgelagerte Abhängigkeiten						Wenn der vorgelagerte Wert niedriger ist als aktuelle Werte und der nachgelagerte Wert größer ist,
[3b] Abgeglichene nachgelagerte Abhängigkeiten						arbeiten Sie mit Abhängigkeiten, um auszugleichen und hier ausgeglichene Werte einzugeben.
[3] Abhängigkeiten						Werte senken, um vorgelagerte Abhängigkeiten zu erfüllen oder die basierend auf nachgelagerten Abhängigkeitsfähigkeiten zu erhöhen
Schritt 2: zusätzliche Fragen						Geben Sie an, ob die Frage zutrifft. Falls nicht, überspringen Sie sie.
Basis-RTO/-RPO						Übertragen Sie die RTO- und RPO-Werte von oben nach hier unten.
[4] Art des Ausfalls	[ ]J/[ ]N					Geben Sie Wiederherstellungsziele für Ereignisarten mit strengsten Anforderungen ein.
[5] Spezifische zeitbasierte Ziele	[ ]J/[ ]N					Geben Sie Wiederherstellungsziele für Zeiten mit strengsten Anforderungen ein.
[6] Unterbrechungen bei Kunden	[ ]J/[ ]N					Grafische Darstellung der betroffenen Kunden in Abhängigkeit von der Ausfallzeit oder dem Datenverlust. Verwenden Sie dies, um das maximal zulässige RTO und RPO auf der Grundlage der Kundenauswirkungen einzugeben.
[7] Auswirkungen auf den Ruf	[ ]J/[ ]N					Mit dem Unternehmen arbeiten, um die maximale RTO und den maximalen RPO basierend auf der Auswirkung auf die Reputation zu bestimmen
[8] Betriebliche Auswirkungen	[ ]J/[ ]N					Geben Sie das maximale RTO und RPO auf der Grundlage der betrieblichen Auswirkungen ein.
[9] Organisatorische Ausrichtung	[ ]J/[ ]N					Geben Sie das maximale RTO und RPO für Workloads dieses Typs gemäß den LOB- und Organisationsanforderungen ein.
[10] Vertragliche Verpflichtungen	[ ]J/[ ]N					Geben Sie das maximale RTO und RPO auf der Grundlage der vertraglichen Verpflichtungen ein.
[11] Gesetzliche Vorschriften	[ ]J/[ ]N					Geben Sie das maximale RTO und RPO auf der Grundlage der geltenden gesetzlichen Bestimmungen ein.
Ziel basierend auf zusätzlichen Fragen						Nehmen Sie den Mindestwert (strengerer Wert) aus den Fragen 4–11 und geben Sie ihn hier ein.
Angepasstes Ziel						Wenn die Ziele in der obigen Zeile nicht erreicht werden können, arbeiten Sie mit den Beteiligten zusammen, um die Beschränkungen zu lockern, und geben Sie hier ein neues Minimum ein.
RTO/RPO angepasst						Geben Sie die Basis-RPO-/RTO-Werte oder das angepasste Ziel ein, je nachdem, welcher Wert niedriger ist.
Schritt 3						
Zuordnung zu vordefinierter Kategorie oder Stufe						Senken Sie beide Werte (machen Sie sie strenger), um sie an die nächstgelegene definierte Stufe anzupassen.

Arbeitsblatt

Grad des Aufwands für den Implementierungsplan: Niedrig

Ressourcen

Ähnliche bewährte Methoden:

- the section called “REL09-BP04 Verifizieren der Sicherungsintegrität und -verfahren durch regelmäßiges Wiederherstellen der Daten”
- the section called “REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen”
- the section called “REL13-BP03 Testen der Implementierung der Notfallwiederherstellung zur Validierung:”

Zugehörige Dokumente:

- AWS Architecture Blog: Notfallwiederherstellungsserie

- Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud (AWS-Whitepaper)
- Verwalten von Ausfallsicherheit mit AWS Resilience Hub
- APN-Partner: Partner, die Sie bei der Notfallwiederherstellung unterstützen können
- AWS Marketplace: Für die Notfallwiederherstellung geeignete Produkte

Relevante Videos

- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2) (Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen)
- Notfallwiederherstellung von Workloads auf AWS

REL13-BP02: Verwenden von definierten Wiederherstellungsstrategien, um die Wiederherstellungsziele zu erreichen

Definieren Sie eine Notfallwiederherstellungsstrategie (Disaster Recovery, DR), die den Wiederherstellungszielen Ihrer Workloads entspricht. Wählen Sie eine Strategie aus, z. B. Backup und Wiederherstellung, Standby (aktiv/passiv) oder Aktiv/Aktiv.

Eine DR-Strategie beruht auf der Fähigkeit, Ihre Workload an einem Wiederherstellungsstandort bereitzustellen, wenn Ihr primärer Standort nicht mehr in der Lage ist, den Workload auszuführen. Die häufigsten Wiederherstellungsziele sind RTO und RPO, wie besprochen in REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten:.

Eine DR-Strategie, die mehrere Availability Zones (AZs) innerhalb eines einzigen AWS-Region umfasst, kann Katastrophenereignisse wie Brände, Überschwemmungen und größere Stromausfälle abfedern. Wenn es erforderlich ist, einen Schutz gegen ein unwahrscheinliches Ereignis zu implementieren, das verhindert, dass Ihre Workload in einer bestimmten AWS-Region ausgeführt werden kann, können Sie eine DR-Strategie verwenden, die mehrere Regionen nutzt.

Wenn Sie eine DR-Strategie für mehrere Regionen entwickeln, sollten Sie eine der folgenden Strategien wählen. Sie werden nach zunehmenden Kosten und zunehmender Komplexität und abnehmender RTO und RPO aufgelistet. Wiederherstellungsregion bezieht sich auf einen AWS-Region anders als der primäre, der für Ihre Workload verwendet wird.

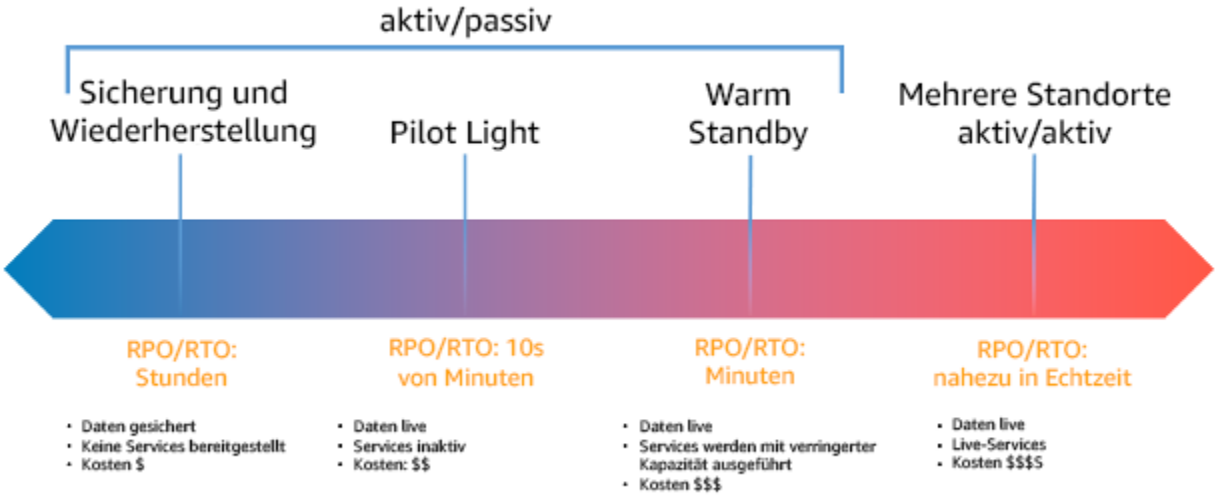


Abbildung 17: Notfallwiederherstellungsstrategien (DR)

- Sicherung und Wiederherstellung (RPO in Stunden, RTO in 24 Stunden oder weniger): Sichern Sie Ihre Daten und Anwendungen in der Wiederherstellungsregion. Die Verwendung automatisierter oder kontinuierlicher Backups ermöglicht eine zeitpunktgenaue Wiederherstellung, wodurch das RPO in einigen Fällen auf bis zu 5 Minuten gesenkt werden kann. Im Falle eines Notfalls stellen Sie Ihre Infrastruktur bereit (wobei Sie Infrastruktur als Code verwenden, um die RTO zu verkürzen), stellen Ihren Code bereit und stellen die gesicherten Daten wieder her, um eine Wiederherstellung nach einem Notfall in der Wiederherstellungsregion zu erfahren.
- Pilot Light (RPO in Minuten, RTO in zehn Minuten): Bereitstellung einer Kopie Ihrer Kern-Workload-Infrastruktur in der Wiederherstellungsregion. Replizieren Sie Ihre Daten in die Wiederherstellungsregion und erstellen Sie dort Sicherungskopien der Daten. Ressourcen, die zur Unterstützung der Datenreplikation und -sicherung erforderlich sind, wie Datenbanken und Objektspeicher, sind immer eingeschaltet. Andere Elemente wie Anwendungsserver oder Serverless Compute werden nicht bereitgestellt, sondern können bei Bedarf mit der erforderlichen Konfiguration und dem Anwendungscode erstellt werden.
- Warm Standby (RPO in Sekunden, RTO in Minuten): Behalten Sie eine verkleinerte, aber voll funktionsfähige Version Ihres Workloads in der Wiederherstellungsregion bei. Geschäftskritische Systeme sind vollständig dupliziert und ständig aktiv, aber mit herunterskalierter Infrastruktur. Die Daten werden repliziert und sind in der Wiederherstellungsregion live. Wenn eine Wiederherstellung erforderlich ist, wird das System zur Bewältigung der Produktionslast schnell hochskaliert. Je höher die Skalierung des Warm Standby, desto geringer ist die Abhängigkeit von RTO und Kontrollebene. Wenn voll skaliert, wird dies als Hot Standby bezeichnet.

- Multi-Region (Multi-Site) aktiv-aktiv (RPO nahe Null, RTO potenziell Null): Ihre Workload wird auf mehreren AWS-Regionen bereitgestellt und bedient aktiv Datenverkehr von diesen. Bei dieser Strategie müssen Sie die Daten zwischen den Regionen synchronisieren. Mögliche Konflikte, die durch Schreibvorgänge auf denselben Datensatz in zwei verschiedenen regionalen Repliken verursacht werden, müssen vermieden oder behandelt werden, was sehr komplex sein kann. Die Datenreplikation ist nützlich für die Datensynchronisation und schützt Sie vor einigen Arten von Notfällen, aber sie schützt Sie nicht vor Datenbeschädigung oder -zerstörung, es sei denn, Ihre Lösung umfasst auch Optionen für eine zeitpunktgenaue Wiederherstellung.

> **Note**
>
> Der Unterschied zwischen Pilot Light und Warm Standby ist oft nicht sofort klar. Beide beinhalten eine Umgebung in Ihrer Wiederherstellungsregion mit Kopien der Assets Ihrer Primärregion. Der Unterschied besteht darin, dass Pilot Light keine Anfragen bearbeiten kann, ohne dass zuvor zusätzliche Maßnahmen ergriffen werden, während Warm Standby den Datenverkehr (mit reduzierter Kapazität) sofort bearbeiten kann. Bei Pilot Light müssen Sie die Server einschalten, möglicherweise zusätzliche (nicht zum Kerngeschäft gehörende) Infrastruktur bereitstellen und die Leistung hochskalieren, während Sie bei Warm Standby nur die Leistung hochskalieren müssen (alles ist bereits bereitgestellt und läuft). Wählen Sie je nach RTO- und RPO-Anforderungen zwischen diesen Varianten.

Gewünschtes Ergebnis:

Für jede Workload gibt es eine definierte und implementierte DR-Strategie, die es dieser Workload ermöglicht, die DR-Ziele zu erreichen. DR-Strategien zwischen Workloads nutzen wiederverwendbare Muster (wie die zuvor beschriebenen Strategien),

Gängige Antimuster:

- Implementierung von inkonsistenten Wiederherstellungsprozeduren für Workloads mit ähnlichen DR-Zielen.
- Die DR-Strategie muss im Notfall Ad-hoc umgesetzt werden.
- Kein Plan verfügbar für DR.
- Abhängigkeit von Vorgängen auf der Steuerungsebene während der Wiederherstellung.

Vorteile der Einführung dieser bewährten Methode:

- Durch die Nutzung definierter Wiederherstellungsstrategien können Sie verbreitet verwendete Tools und Testverfahren verwenden.
- Die Verwendung definierter Wiederherstellungsstrategien ermöglicht einen effizienteren Wissensaustausch zwischen den Teams und eine einfachere Implementierung von DR für die eigenen Workloads.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

- Ohne eine geplante, implementierte und getestete DR-Strategie ist es unwahrscheinlich, dass Sie Ihre Wiederherstellungsziele im Falle eines Notfalls erreichen.

Implementierungsleitfaden

Zu jedem dieser Schritte finden Sie unten weitere Informationen.

1. Bestimmen Sie eine DR-Strategie, die die Wiederherstellungsanforderungen für diese Workload erfüllt.
2. Überprüfen Sie die Muster, wie die ausgewählte DR-Strategie umgesetzt werden kann.
3. Beurteilen Sie die Ressourcen Ihrer Workloads und deren Konfiguration in der Wiederherstellungsregion vor dem Failover (während des normalen Betriebs).
4. Legen Sie fest, wie Sie Ihre Wiederherstellungsregion bei Bedarf (während eines Notfallereignisses) für einen Failover bereit machen wollen, und setzen Sie diese um.
5. Legen Sie fest und implementieren Sie, wie Sie den Datenverkehr bei Bedarf (im Notfall) zum Failover umleiten werden.
6. Entwerfen Sie einen Plan, wie Ihre Workload zurückgehen wird.

Implementierungsschritte

1. Bestimmen Sie eine DR-Strategie, die die Wiederherstellungsanforderungen für diese Workload erfüllt.

Die Wahl einer DR-Strategie ist eine Abwägung zwischen der Reduzierung von Ausfallzeiten und Datenverlusten (RTO und RPO) und den Kosten und der Komplexität der Implementierung der Strategie. Sie sollten vermeiden, eine Strategie zu verfolgen, die strikter ist als nötig, da dies unnötige Kosten verursacht.

Im folgenden Diagramm hat das Unternehmen beispielsweise seine maximal zulässige RTO sowie die Grenze der Ausgaben für seine Strategie zur Wiederherstellung von Diensten festgelegt. In Anbetracht der Ziele des Unternehmens erfüllen die DR-Strategien Pilot Light oder Warm Standby sowohl die RTO- als auch die Kostenkriterien.

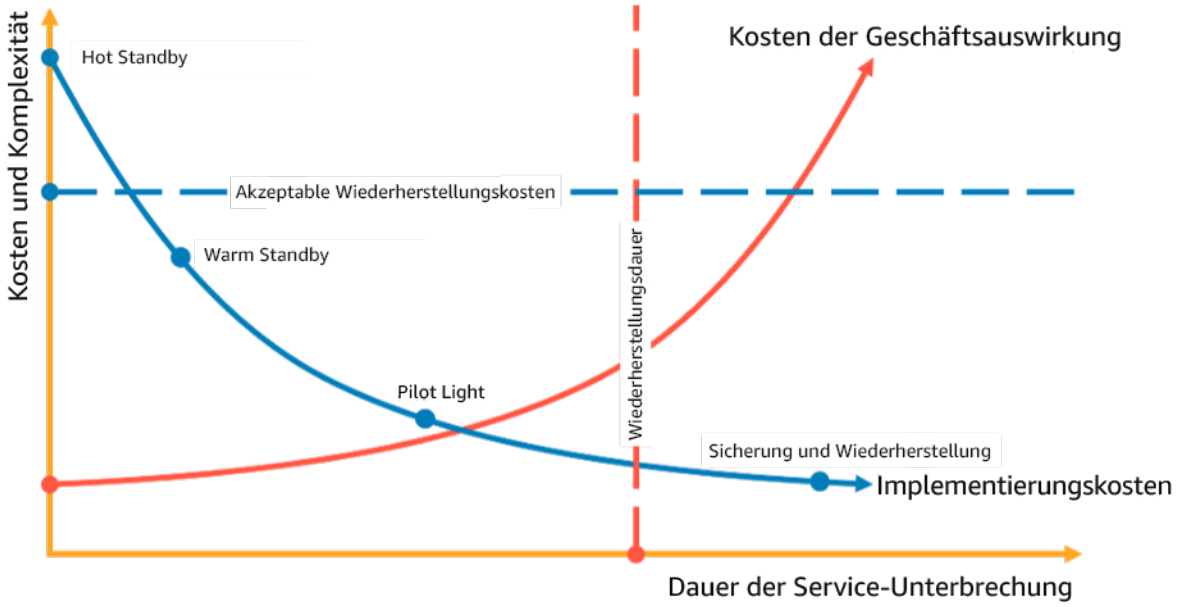


Abbildung 18: Auswahl einer DR-Strategie auf der Grundlage von RTO und Kosten

Weitere Information finden Sie unter Betriebskontinuitätsplan (BCP).

2. Überprüfen Sie die Muster, wie die ausgewählte DR-Strategie umgesetzt werden kann.

In diesem Schritt geht es darum, zu verstehen, wie Sie die gewählte Strategie umsetzen wollen. Die Strategien werden durch die Verwendung von AWS-Regionen als primäre und Wiederherstellungsstandort erläutert. Sie können jedoch auch Verfügbarkeitszonen innerhalb einer einzigen Region als DR-Strategie verwenden, die Elemente mehrerer dieser Strategien nutzt.

In den darauf folgenden Schritten werden Sie die Strategie auf Ihre spezifische Workload anwenden.

Sicherung und Wiederherstellung

Sicherung und Wiederherstellung ist die am einfachsten zu implementierende Strategie, erfordert aber mehr Zeit und Aufwand für die Wiederherstellung der Workload, was zu höheren RTO- und RPO-Werten führt. Es ist eine gute Vorgehensweise, immer Sicherungskopien Ihrer Daten zu erstellen und diese auf einen anderen Standort (z. B. einen anderen AWS-Region) zu kopieren.

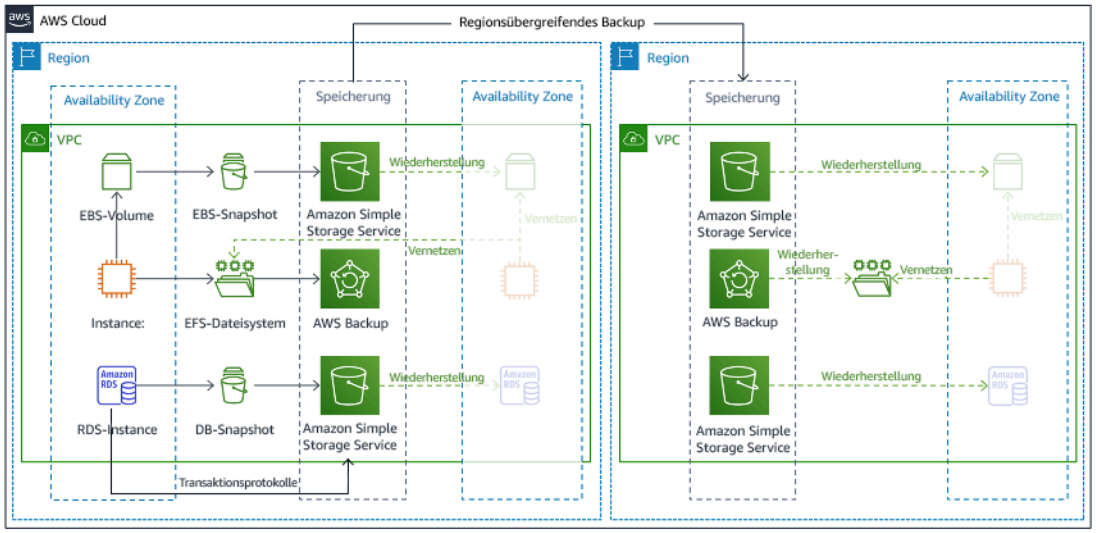


Abbildung 19: Sicherungs- und Wiederherstellungsarchitektur

Weitere Details zu dieser Strategie finden Sie unter Notfallwiederherstellungsarchitektur (DR) auf AWS, Teil II: Sicherung und Wiederherstellung mit Rapid Recovery.

Pilot Light

Mit dem Pilot Light-Verfahren replizieren Sie Ihre Daten von Ihrer primären Region auf Ihre Wiederherstellungsregion. Die Kernressourcen, die für die Workload-Infrastruktur verwendet werden, werden in der Wiederherstellungsregion bereitgestellt, jedoch werden noch zusätzliche Ressourcen und Abhängigkeiten benötigt, um diesen Stack funktionsfähig zu machen. In Abbildung 20 werden zum Beispiel keine Compute Instances bereitgestellt.

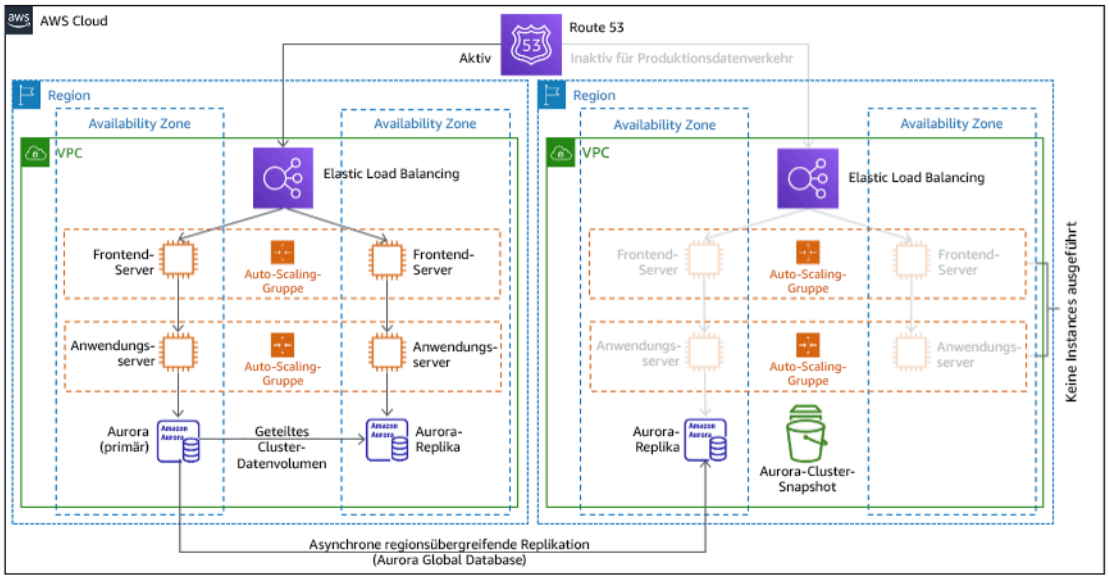


Abbildung 20: Pilot Light-Architektur

Weitere Details zu dieser Strategie finden Sie unter [Notfallwiederherstellungsarchitektur (DR) auf AWS, Teil III: Pilot Light und Warm Standby].

Warm Standby

Das Warm Standby Verfahren stellt sicher, dass eine verkleinerte, aber voll funktionsfähige Kopie Ihrer Produktionsumgebung in einer anderen Region vorhanden ist. Dieser Ansatz erweitert das Konzept des Pilot Light und verkürzt die Zeit bis zur Wiederherstellung, da die Workload in einer anderen Region ständig präsent ist. Wenn die Wiederherstellungsregion mit voller Kapazität eingesetzt wird, spricht man von einem Hot Standby.

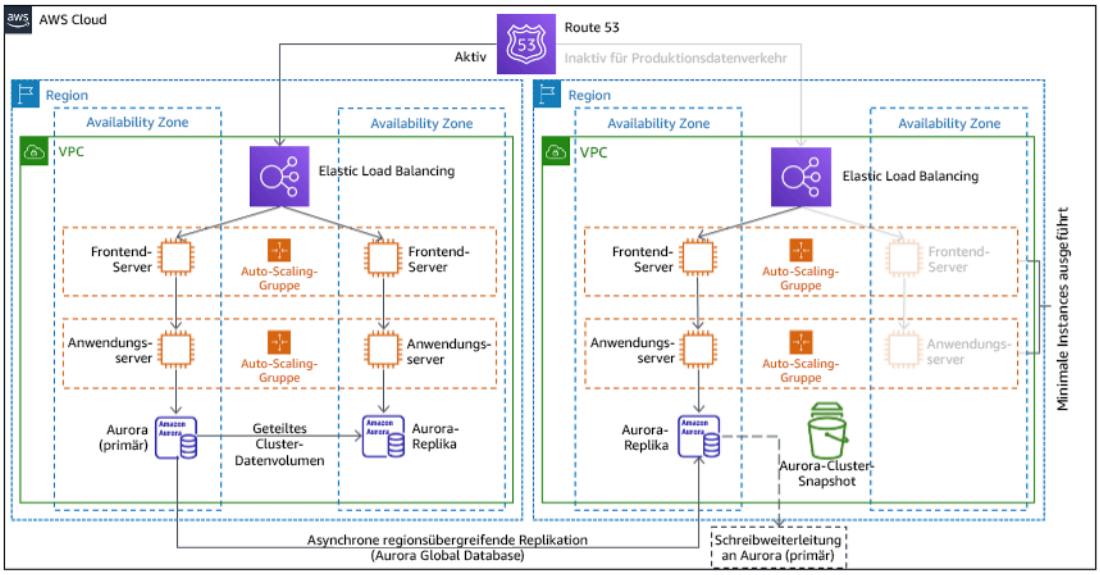


Abbildung 21: Warm Standby-Architektur

Der Einsatz von Warm Standby oder Pilot Light erfordert ein Hochskalieren der Ressourcen in der Wiederherstellungsregion. Um sicherzustellen, dass die Kapazität bei Bedarf zur Verfügung steht, sollten Sie die Verwendung für Kapazitätsreservierungen für EC2-Instances erwägen. Wenn Sie AWS Lambda verwenden, können Sie mit Hilfe der Gleichzeitigkeit Ausführungsumgebungen sicherstellen, die darauf vorbereitet sind, sofort auf die Aufrufe Ihrer Funktion zu reagieren.

Weitere Details zu dieser Strategie finden Sie unter Notfallwiederherstellungsarchitektur (DR) auf AWS, Teil III: Pilot Light und Warm Standby.

Multi-Site Aktiv/Aktiv

Sie können Ihre Workload gleichzeitig in mehreren Regionen als Teil einer Multi-Site Aktiv/Aktiv-Strategie ausführen. Multi-Site Aktiv/Aktiv bedient den Datenverkehr aus allen Regionen, in denen es eingesetzt wird. Kunden können diese Strategie aus anderen Gründen als DR wählen. Sie kann zur Erhöhung der Verfügbarkeit oder bei der Bereitstellung einer Workload für eine globale Zielgruppe verwendet werden (um den Endpunkt näher an die Benutzer zu bringen und/oder um Stacks bereitzustellen, die für die Zielgruppe in dieser Region lokalisiert sind). Bei einer DR-Strategie wird, wenn die Workload in einer der AWS-Regionen, in denen sie bereitgestellt wird, nicht unterstützt werden kann, diese Region evakuiert und die verbleibende(n) Region(en) wird (werden) zur Aufrechterhaltung der Verfügbarkeit verwendet. Multi-Site Aktiv/Aktiv ist die betrieblich komplexeste der DR-Strategien und sollte nur dann gewählt werden, wenn die Geschäftsanforderungen dies erfordern.

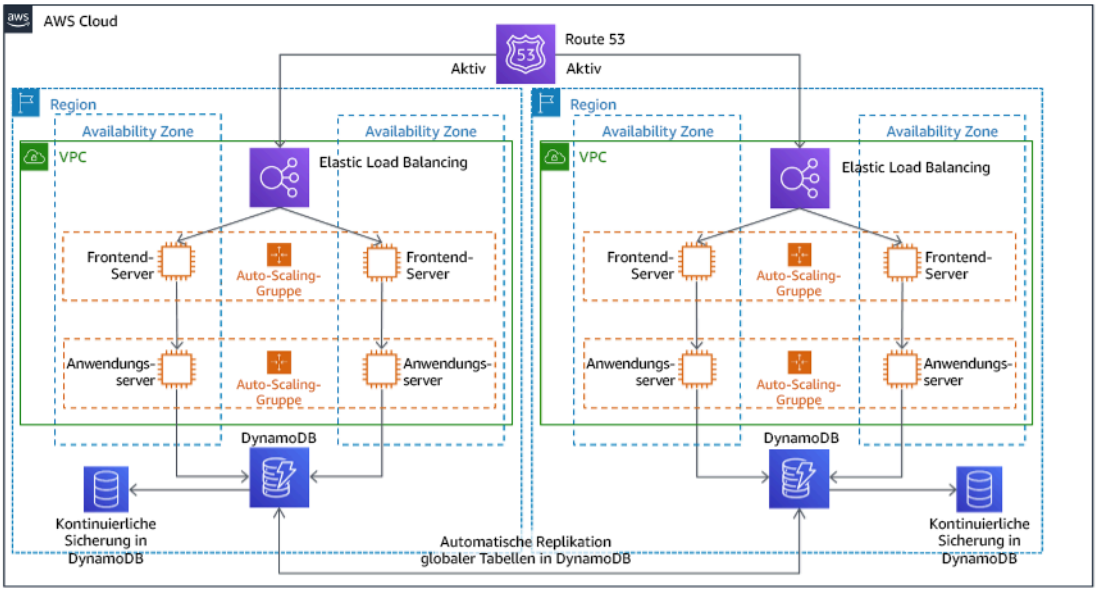


Abbildung 22: Multi-Site Aktiv/Aktiv Architektur

Weitere Details zu dieser Strategie finden Sie unter Architektur für die Notfallwiederherstellung (Disaster Recovery, DR) auf AWS, Teil IV: Multi-Site Aktiv-Aktiv.

Zusätzliche Praktiken zum Schutz von Daten

Bei allen Strategien müssen Sie sich auch gegen einen Datennotfall wappnen. Kontinuierliche Datenreplikation schützt Sie vor einigen Arten von Notfällen, aber sie schützt Sie möglicherweise nicht vor Datenbeschädigung oder -zerstörung, es sei denn, Ihre Strategie umfasst auch die Versionsverwaltung gespeicherter Daten oder Optionen für eine zeitpunktgenaue Wiederherstellung. Sie müssen auch die replizierten Daten in der Wiederherstellungssite sichern, um zusätzlich zu den Replikaten zeitpunktgenaue Sicherungen zu erstellen.

Verwendung von mehreren Availability Zones (AZs) innerhalb einer einzigen AWS-Region

Wenn Sie mehrere AZs in einer einzigen Region verwenden, nutzt Ihre DR-Implementierung mehrere Elemente der oben genannten Strategien. Zunächst müssen Sie eine Hochverfügbarkeitsarchitektur (High Availability / HA) mit mehreren AZs erstellen, wie in Abbildung 23 dargestellt. Diese Architektur nutzt einen Multi-Site Aktiv/Akitv-Ansatz als Amazon EC2-Instances und Elastic Load Balancer stellen Ressourcen in mehreren AZs bereit, die Anfragen bearbeiten. Die Architektur demonstriert auch Hot Standby, wo, wenn die primäre Amazon RDS Instance ausfällt (oder die AZ selbst ausfällt), die Standby-Instance dann zur primären Instance hochgestuft wird.

Abbildung 23: Multi-AZ-Architektur

Zusätzlich zu dieser HA-Architektur müssen Sie Backups aller Daten hinzufügen, die für die Ausführung Ihrer Workloads erforderlich sind. Dies ist besonders wichtig für Daten, die sich auf eine einzige Zone beschränken, wie z. B. Amazon EBS-Volumes oder Amazon Redshift-Cluster. Wenn ein AZ ausfällt, müssen Sie diese Daten auf einem anderen AZ wiederherstellen. Wenn möglich, sollten Sie auch Datensicherungen auf einen anderen AWS-Region kopieren, um eine zusätzliche Sicherheit zu gewährleisten.

Ein weniger gebräuchlicher alternativer Ansatz für eine einzelne Region, Multi-AZ DR, wird in diesem Blogbeitrag vorgestellt, Entwickeln hoch resilienter Anwendungen mit Amazon Route 53 Application Recovery Controller, Teil 1: Stack für eine einzelne Region. Hier besteht die Strategie darin, so viel Isolation wie möglich zwischen den AZs aufrechtzuerhalten, ähnlich wie bei den Regionen. Bei dieser alternativen Strategie können Sie sich für einen aktiv/aktiven oder aktiv/passiven Ansatz entscheiden.

Hinweis: Für einige Workloads gibt es gesetzliche Anforderungen an die Datenaufbewahrung. Wenn dies auf Ihre Workload in einer Region zutrifft, in der es derzeit nur eine AWS-Region gibt, dann ist die Multi-Region für Ihre geschäftlichen Anforderungen nicht geeignet. Multi-AZ-Strategien bieten einen guten Schutz gegen die meisten Notfälle.

3. Beurteilen Sie die Ressourcen Ihrer Workloads und deren Konfiguration in der Wiederherstellungsregion vor dem Failover (während des normalen Betriebs).

Für Infrastruktur und AWS-Ressourcen verwenden Sie Infrastruktur als Code oder Tools AWS CloudFormation von Drittanbietern wie Hashicorp Terraform. Für die Bereitstellung auf mehrere Konten und Regionen mit einem einzigen Vorgang können Sie AWS CloudFormation-StackSets verwenden. Bei Multi-Site-Aktiv/Aktiv- und Hot Standby-Strategien verfügt die in Ihrer Wiederherstellungsregion bereitgestellte Infrastruktur über dieselben Ressourcen wie Ihre Primärregion. Bei den Strategien Pilot Light und Warm Standby sind zusätzliche Maßnahmen erforderlich, um die Infrastruktur produktionsreif zu machen. Mit der Hilfe von CloudFormation Parametern und bedingter Logikkönnen Sie mit einer einzigen Vorlage steuern, ob ein bereitgestellter Stack aktiv oder in Bereitschaft ist. Ein Beispiel für eine solche CloudFormation-Vorlage ist in diesem Blogbeitrag enthalten..

Alle DR-Strategien setzen voraus, dass die Datenquellen innerhalb der AWS-Region gesichert werden und diese Sicherungen dann in die Wiederherstellungsregion kopiert werden. AWS Backup bietet eine zentrale Ansicht, in der Sie Backups für diese Ressourcen konfigurieren, planen und überwachen können. Bei Pilot Light, Warm Standby und Multi-Site Aktiv/Aktiv sollten Sie auch Daten aus der primären Region auf Datenressourcen in der Wiederherstellungsregion replizieren, z. B. Amazon Relational Database Service (Amazon RDS) DB-Instances oder Amazon DynamoDB -Tabellen. Diese Datenressourcen sind daher aktiv und bereit, Anfragen in der Wiederherstellungsregion zu bedienen.

Weitere Informationen darüber, wie AWS-Dienste regionenübergreifend funktionieren, finden Sie in dieser Blogserie über das Erstellen einer regionenübergreifenden Anwendung mit AWS-Services.

4. Legen Sie fest, wie Sie Ihre Wiederherstellungsregion bei Bedarf (während eines Notfallereignisses) für einen Failover bereit machen wollen, und setzen Sie diese um.

Bei Multi-Site Aktiv/Aktiv bedeutet Failover, dass eine Region evakuiert wird und die verbleibenden aktiven Regionen genutzt werden. Im Allgemeinen sind diese Regionen bereit, Datenverkehr aufzunehmen. Bei den Strategien „Pilot Light“ und „Warm Standby“ müssen Ihre Wiederherstellungsmaßnahmen die fehlenden Ressourcen bereitstellen, z. B. die EC2-Instances in Abbildung 20, sowie alle anderen fehlenden Ressourcen.

Bei allen oben genannten Strategien müssen Sie möglicherweise schreibgeschützte Instances von Datenbanken zur primären Lese-/Schreibinstance machen.

Bei der Sicherung und Wiederherstellung werden durch die Wiederherstellung von Daten aus der Sicherung Ressourcen für diese Daten wie EBS-Volumes, RDS-DB-Instances und DynamoDB-Tabellen erstellt. Außerdem müssen Sie die Infrastruktur wiederherstellen und Code bereitstellen. Sie

können AWS Backup Daten in der Wiederherstellungsregion wiederherstellen. Transparenz REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen für weitere Informationen. Der Wiederaufbau der Infrastruktur umfasst die Erstellung von Ressourcen wie EC2-Instances zusätzlich zu den Amazon Virtual Private Cloud (Amazon VPC),Subnetzen und benötigten Sicherheitsgruppen. Sie können einen Großteil des Wiederherstellungsprozesses automatisieren. Um zu erfahren wie das gemacht wird, sehen Sie sich diesen Blogbeitrag an..

5. Legen Sie fest und implementieren Sie, wie Sie den Datenverkehr bei Bedarf (im Notfall) zum Failover umleiten werden.

Dieser Failover-Vorgang kann entweder automatisch oder manuell eingeleitet werden. Ein automatisch eingeleitetes Failover auf der Grundlage von Zustandsprüfungen oder Alarmen ist mit Vorsicht zu genießen, da ein unnötiges Failover (Fehlalarm) Kosten wie Nichtverfügbarkeit und Datenverlust verursacht. Daher wird häufig eine manuell initiiertes Failover verwendet. In diesem Fall sollten Sie die Schritte für das Failover dennoch automatisieren, so dass die manuelle Auslösung wie ein Knopfdruck wirkt.

Bei der Inanspruchnahme von AWS-Services gibt es mehrere Optionen für die Verwaltung des Datenverkehrs zu berücksichtigen. Eine Option ist die Verwendung von Amazon Route 53. Mit Amazon Route 53 können Sie mehrere IP-Endpunkte in einem oder mehreren AWS-Regionen mit einem Route 53-Domainnamen verknüpfen. Um ein manuell initiiertes Failover zu implementieren, können Sie den Amazon Route 53 Application Recovery Controllerverwenden, der eine hochverfügbare API für die Datenebene bietet, um den Datenverkehr in die Wiederherstellungsregion umzuleiten. Verwenden Sie bei der Implementierung von Failover Vorgänge auf der Datenebene und vermeiden Sie solche auf der Steuerungsebene, wie in beschriebe in REL11-BP04 Nutzen der Datenebene und nicht der Steuerebene während der Wiederherstellung.

Mehr Information darüber sowie andere Optionen finden Sie in diesem Abschnitt des Disaster Recovery Whitepaper.

6. Entwerfen Sie einen Plan, wie Ihre Workload zurückgehen wird.

Failback bedeutet, dass Sie den Workload-Betrieb in der primären Region wieder aufnehmen, nachdem ein Notfallereignis abgeklungen ist. Die Bereitstellung von Infrastruktur und Code für die primäre Region erfolgt im Allgemeinen in denselben Schritten wie ursprünglich, wobei Infrastruktur als Code und Code-Bereitstellungspipelines verwendet werden. Die Herausforderung beim Failback ist

die Wiederherstellung von Datenspeichern und die Sicherstellung ihrer Konsistenz mit der in Betrieb befindlichen Wiederherstellungsregion.

Im ausgefallenen Zustand sind die Datenbanken in der Wiederherstellungsregion aktiv und verfügen über die aktuellen Daten. Ziel ist es dann, eine erneute Synchronisierung von der Wiederherstellungsregion mit der primären Region vorzunehmen, um sicherzustellen, dass diese auf dem neuesten Stand ist.

Einige AWS-Services werden das automatisch tun. Bei der Verwendung Amazon DynamoDB globaler Tabellen,setzt DynamoDB die Übertragung aller anstehenden Schreibvorgänge fort, auch wenn die Tabelle in der primären Region nicht mehr verfügbar ist, sobald sie wieder online ist. Bei der Verwendung von Amazon Aurora Global Database und verwaltetem geplanten Failoverwird die bestehende Replikationstopologie der Aurora globalen Datenbank wird beibehalten. Daher wird die ehemalige Lese-/Schreibinstance in der primären Region zu einem Replikat und erhält Aktualisierungen von der Wiederherstellungsregion.

In Fällen, in denen dies nicht automatisch geschieht, müssen Sie die Datenbank in der primären Region als Replikat der Datenbank in der Wiederherstellungsregion neu einrichten. In vielen Fällen bedeutet dies, dass die alte primäre Datenbank gelöscht und neue Replikate erstellt werden müssen. Eine Anleitung, wie Sie dies mit Amazon Aurora Global Database unter der Annahme eines ungeplanten Failover tun können, finden Sie beispielsweise in dieser Übung: Fail-Back einer globalen Datenbank.

Wenn Sie nach einem Failover in Ihrer Wiederherstellungsregion weiterarbeiten können, sollten Sie diese zur neuen Primärregion machen. Sie würden trotzdem alle oben genannten Schritte durchführen, um die ehemalige Primärregion in eine Wiederherstellungsregion zu verwandeln. Einige Unternehmen führen eine planmäßige Rotation durch und tauschen ihre Primär- und Wiederherstellungsregionen in regelmäßigen Abständen aus (z. B. alle drei Monate).

Alle für Failover und Failback erforderlichen Schritte sollten in einem Playbook festgehalten werden, das allen Teammitgliedern zur Verfügung steht und regelmäßig überprüft wird.

Grad des Aufwands für den Implementierungsplan:Hoch

Ressourcen

Ähnliche bewährte Methoden:

- the section called “REL09-BP01 Ermitteln und Sichern aller zu sichernden Daten oder Reproduzieren der Daten aus Quellen”

- the section called “REL11-BP04 Nutzen der Datenebene und nicht der Steuerebene während der Wiederherstellung”
- the section called “REL13-BP01 Definieren von Wiederherstellungszielen bei Ausfällen und Datenverlusten:”

Zugehörige Dokumente:

- AWS Architecture Blog: Notfallwiederherstellungsserie
- Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud (AWS-Whitepaper)
- Optionen für die Notfallwiederherstellung in der Cloud
- Entwickeln Sie eine Multi-Region-Serverless-Backend-Lösung, die aktiv/aktiv ist.
- Multi-Region-Serverless-Backend – neu aufgelegt
- RDS: Regionsübergreifendes Replizieren von Lesereplikaten
- Route 53: Konfigurieren von DNS-Failover
- S3: Regionsübergreifende Replikation
- Was ist AWS Backup?
- Was ist Route 53 Application Recovery Controller?
- AWS Elastic Disaster Recovery
- HashiCorp Terraform: Get Started – AWS (Erste Schritte)
- APN-Partner: Partner, die Sie bei der Notfallwiederherstellung unterstützen können
- AWS Marketplace: Für die Notfallwiederherstellung geeignete Produkte

Relevante Videos:

- Notfallwiederherstellung von Workloads auf AWS
- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2) (Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen)
- Erste Schritte mit AWS Elastic Disaster Recovery | Amazon Web Services

Ähnliche Beispiele:

- AWS Well-Architected Labs – Disaster Recovery (Notfallwiederherstellung) – Eine Reihe von Workshops zur Veranschaulichung der DR-Strategien

REL13-BP03 Testen der Implementierung der Notfallwiederherstellung zur Validierung:

Testen Sie regelmäßig den Failover zu Ihrem Wiederherstellungsstandort, um den ordnungsgemäßen Betrieb und die Einhaltung von RTO und RPO sicherzustellen.

Vom Erstellen selten durchgeführter Wiederherstellungspfade wird abgeraten. So könnten Sie beispielsweise einen zweiten Datenspeicher unterhalten, der nur für Leseabfragen verwendet wird. Wenn Sie Daten in einen Datenspeicher schreiben und der primäre Datenspeicher einen Fehler ausgibt, können Sie einen Failover auf den zweiten Datenspeicher durchführen. Wenn Sie diesen Failover nicht regelmäßig testen, werden Sie möglicherweise feststellen, dass Ihre Annahmen zu den Möglichkeiten des sekundären Datenspeichers unzutreffend sind. Die Kapazität des zweiten Datenspeichers, die bei den letzten Tests möglicherweise noch ausreichend war, genügt möglicherweise nicht mehr den Anforderungen dieses Szenarios. Unsere Erfahrungen haben gezeigt, dass bei einer Wiederherstellung nach einem Fehler nur der Pfad funktioniert, den Sie regelmäßig testen. Daher ist es ratsam, mehrere Wiederherstellungspfade zu pflegen. Sie können Wiederherstellungsmuster erstellen und diese regelmäßig testen. Auch komplexe oder kritische Wiederherstellungspfade müssen regelmäßig mittels Fehlersimulationen in der Produktion durchgeführt werden, um sicherzustellen, dass sie funktionieren. In dem gerade besprochenen Beispiel sollten Sie regelmäßig und unabhängig von der Erfordernis einen Failover auf die Standby-Ressourcen durchführen.

Gängige Antimuster:

- Failover werden nie in der Produktion durchgeführt.

Vorteile der Einführung dieser bewährten Methode: Durch regelmäßige Tests des Notfallwiederherstellungsplans wird sichergestellt, dass er bei Bedarf funktioniert und vom Team umgesetzt werden kann.

Risikostufe, falls diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

- Workloads für die Wiederherstellung auslegen. Testen Sie regelmäßig Ihre Wiederherstellungspfade. Mithilfe des Recovery Oriented Computing (wiederherstellungsorientiertes Computing) können Sie die für die Wiederherstellung

förderlichen Merkmale in Systemen identifizieren. Hierzu zählen: Isolation und Redundanz, die systemweite Fähigkeit zum Zurücksetzen von Änderungen, das Überwachen und Ermitteln des Systemzustands, die Bereitstellung von Diagnosen, automatisierte Wiederherstellungen, ein modulares Design und die Möglichkeit von Neustarts. Erproben Sie den Wiederherstellungspfad, um sicherzustellen, dass die Wiederherstellung innerhalb des vorgegebenen Zeitraums erfolgt und der vorgegebene Zustand erreicht wird. Dokumentieren Sie während dieser Wiederherstellung auftretende Probleme in Ihren Runbooks und suchen Sie vor dem nächsten Test nach Lösungen.

  - The Berkeley/Stanford Recovery-Oriented Computing (ROC) Project
- Verwenden Sie CloudEndure Disaster Recovery zum Implementieren und Testen Ihrer Strategie für die Notfallwiederherstellung.
  - Testen der Lösung für die Notfallwiederherstellung mit CloudEndure
  - CloudEndure Disaster Recovery
  - CloudEndure Disaster Recovery auf AWS

Ressourcen

Zugehörige Dokumente:

- APN-Partner: Partner, die Sie bei der Notfallwiederherstellung unterstützen können
- AWS Architecture Blog: Notfallwiederherstellungsserie
- AWS Marketplace: Für die Notfallwiederherstellung geeignete Produkte
- CloudEndure Disaster Recovery
- Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud (AWS-Whitepaper)
- Testen der Lösung für die Notfallwiederherstellung mit CloudEndure
- The Berkeley/Stanford Recovery-Oriented Computing (ROC) Project
- Was ist AWS Fault Injection Simulator?

Relevante Videos:

- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2) (Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen)
- AWS re:Invent 2019: Lösungen für Sicherung, Wiederherstellung und Notfallwiederherstellung mit AWS AWS (STG208)

Ähnliche Beispiele:

- AWS Well-Architected Labs: Testen der Ausfallsicherheit

REL13-BP04 Verwalten der Konfigurationsabweichungen am Standort oder in der Region der Notfallwiederherstellung:

Stellen Sie sicher, dass die Infrastruktur, die Daten und die Konfiguration am Standort oder in der Region der Notfallwiederherstellung den Anforderungen entsprechen. Sie sollten beispielsweise prüfen, ob AMIs und Service Quotas auf dem neuesten Stand sind.

AWS Config überwacht und zeichnet Ihre AWS-Ressourcenkonfigurationen kontinuierlich auf. Es kann Abweichungen erkennen und als Auslöser für AWS Systems Manager Automation dienen, um diese zu beheben und Warnmeldungen zu senden. AWS CloudFormation kann zusätzlich Abweichungen in bereitgestellten Stacks erkennen.

Gängige Antimuster:

- Versäumnis, Aktualisierungen an Ihren Wiederherstellungsstandorten vorzunehmen, wenn Sie Konfigurations- oder Infrastrukturänderungen an Ihren Hauptstandorten vornehmen.
- Mögliche Einschränkungen (z. B. Serviceunterschiede) an Ihren primären Standorten und den Standorten für die Notfallwiederherstellung werden nicht berücksichtigt.

Vorteile der Einführung dieser bewährten Methode: Wenn Ihre Umgebung für die Notfallwiederherstellung mit der vorhandenen Umgebung konsistent ist, gewährleisten dies eine vollständige Wiederherstellung.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Sicherstellen, dass die Bereitstellung an Haupt- und Sicherungsstandorte erfolgt Pipelines für die Bereitstellung von Anwendungen in der Produktion müssen die Anwendungen an alle Standorte verteilen, die in der Strategie für die Notfallwiederherstellung angegeben sind. Dazu gehören auch Entwicklungs- und Testumgebungen.
- Aktivieren von AWS Config zum Verfolgen von Standorten mit möglichen Abweichungen. Erstellen Sie mithilfe von AWS Config Regeln Systeme, die Ihre Strategien für die Notfallwiederherstellung durchsetzen und bei Erkennung von Abweichungen Warnungen generieren.
  - Korrigieren von nicht konformen AWS-Ressourcen mit AWS-Config-Regeln

  - AWS Systems Manager Automation
- Verwenden Sie AWS CloudFormation zur Bereitstellung Ihrer Infrastruktur. AWS CloudFormation kann Abweichungen zwischen den Angaben in den CloudFormation-Vorlagen und der tatsächlichen Bereitstellung erkennen.
  - AWS CloudFormation: Ermitteln von Abweichungen im gesamten CloudFormation-Stack

Ressourcen

Zugehörige Dokumente:

- APN-Partner: Partner, die Sie bei der Notfallwiederherstellung unterstützen können
- AWS Architecture Blog: Notfallwiederherstellungsserie
- AWS CloudFormation: Ermitteln von Abweichungen im gesamten CloudFormation-Stack
- AWS Marketplace: Für die Notfallwiederherstellung geeignete Produkte
- AWS Systems Manager Automation
- Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud (AWS-Whitepaper)
- Wie implementiere ich eine Lösung für die Verwaltung der Infrastrukturkonfiguration in AWS?
- Korrigieren von nicht konformen AWS-Ressourcen mit AWS-Config-Regeln

Relevante Videos:

- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2) (Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen)

REL13-BP05: Automatisieren der Wiederherstellung

Automatisieren Sie mit Tools von AWS oder Drittanbietern die Systemwiederherstellung und leiten Sie Datenverkehr zum Standort oder zur Region der Notfallwiederherstellung weiter.

Basierend auf konfigurierten Zustandsprüfungen können AWS-Services wie Elastic Load Balancing und AWS Auto Scaling die Last auf fehlerfreie Availability Zones verteilen während Services wie z. B, Amazon Route 53 und AWS Global Accelerator, die Last an fehlerfreie AWS-Regionen leiten können. Amazon Route 53 Application Recovery Controller hilft Ihnen, mithilfe von Bereitschaftsprüfungen und Routing-Steuerungsfunktionen Failover-Vorgänge zu verwalten und zu koordinieren. Diese Funktionen überwachen kontinuierlich die Fähigkeit Ihrer Anwendung, eine Wiederherstellung nach

Fehlern durchzuführen, so dass Sie die Wiederherstellung der Anwendung über mehrere AWS-Regionen, Availability Zones und On-Premises kontrollieren können.

Für Workloads in bestehenden physischen oder virtuellen Rechenzentren oder privaten Clouds, AWS Elastic Disaster Recovery,verfügbar durch AWS Marketplace, ermöglicht es Unternehmen, eine automatisierte Notfallwiederherstellungsstrategie auf AWS einzurichten. CloudEndure unterstützt auch die regions- bzw. AZ-übergreifende Notfallwiederherstellung in AWS.

Gängige Antimuster:

- Die Implementierung von identischem automatisiertem Failover und Failback kann bei einem Fehler zu Flapping führen.

Vorteile der Einführung dieser bewährten Methode: Die automatisierte Wiederherstellung verkürzt die Wiederherstellungszeit, da manuelle Fehler nicht mehr möglich sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Automatisieren von Wiederherstellungspfaden. Wenn in Szenarien mit hoher Verfügbarkeit kurze Wiederherstellungszeiten erforderlich sind, sind menschliche Beurteilungen und Aktionen zu langsam. Das System sollte in jeder Situation in der Lage sein, eine Wiederherstellung durchzuführen.
  - Verwenden Sie CloudEndure Disaster Recovery für automatisiertes Failover und Failback. CloudEndure Disaster Recovery repliziert Ihre Computer (einschließlich Betriebssystem, Systemstatuskonfiguration, Datenbanken, Anwendungen und Dateien) kontinuierlich in einen kostengünstigen Staging-Bereich in Ihrem AWS-Konto-Zielkonto und in Ihrer bevorzugten Region. Bei einem Notfall können Sie CloudEndure Disaster Recovery anweisen, innerhalb weniger Minuten automatisch Tausende Ihrer virtuellen Maschinen vollständig bereitgestellt zu starten.
    - Ausführen von Failover und Failback bei Notfallwiederherstellungen
    - CloudEndure Disaster Recovery

Ressourcen

Zugehörige Dokumente:

- APN-Partner: Partner, die Sie bei der Notfallwiederherstellung unterstützen können

- AWS Architecture Blog: Notfallwiederherstellungsserie
- AWS Marketplace: Für die Notfallwiederherstellung geeignete Produkte
- AWS Systems Manager Automation
- CloudEndure Disaster Recovery auf AWS
- Notfallwiederherstellung von Workloads auf AWS: Wiederherstellung in der Cloud (AWS-Whitepaper)

Relevante Videos:

- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2) (Architekturmuster für Aktiv-Aktiv-Anwendungen in mehreren Regionen)

Leistungseffizienz

Themen

- Auswahl
- Prüfen Sie die Angaben.
- Überwachung
- Kompromisse

Auswahl

Fragen

- LEIST 1 Was ist bei der Wahl einer leistungsfähigen Architektur zu beachten?
- LEIST 2 Was ist bei der Wahl der Datenverarbeitungslösung zu beachten?
- LEIST 3 Was ist bei der Wahl der Speicherlösung zu beachten?
- LEIST 4 Was ist bei der Wahl der Datenbanklösung zu beachten?
- LEIST 5 Was ist beim Konfigurieren der Netzwerklösung zu beachten?

LEIST 1 Was ist bei der Wahl einer leistungsfähigen Architektur zu beachten?

Oft sind mehrere Ansätze erforderlich, um die optimale Leistung für eine Workload zu erzielen. Gut geplante Systeme nutzen mehrere Lösungen und Funktionen zur Leistungsoptimierung.

Bewährte Methoden

- PERF01-BP01 Verstehen von verfügbaren Services und Ressourcen
- PERF01-BP02 Definieren eines Prozesses für die Wahl der Architektur
- PERF01-BP03 Einbeziehen von Kostenanforderungen in Entscheidungen
- PERF01-BP04 Verwenden von Richtlinien oder Referenzarchitekturen
- PERF01-BP05 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner
- PERF01-BP06 Benchmarking vorhandener Workloads
- PERF01-BP07 Durchführen von Lasttests für den Workload

PERF01-BP01 Verstehen von verfügbaren Services und Ressourcen

Informieren Sie sich über die vielfältigen Services und Ressourcen, die Ihnen in der Cloud zur Verfügung stehen. Bestimmen Sie die für Ihre Workload relevanten Services und Konfigurationsoptionen und bringen Sie in Erfahrung, wie Sie damit eine optimale Leistung erzielen.

Wenn Sie einen vorhandenen Workload evaluieren, müssen Sie einen Bestand der verschiedenen Services-Ressourcen generieren, den er verbraucht. Mit diesem Bestand können Sie prüfen, welche Komponenten durch verwaltete Services und neuere Technologien ersetzt werden können.

Gängige Antimuster:

- Sie verwenden die Cloud als gemeinsam genutztes Rechenzentrum.
- Sie nutzen freigegebenen Speicher für alle Objekte, die einen persistenten Speicher benötigen.
- Sie verwenden keine automatische Skalierung.
- Sie verwenden Instance-Typen, die am besten zu Ihren aktuellen Standards passen, bei Bedarf jedoch größer sind.
- Von Ihnen werden Technologien bereitgestellt und verwaltet, die als verwaltete Services verfügbar sind.

Vorteile der Einführung dieser bewährten Methode: Indem Sie unbekannte Services in Betracht ziehen, können Sie unter Umständen die Kosten der Infrastruktur und den Wartungsaufwand für Ihre Services erheblich reduzieren. Möglicherweise können Sie durch Bereitstellung neuer Services und Funktionen Markteinführungen beschleunigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Inventarisieren der Workload-Software und -Architektur für verwandte Services: Erstellen Sie ein Inventar Ihrer Workload und entscheiden Sie, über welche Kategorie von Produkten Sie mehr erfahren möchten. Ermitteln Sie die Workload-Komponenten, die zur Leistungssteigerung und Verminderung der betrieblichen Komplexität durch verwaltete Services ersetzt werden können.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network
- AWS-Lösungsbibliothek
- AWS Knowledge Center

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture: Expedia

Zugehörige Beispiele:

- AWS-Beispiele
- AWS-SDK-Beispiele

PERF01-BP02 Definieren eines Prozesses für die Wahl der Architektur

Nutzen Sie interne Erfahrungen und Kenntnisse im Zusammenhang mit der Cloud oder ziehen Sie externe Ressourcen heran, wie etwa veröffentlichte Anwendungsbeispiele, relevante Dokumentation oder Whitepapers, um einen Prozess zur Auswahl der geeigneten Ressourcen und Services festzulegen. Sie sollten einen Prozess definieren, der das Experimentieren und Benchmarking mit den Services fördert, die in Ihrer Workload verwendet werden könnten.

Berücksichtigen Sie beim Erstellen kritischer Benutzerszenarien für Ihre Architektur die Leistungsanforderungen. Geben Sie beispielsweise an, wie schnell jedes der kritischen Benutzerszenarien ausgeführt werden soll. Implementieren Sie für diese kritischen Szenarien

zusätzliche skriptbasierte Benutzerreisen, um ihre Leistung mit Ihren Anforderungen vergleichen zu können.

Gängige Antimuster:

- Sie gehen davon aus, dass Ihre aktuelle Architektur unverändert bleibt und im Laufe der Zeit nicht aktualisiert wird.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie begründen.

Vorteile der Einführung dieser bewährten Methode: Durch einen definierten Prozess zum Ändern der Architektur erhalten Sie die Möglichkeit, die gesammelten Daten langfristig in die Gestaltung der Workload einfließen zu lassen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Auswählen eines Architekturansatzes: Machen Sie die Art von Architektur ausfindig, die Ihre Leistungsanforderungen erfüllt. Ermitteln Sie Einschränkungen, etwa in Bezug auf die Medien für die Bereitstellung (Desktop, Web, Mobilgeräte, IoT), Anforderungen für Legacy-Systeme und Integrationen. Bestimmen Sie die Möglichkeiten der Wiederverwendung, einschließlich Refactoring. Konsultieren Sie andere Teams, Architekturdiagramme und Ressourcen wie AWS Solution Architects, AWS-Referenzarchitekturen und AWS-Partner, damit Ihnen die Wahl der Architektur leichter fällt.

Definieren von Leistungsanforderungen: Ermitteln Sie anhand der Kundenerfahrungen die wichtigsten Metriken. Identifizieren Sie für jede Kennzahl Ziel, Messverfahren und Priorität. Definieren Sie das Kundenerlebnis. Dokumentieren Sie die vom Kunden erwartete Leistung. Berücksichtigen Sie hierbei auch, wie Kunden die Leistung der Workload beurteilen. Räumen Sie bei kritischen User Stories problematischen Erlebnissen Priorität ein. Beziehen Sie Leistungsanforderungen mit ein und implementieren Sie skriptbasierte User Journeys, damit Sie nachvollziehen können, wie die Stories verglichen mit Ihren Anforderungen abschneiden.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network

- AWS-Lösungsbibliothek
- AWS Knowledge Center

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture: Expedia

Zugehörige Beispiele:

- AWS-Beispiele
- AWS-SDK-Beispiele

PERF01-BP03 Einbeziehen von Kostenanforderungen in Entscheidungen

Für den Betrieb von Workloads gelten oft bestimmte Kostenanforderungen. Verwenden Sie interne Kostenkontrollen, um Ressourcentypen und -größen entsprechend dem prognostizierten Ressourcenbedarf auszuwählen.

Ermitteln Sie, welche Workload-Komponenten durch vollständig verwaltete Services wie verwaltete Datenbanken, In-Memory-Caches und ETL-Services ersetzt werden können. Durch eine Reduzierung Ihrer betrieblichen Workload können Ressourcen vorwiegend auf Geschäftsergebnisse ausgerichtet werden.

Bewährte Methoden für Kostenanforderungen finden Sie im Abschnitt Kostengünstige Ressourcen im Whitepaper zur Säule der Kostenoptimierung.

Gängige Antimuster:

- Sie verwenden nur eine Instance-Familie.
- Sie bewerten keine lizenzierten Lösungen im Vergleich zu Open-Source-Lösungen.
- Sie nutzen nur Blockspeicher.
- Sie stellen gängige Software in EC2-Instances sowie in Amazon EBS- oder flüchtigen Volumes bereit, die als verwalteter Service verfügbar sind.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie die Kosten bei der Auswahl berücksichtigen, können Sie andere Investitionen tätigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Optimieren der Workload-Komponenten zur Kostensenkung: Dimensionieren Sie Workload-Komponenten richtig und ermöglichen Sie Elastizität, um Kosten zu senken und die Effizienz der Komponenten zu maximieren. Ermitteln Sie, welche Workload-Komponenten gegebenenfalls durch verwaltete Services ersetzt werden können, z. B. verwaltete Datenbanken, In-Memory-Caches und Reverse-Proxys.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network
- AWS-Lösungsbibliothek
- AWS Knowledge Center
- AWS Compute Optimizer

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture: Expedia
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)

Zugehörige Beispiele:

- AWS-Beispiele
- AWS-SDK-Beispiele
- Die richtige Dimensionierung ermitteln, wenn Compute Optimizer und die Arbeitsspeicherauslastung aktiviert sind
- AWS Compute Optimizer-Demo-Code

PERF01-BP04 Verwenden von Richtlinien oder Referenzarchitekturen

Maximieren Sie die Leistung und Effizienz, indem Sie interne Richtlinien und vorhandene Referenzarchitekturen evaluieren und anhand Ihrer Analyse Services und Konfigurationen für Ihre Workload auswählen.

Gängige Antimuster:

- Sie erlauben eine Auswahl vielfältiger Technologien, was sich auf den Verwaltungsaufwand Ihres Unternehmens auswirken kann.

Vorteile der Einführung dieser bewährten Methode: Durch Festlegung einer Richtlinie für die Architektur-, Technologie und Anbieterauswahl können Entscheidungen schnell getroffen werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Bereitstellen des Workloads mithilfe vorhandener Richtlinien und Referenzarchitekturen: Integrieren Sie die Services in Ihre Cloud-Bereitstellung. Stellen Sie anschließend anhand von Leistungstests sicher, dass Sie die eigenen Leistungsanforderungen weiterhin erfüllen können.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network
- AWS-Lösungsbibliothek
- AWS Knowledge Center

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture: Expedia

Zugehörige Beispiele:

- AWS-Beispiele

- AWS-SDK-Beispiele

PERF01-BP05 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner

Greifen Sie bei Ihren Entscheidungen auf die Ressourcen von Cloud-Unternehmen, wie etwa Lösungsarchitekten, oder auf professionelle Services oder einen geeigneten Partner zurück. Diese Ressourcen können Ihnen dabei helfen, Ihre Architektur zu überprüfen und zu verbessern, um so die Leistung zu optimieren.

Wenden Sie sich an AWS, wenn Sie zusätzliche Anleitungen oder Produktinformationen benötigen. AWS Solutions Architects und AWS Professional Services liefern Ratschläge für die Implementierung von Lösungen. AWS-Partner bieten AWS-Fachwissen, damit Sie in Ihrem Unternehmen flexibel agieren und Innovationen nutzen können.

Gängige Antimuster:

- Sie nutzen AWS als üblichen Anbieter von Rechenzentren.
- Sie verwenden AWS-Services auf unvorgesehene Weise.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie sich mit Ihrem Anbieter oder einem Partner beraten, können Sie Entscheidungen mit größerer Zuversicht treffen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Anfordern von Unterstützung bei AWS-Ressourcen: AWS Solutions Architects und Professional Services liefern Ratschläge für die Implementierung von Lösungen. APN-Partner bieten AWS-Fachwissen, damit Sie in Ihrem Unternehmen flexibel agieren und Innovationen nutzen können.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network
- AWS-Lösungsbibliothek
- AWS Knowledge Center

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture: Expedia

Zugehörige Beispiele:

- AWS-Beispiele
- AWS-SDK-Beispiele

PERF01-BP06 Benchmarking vorhandener Workloads

Führen Sie einen Benchmark-Vergleich für eine vorhandene Workload durch, um sich ein Bild über deren Leistung in der Cloud zu verschaffen. Nutzen Sie die beim Benchmarking erfassten Daten als Grundlage für architektonische Entscheidungen.

Kombinieren Sie Benchmarking mit synthetischen Tests und der Überwachung echter Benutzer, um Daten zur Leistung Ihrer Workload-Komponenten zu generieren. Benchmarking lässt sich in der Regel schneller als Lasttests einrichten und dient zur Bewertung der Technologie einer bestimmten Komponente. Ein Benchmarking wird oft zu Beginn eines neuen Projekts durchgeführt, wenn Sie noch keine vollständige Lösung für einen Lasttest haben.

Sie können wahlweise eigene Benchmark-Tests erstellen oder branchenübliche Standardtests verwenden, wie etwa TPC-DS für das Benchmarking Ihrer Data-Warehousing-Workloads. Branchen-Benchmarks sind zum Vergleich von Umgebungen nützlich. Benutzerdefinierte Benchmarks eignen sich zum Prüfen spezieller Arten von Vorgängen, die Sie in der Architektur ausführen möchten.

Beim Benchmarking ist es wichtig, die Testumgebung entsprechend vorzubereiten, um aussagekräftige Ergebnisse zu erzielen. Führen Sie zur Ermittlung aller Varianzen im Laufe der Zeit mehrmals denselben Benchmark-Test aus.

Da sich Benchmarks in der Regel schneller als Lasttests ausführen lassen, können Sie früher in der Bereitstellungspipeline eingesetzt werden und schneller Feedback zu Leistungsabweichungen liefern. Wenn Sie eine wesentliche Veränderung einer Komponente oder eines Services bewerten, können Sie schnell ermitteln, ob der Aufwand für die Korrektur gerechtfertigt ist. Die Verwendung von Benchmarking in Verbindung mit Lasttests ist wichtig, da letztere Auskunft über die Leistung der Workload in der Produktion geben.

Gängige Antimuster:

- Sie verlassen sich auf gängige Benchmarks, die für Ihre Workload-Merkmale nicht aufschlussreich sind.
- Sie verlassen sich auf Kundenfeedback und Kundenwahrnehmung als einzige Benchmark.

Vorteile der Einführung dieser bewährten Methode: Durch das Benchmarking Ihrer aktuellen Implementierung können Sie die Leistungssteigerung messen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Leistung während der Entwicklung überwachen: Implementieren Sie Prozesse, die Ihnen Einblick in die Leistung gewähren, während sich Ihr Workload entwickelt.

Integrieren in eigene Bereitstellungspipeline: Führen Sie automatisch Lasttests in Ihrer Bereitstellungspipeline aus. Vergleichen Sie die Testergebnisse mit vordefinierten Key Performance Indicators (KPIs, Leistungskennzahlen) und Schwellenwerten, damit die Leistungsanforderungen weiterhin erfüllt werden.

Testen von User Journeys: Verwenden Sie für Lasttests synthetische oder bereinigte Daten (d. h. entfernen Sie sensible oder personenbezogene Informationen). Testen Sie die gesamte Architektur intensiv, indem Sie wiedergegebene oder vorprogrammierte Benutzerreisen durch Ihre Anwendung verwenden.

Überwachung echter Benutzer: Verwenden Sie CloudWatch RUM, um clientseitige Daten über Ihre Anwendungsleistung zu erfassen und anzuzeigen. Verwenden Sie diese Daten, um die Leistungs-Benchmarks für echte Benutzer festzulegen.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- AWS Partner Network
- AWS-Lösungsbibliothek
- AWS Knowledge Center
- Amazon CloudWatch RUM

- Amazon CloudWatch Synthetics

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)
- Demo of Amazon CloudWatch Synthetics (Demo von CW Synthetics)

Zugehörige Beispiele:

- AWS-Beispiele
- AWS-SDK-Beispiele
- Verteilte Belastungstests
- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client

PERF01-BP07 Durchführen von Lasttests für den Workload

Stellen Sie Ihre neueste Workload-Architektur mit verschiedenen Ressourcentypen und -größen in der Cloud bereit. Überwachen Sie die Bereitstellung, um Leistungsmetriken zu erfassen, die Engpässe oder überschüssige Kapazität erkennen lassen. Nutzen Sie diese Leistungsdaten, um die Architektur zu entwerfen oder zu verbessern und eine bessere Auswahl von Ressourcen zu treffen.

Bei Lasttests wird der tatsächliche Workload herangezogen. So lässt sich feststellen, wie leistungsfähig Ihre Lösung in einer Produktionsumgebung ist. Verwenden Sie für Lasttests synthetische oder bereinigte Daten und entfernen Sie sensible oder personenbezogene Informationen. Verwenden Sie progressiv wiedergegebene oder vorprogrammierte Benutzerreisen durch Ihre Workload, um die gesamte Architektur zu testen. Führen Sie automatisch Lasttests als Teil Ihrer Bereitstellungs-Pipeline durch und vergleichen Sie die Ergebnisse mit vordefinierten KPIs und Schwellenwerten. So wird sichergestellt, dass Sie weiterhin die erforderliche Leistung erreichen.

Gängige Antimuster:

- Sie führen Lasttests für einzelne Teile der Workload durch, aber nicht für die gesamte Workload.

- Sie führen Lasttests in einer Infrastruktur durch, die sich von Ihrer Produktionsumgebung unterscheidet.
- Sie führen Lasttests nur für die erwartete Last durch und nicht für noch größere Lasten, um mögliche künftige Probleme besser vorherzusehen.
- Sie führen Lasttests durch, ohne den AWS Support zu informieren. Die Tests sind jedoch nutzlos, da sie wie Denial-of-Service-Vorfälle aussehen.

Vorteile der Einführung dieser bewährten Methode: Die Messung der Leistung im Rahmen eines Lasttests gibt Aufschluss darüber, wo bei zunehmender Last mit Auswirkungen zu rechnen ist. Auf diese Weise können Sie erforderliche Änderungen vorhersehen, bevor sie sich auf Ihre Workload auswirken.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Validieren des Ansatzes mittels Lasttests: Führen Sie einen Lasttest für einen Machbarkeitsnachweis durch, um festzustellen, ob die Leistungsanforderungen erfüllt werden. Mithilfe von AWS-Services können Sie Umgebungen im Produktionsmaßstab ausführen und damit Ihre Architektur testen. Da Sie für die Testumgebung nur bei Nutzung bezahlen, können Sie umfassende Tests zu einem Bruchteil der Kosten durchführen, die bei Verwendung einer lokalen Umgebung anfallen würden.

Überwachen von Metriken: Mithilfe von CloudWatch lassen sich Kennzahlen aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können auch benutzerdefinierte Kennzahlen erfassen und in Oberflächen-, Geschäfts- oder abgeleiteten Kennzahlen veröffentlichen. Richten Sie mit CloudWatch oder mit Lösungen von Drittanbietern Alarme ein, die auf das Überschreiten von Schwellenwerten hinweisen.

Bedarfsgerechte Tests: Bei Lasttests wird die tatsächliche Workload herangezogen. So lässt sich feststellen, wie leistungsfähig Ihre Lösung in einer Produktionsumgebung ist. Mithilfe von AWS-Services können Sie Umgebungen im Produktionsmaßstab ausführen und damit Ihre Architektur testen. Da Sie für die Testumgebung nur bei Nutzung bezahlen, können Sie umfassende Tests zu geringeren Kosten durchführen, als bei Verwendung einer lokalen Umgebung anfallen würden. Testen Sie Ihren Workload mithilfe der AWS Cloud, um zu ermitteln, an welcher Stelle er nicht skalierbar ist oder ob die Skalierung nicht-linear erfolgt. Nutzen Sie beispielsweise Spot Instances, um kostengünstig Lasten zu erzeugen und Engpässe zu identifizieren, bevor diese in der Produktionsumgebung auftreten.

Ressourcen

Zugehörige Dokumente:

- AWS CloudFormation
- Building AWS CloudFormation Templates using CloudFormer (Erstellen von CFN-Vorlagen mit CloudFormer)
- Amazon CloudWatch RUM
- Amazon CloudWatch Synthetics
- Distributed Load Testing on AWS (Verteilte Lasttests auf AWS)

Relevante Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)
- Demo of Amazon CloudWatch Synthetics (Demo von CW Synthetics)

Zugehörige Beispiele:

- Distributed Load Testing on AWS (Verteilte Lasttests auf AWS)

LEIST 2 Was ist bei der Wahl der Datenverarbeitungslösung zu beachten?

Die optimale Datenverarbeitungslösung für eine Workload ist vom Anwendungsdesign sowie von Nutzungsmustern und Konfigurationseinstellungen abhängig. Architekturen können unterschiedliche Datenverarbeitungslösungen für verschiedene Komponenten verwenden und unterschiedliche Funktionen zur Leistungsverbesserung unterstützen. Die Wahl der falschen Datenverarbeitungslösung für eine Architektur kann die Leistungseffizienz schmälern.

Bewährte Methoden

- PERF02-BP01 Prüfen von verfügbaren Datenverarbeitungsoptionen
- PERF02-BP02 Verstehen verfügbarer Konfigurationsoptionen für die Datenverarbeitung
- PERF02-BP03 Erfassen von Datenverarbeitungsmetriken
- PERF02-BP04 Bestimmen der erforderlichen Konfiguration durch Dimensionieren

- PERF02-BP05 Nutzen verfügbarer Elastizität von Ressourcen
- PERF02-BP06 Neue Bewertung von Datenverarbeitungsbedarf anhand von Metriken

PERF02-BP01 Prüfen von verfügbaren Datenverarbeitungsoptionen

Erfahren Sie, wie Ihre Workload vom Einsatz unterschiedlicher Datenverarbeitungsoptionen wie Instances, Container und Funktionen profitieren kann.

Gewünschtes Ergebnis: Indem Sie alle verfügbaren Datenverarbeitungsoptionen verstehen, erkennen Sie die Möglichkeiten zur Leistungsverbesserung, zum Verringern von unnötigen Infrastrukturkosten und zum Reduzieren des Aufwands, um Ihre Workload zu verwalten. Zudem können Sie durch Bereitstellung neuer Services und Funktionen Markteinführungen beschleunigen.

Gängige Antimuster:

- Verwenden der gleichen Datenverarbeitungslösung bei einer Post-Migration-Workload, die On-Premises eingesetzt wurde.
- Fehlendes Bewusstsein für Cloud-Datenverarbeitungslösungen und wie diese Lösungen Ihre Datenverarbeitungsleistung verbessern können.
- Überdimensionieren einer bestehenden Datenverarbeitungslösung, um Skalierungs- oder Leistungsanforderungen zu erfüllen, wenn eine alternative Datenverarbeitungslösung Ihren Workload-Merkmalen besser entsprechen würde.

Vorteile der Einführung dieser bewährten Methode: Indem Sie die Datenverarbeitungsanforderungen ermitteln und die verfügbaren Datenverarbeitungslösungen evaluieren, verstehen Business-Stakeholder und Entwicklungsteams die Vorteile und Einschränkungen der ausgewählten Datenverarbeitungslösung. Die ausgewählte Datenverarbeitungslösung sollte den Kriterien für die Workload-Leistung entsprechen. Wesentliche Kriterien umfassen Anforderungen an Datenverarbeitung, Datenverkehrsmuster, Datenzugriffsmuster, Skalierung und Latenz.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Machen Sie sich mit den Lösungen zur Virtualisierung, Containerisierung und Verwaltung vertraut, von denen Ihre Workload profitieren kann und die Ihren Leistungsanforderungen entsprechen. Eine Workload kann unterschiedliche Arten von Datenverarbeitungslösungen enthalten. Jede Datenverarbeitungslösung zeichnet sich durch andere Eigenschaften aus. Basierend auf der Skala

Ihrer Workload und Ihrer Datenverarbeitungsanforderungen kann eine Datenverarbeitungslösung ausgewählt und für Ihre Bedürfnisse konfiguriert werden. Der Cloud-Architekt sollte die Vorteile und Nachteile von Instances, Containern und Funktionen kennenlernen. Die folgenden Schritte helfen Ihnen beim Auswählen Ihrer Datenverarbeitungslösung, die Ihren Workload-Eigenschaften und Leistungsanforderungen entspricht.

Typ	Server	Container	Funktion
AWS-Service	Virtuelle Server-Instances in der Amazon Elastic Compute Cloud (Amazon EC2)	Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS)	AWS Lambda
Schlüsselmerkmale	Es gibt eine dediziert e Option für die Anforderungen an Hardwarel izenzen, Platzieru ngsoptionen und eine große Auswahl von unterschiedlichen Instance-Familien basierend auf Datenverarbeitungs metriken	Einfache Bereitste llung, konsisten te Umgebungen, wird auf EC2-Insta nces ausgeführt, ist skalierbar	Kurze Laufzeit (15 Minuten oder kürzer), der maximale Arbeitsspeicher und die CPU sind nicht so hoch wie bei anderen Services, verwaltete Hardwaree bene, skaliert auf Millionen gleichzei tiger Anforderungen
Gängige Anwendung sfälle	Lift-and-Shift-Mig rationen, monolithi sche Anwendung, hybride Umgebungen, Enterprise-Anwendu ngen	Microservices, Hybrid-Umgebungen	Microservices, ereignisgesteuerte Anwendungen

Implementierungsschritte:

1. Wählen Sie den Ort aus, an dem sich die Datenverarbeitungslösung befinden soll, indem Sie the section called “PERF05-BP06 Auswählen des Workload-Standortes entsprechend den Netzwerkanforderungen”evaluieren. Dieser Standort schränkt die für Sie verfügbaren Arten von Rechenlösungen ein.
2. Identifizieren Sie die Art der Datenverarbeitungslösung, die am besten mit den Anforderungen an den Standort und die Anwendung funktioniert.
    a. Virtuelle Server-Instances in der Amazon Elastic Compute Cloud (Amazon EC2) sind in vielen unterschiedlichen Familien und Größen verfügbar. Sie bieten eine Vielzahl von Optionen wie Solid-State-Laufwerken (SSDs) und Grafikprozessoren (Graphics Processing Units, GPUs). EC2-Instances bieten bei der Auswahl von Instances die größte Flexibilität. Wenn Sie eine EC2-Instance starten, wird anhand des von Ihnen festgelegten Instance-Typs die Hardware für Ihre Instance ermittelt. Jeder Instance-Typ umfasst andere Datenverarbeitungs-, Arbeitsspeicher- und Speicheroptionen. Instance-Typen werden anhand dieser Optionen in Instance-Familien gruppiert. Typische Anwendungsfälle umfassen: das Ausführen von Enterprise-Anwendungen, High Performance Computing (HPC), das Trainieren und Bereitstellen von Machine-Learning-Anwendungen und das Ausführen von cloudnativen Anwendungen.
    b. Amazon Elastic Container Service (Amazon ECS) ist ein vollständig verwalteter Service zur Container-Orchestrierung, mit dem Sie Container in einem Cluster aus EC2-Instances oder Serverless-Instances mit AWS Fargate automatisch ausführen und verwalten können. Sie können Amazon ECS zusammen mit anderen Services wie Amazon Route 53, Secrets Manager, AWS Identity and Access Management (IAM) und Amazon CloudWatch verwenden. Amazon ECS ist empfehlenswert, wenn Ihre Anwendung containerisiert ist und Ihr Entwicklungsteam Docker-Container bevorzugt.
    c. Amazon Elastic Kubernetes Service (Amazon EKS) ist ein vollständig verwalteter Kubernetes-Service. Sie können Ihre EKS-Cluster mit AWS Fargate ausführen, sodass keine Server mehr bereitgestellt und verwaltet werden müssen. Die Verwaltung von Amazon EKS wird durch Integrationen mit AWS-Services wie Amazon CloudWatch, Auto-Scaling-Gruppen, AWS Identity and Access Management (IAM) und Amazon Virtual Private Cloud (VPC) vereinfacht. Wenn Sie Container einsetzen, müssen Sie Datenverarbeitungsmetriken verwenden, um den optimalen Typ für Ihre Workload zu ermitteln, ähnlich wie Sie Ihre Datenverarbeitungsmetriken verwenden, um Ihre EC2- oder AWS Fargate-Instance-Typen auszuwählen. Amazon EKS wird empfohlen, wenn Ihre Anwendung containerisiert ist und Ihr Entwicklungsteam Kubernetes-Container gegenüber Docker-Containern bevorzugt.
    d. Sie können AWS Lambda verwenden, um Code auszuführen, der die erlaubte Laufzeit, den Speicher und die CPU-Optionen unterstützt. Laden Sie einfach Ihren Code hoch und AWS

Lambda verwaltet alles, was zum Ausführen und Skalieren des Codes erforderlich ist. Ihr Code kann automatisch über andere AWS-Services ausgelöst werden oder Sie können ihn direkt aufrufen. Lambda wird für kurz ausgeführte Microservice-Architekturen empfohlen, die für die Cloud entwickelt wurden.

3. Nachdem Sie mit Ihrer neuen Datenverarbeitungslösung experimentiert haben, planen Sie Ihre Migration und überprüfen Sie Ihre Leistungsmetriken. Dies ist ein kontinuierlicher Prozess, siehe the section called “PERF02-BP04 Bestimmen der erforderlichen Konfiguration durch Dimensionieren”evaluieren.

Grad des Aufwands für den Implementierungsplan: Wenn eine Workload von einer Datenverarbeitungslösung zu einer anderen verschoben wird, stellt dies möglicherweise einen mittleren Grad des Aufwands beim Faktorwechsel der Anwendung dar.

Ressourcen

Ähnliche Dokumente:

- Cloud Computing mit AWS
- EC2-Instance-Typen
- Steuerung des Prozessorzustands für Ihre EC2-Instance
- EKS-Container: EKS-Worker-Knoten
- Amazon ECS-Container: Amazon ECS-Container-Instances
- Funktionen: Lambda-Funktionskonfiguration
- Prescriptive Guidance für Container
- Prescriptive Guidance für Serverless

Ähnliche Videos:

- Datenverarbeitungsoptionen auswählen
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)
- Amazon EC2-Grundlagen (CMP211-R2)
- Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems
- Bereitstellen leistungsstarker ML-Inferenzen mit AWS Inferentia (CMP324-R1)
- Bessere, schnellere und kostengünstigere Datenverarbeitung: Kostenoptimierung bei Amazon EC2 (CMP202-R1)

Ähnliche Beispiele:

- Migration der Webanwendung zu Containern
- Ausführen eines Serverless-„Hello World“

PERF02-BP02 Verstehen verfügbarer Konfigurationsoptionen für die Datenverarbeitung

Jede Datenverarbeitungslösung hat verfügbare Optionen und Konfigurationen, um die Merkmale Ihrer Workload zu unterstützen. Erfahren Sie, wie die verschiedenen Optionen Ihre Workloads ergänzen und welche Konfigurationsoptionen am besten für Ihre Anwendung geeignet sind. Beispiele für diese Optionen sind Instance-Familien, -Größen, -Merkmale (GPU, I/O), Bursting, Zeitüberschreitungen, Funktionsgrößen, Container-Instances und Gleichzeitigkeit.

Gewünschtes Ergebnis: Die Workload-Merkmale, einschließlich CPU, Arbeitsspeicher, Netzwerkdurchsatz, GPU, IOPS, Datenverkehrsmuster und Datenzugriffsmuster, werden dokumentiert und verwendet, um die Datenverarbeitungslösung so zu konfigurieren, dass Sie den Workload-Merkmalen entspricht. Jede dieser Metriken sowie benutzerspezifische Metriken, die für Ihre Workload spezifisch sind, werden aufgezeichnet, überwacht und dann verwendet, um die Datenverarbeitungskonfiguration zu optimieren, damit sie bestmöglich Ihre Anforderungen erfüllt.

Gängige Antimuster:

- Verwenden der gleichen Datenverarbeitungslösung, die On-Premises eingesetzt wurde.
- Die Datenverarbeitungsoptionen oder die Instance-Familie werden nicht überprüft, damit sie den Workload-Merkmalen entsprechen.
- Die Datenverarbeitung ist überdimensioniert, um Bursting-Kapazitäten zu gewährleisten.
- Sie verwenden mehrere Plattformen zur Datenverarbeitungsverwaltung für ein und dieselbe Workload.

Vorteile der Einführung dieser bewährten Methode: Sie müssen mit den Datenverarbeitungsangeboten von AWS vertraut sein, damit Sie die richtige Lösung für die einzelnen Workloads bestimmen können. Nachdem Sie die Datenverarbeitungsangebote für Ihre Workload ausgewählt haben, können Sie anhand von schnellen Experimenten mit diesen Angeboten feststellen, wie gut sie Ihren Workload-Anforderungen entsprechen. Eine Datenverarbeitungslösung, die für Ihre Workload-Eigenschaften optimiert ist, steigert Ihre Leistung, verringert Ihre Kosten und erhöht Ihre Zuverlässigkeit.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Wenn Ihre Workload die gleiche Rechenoption für mehr als vier Wochen verwendet hat und sie davon ausgehen, dass die Eigenschaften in Zukunft gleich bleiben, können Sie AWS Compute Optimizer verwenden, um eine Empfehlung basierend auf Ihren Rechenmerkmalen zu erhalten. Wenn AWS Compute Optimizer nicht in Frage kommt, da Metriken fehlen, es sich im einen nicht unterstützten Instance-Typ handelt oder sich eine vorhersehbare Änderung in Ihren Merkmalen ereignen kann, müssen Sie Ihre Metriken basierend auf Lasttests und Experimenten vorhersagen.

Implementierungsschritte:

1. Führen Sie EC2-Instances oder -Container mit dem EC2-Starttyp aus?
    a. Kann Ihre Workload GPUs zur Erhöhung der Leistung verwenden?
        i. Beschleunigte Computing-Instances sind GPU-basierte Instances, die die höchste Leistung für Machine-Learning-Training, Inferenz und High Performance Computing bieten.
    b. Führt Ihre Workload Anwendungen zur Machine-Learning-Inferenz aus?
        i. AWS Inferentia (Inf1) – Inf1-Instances wurden entwickelt, um Machine Learning-Inferenzanwendungen zu unterstützen. Mithilfe von Inf1-Instances können Kunden umfangreiche Inferenzanwendungen für Machine Learning wie Bilderkennung, Spracherkennung, Verarbeitung natürlicher Sprache, Personalisierung und Betrugserkennung ausführen. Sie können ein Modell in einem der gängigen Machine Learning-Frameworks wie TensorFlow, PyTorch oder MXNet erstellen und GPU-Instances verwenden, um Ihr Modell zu schulen. Nachdem Ihr Machine Learning-Modell geschult wurde, um Ihre Anforderungen zu erfüllen, können Sie es auf Inf1-Instances bereitstellen. Dazu verwenden Sie AWS Neuron, ein spezialisiertes Software Development Kit (SDK), das aus einem Compiler, einer Laufzeit und Tools zur Profilerstellung besteht, die die Machine Learning-Inferenzleistung von Inferentia-Chips optimieren.
    c. Lässt sich Ihre Workload mit Ihren grundlegenden Hardwarekomponenten integrieren, um die Leistung zu verbessern?
        i. Field Programmable Gate Arrays (FPGAs) – Mit FPGAs können Sie Workloads mithilfe einer benutzerdefinierten Hardwarebeschleunigung für die anspruchsvollsten Workloads optimieren. Zum Definieren der Algorithmen bieten sich gängige unterstützte Programmiersprachen wie C oder Go sowie hardwareorientierte Sprachen wie Verilog oder VHDL an.
    d. Verfügen Sie über mindestens vier Wochen an Metriken und können vorhersagen, dass Ihre Datenverkehrsmuster und -metriken in Zukunft ungefähr gleich bleiben werden?

   i. Verwenden Sie Compute Optimizer, um eine Machine-Learning-Empfehlung dazu zu erhalten, welche Datenverarbeitungskonfiguration am besten Ihren Datenverarbeitungsmerkmalen entspricht.

e. Ist Ihre Workload-Leistung durch CPU-Metriken eingeschränkt?

   i. Rechenoptimierte Instances eignen sich hervorragend für Workloads, die leistungsstarke Prozessoren erfordern.

f. Ist Ihre Workload-Leistung durch Arbeitsspeichermetriken eingeschränkt?

   i. Arbeitsspeicheroptimierte Instances bieten große Mengen an Arbeitsspeicher, um arbeitsspeicherintensive Workloads zu unterstützen.

g. Ist Ihre Workload-Leistung durch IOPS eingeschränkt?

   i. Speicheroptimierte Instances wurden für Workloads entworfen, die hohen, sequenziellen Lese- und Schreibzugriff (IOPS) auf lokalen Speicher erfordern.

h. Stellen Ihre Workload-Eigenschaften einen ausgewogenen Bedarf hinsichtlich aller Metriken dar?

   i. Benötigt Ihre Workload-CPU Burst-Kapazitäten, um Spitzen beim Datenverkehr zu bewältigen?

      A. Instances mit Spitzenlastleistung ähneln für Datenverarbeitung optimierten Instances mit dem Unterschied, dass sie eine Burst-Kapazität über die feste CPU-Baseline hinaus bieten, die in einer für Datenverarbeitung optimierten Instance festgelegt ist.

   ii. Allzweck- Instances bieten eine ausgewogene Mischung aller Merkmale, um unterschiedliche Workloads zu unterstützen.

i. Wird Ihre Datenverarbeitungs-Instance auf Linux ausgeführt und ist durch den Netzwerkdurchsatz auf der Netzwerkschnittstellenkarte eingeschränkt?

   i. Lesen Sie Leistungsfrage 5, Bewährte Methoden 2: Evaluieren der verfügbaren Netzwerkfunktionen, um den entsprechenden Instance-Typ und die Instance-Familie zu ermitteln, die Ihren Leistungsanforderungen entsprechen.

j. Benötigt Ihre Workload konsistente und vorhersehbare Instances in einer bestimmten Availability Zone, an die Sie sich für ein Jahr binden können?

   i. Reserved Instances bestätigen Kapazitätsreservierungen in einer bestimmten Availability Zone. Reserved Instances eignen sich optimal für die erforderliche Rechenleistung in einer bestimmten Availability Zone.

k. Hat Ihre Workload Lizenzen, die dedizierte Hardware erfordern?

   i. Dedicated Hosts unterstützen vorhandene Softwarelizenzen und helfen Ihnen bei der Erfüllung von Compliance-Anforderungen.

l. Verfügt Ihre Datenverarbeitungslösung über eine Burst-Funktion und erfordert sie synchrone Verarbeitung?

   i. Mit On-Demand-Instances können Sie die Datenverarbeitungskapazität nach Sekunde oder Stunde ohne langfristige Verpflichtungen verwenden. Diese Instances eignen sich für sich Bursting über die Leistungsbasis hinaus.

m. Ist Ihre Datenverarbeitungslösung zustandslos, fehlertolerant und asynchron?

   i. Spot Instances erschließen ungenutzte Instance-Kapazitäten für Ihre zustandslosen, fehlertoleranten Workloads.

2. Verwenden Sie Container auf Fargate?

   a. Ist Ihre Task-Leistung durch den Arbeitsspeicher oder die CPU-Leistung eingeschränkt?

      i. Verwenden Sie die Task-Größe, um Ihren Arbeitsspeicher oder Ihre CPU anzupassen.

   b. Wird Ihre Leistung von Ihren Datenverkehr-Bursts beeinträchtigt?

      i. Verwenden Sie die Auto-Scaling-Konfiguration, damit sie Ihren Datenverkehrsmustern entspricht.

3. Befindet sich Ihre Datenverarbeitungslösung auf Lambda?

   a. Verfügen Sie über mindestens vier Wochen an Metriken und können vorhersagen, dass Ihre Datenverkehrsmuster und -metriken in Zukunft ungefähr gleich bleiben werden?

      i. Verwenden Sie Compute Optimizer, um eine Machine-Learning-Empfehlung dazu zu erhalten, welche Datenverarbeitungskonfiguration am besten Ihren Datenverarbeitungsmerkmalen entspricht.

   b. Haben Sie nicht ausreichend Metriken, um AWS Compute Optimizer zu verwenden?

      i. Wenn Sie keine verfügbaren Metriken haben, um Compute Optimizer zu verwenden, nutzen Sie AWS Lambda Power Tuning, um die beste Konfiguration zu finden.

   c. Ist Ihre Funktionsleistung durch den Arbeitsspeicher oder die CPU-Leistung eingeschränkt?

      i. Konfigurieren Sie Ihren Lambda-Arbeitsspeicher, damit er Ihren benötigten Leistungsmetriken entspricht.

   d. Überschreitet Ihre Funktion das Zeitlimit bei der Ausführung?

      i. Ändern Sie die Timeout-Einstellungen.

   e. Wird Ihre Funktionsleistung durch Aktivitäts- und Gleichzeitigkeits-Bursts eingeschränkt?

   i. Konfigurieren Sie die Gleichzeitigkeitseinstellungen, damit sie Ihren Leistungsanforderungen entsprechen.

f. Wird Ihre Funktion asynchron ausgeführt und fällt bei wiederholten Versuchen aus?

   i. Konfigurieren Sie das maximale Alter des Ereignisses und die Höchstzahl von Wiederholungen in den Einstellungen für die asynchrone Konfiguration .

Grad des Aufwands für den Implementierungsplan:

Sie müssen Ihre aktuellen Recheneigenschaften und -metriken kennen, um diese bewährten Methoden einzurichten. Das Erfassen dieser Metriken, Festlegen einer Baseline und Verwenden von Metriken zum Ermitteln der idealen Datenverarbeitungsoption stellt einen niedrigen bis mittleren Grad des Aufwands dar. Die Validierung erfolgt am besten über Lasttests und Experimentieren.

Ressourcen

Ähnliche Dokumente:

- Cloud Computing mit AWS
- AWS Compute Optimizer
- EC2-Instance-Typen
- Steuerung des Prozessorzustands für Ihre EC2-Instance
- EKS-Container: EKS-Worker-Knoten
- Amazon ECS-Container: Amazon ECS-Container-Instances
- Funktionen: Lambda-Funktionskonfiguration

Ähnliche Videos:

- Amazon EC2-Grundlagen (CMP211-R2)
- Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)

Ähnliche Beispiele:

- Rightsizing with Compute Optimizer and Memory utilization enabled (Die richtige Dimensionierung ermitteln, wenn Amazon Compute Optimizer und die Arbeitsspeicherauslastung aktiviert sind)

- AWS Compute Optimizer-Demo-Code

PERF02-BP03 Erfassen von Datenverarbeitungsmetriken

Sie müssen die tatsächliche Nutzung der verschiedenen Ressourcen erfassen und verfolgen, um die Leistung Ihrer Datenverarbeitungsressourcen zu bestimmen. Anhand dieser Daten lassen sich die Ressourcenanforderungen genauer bestimmen.

Workloads können große Mengen an Daten generieren, wie Metriken, Protokolle und Ereignisse. Stellen Sie fest, ob Ihr vorhandener Speicher, Überwachungs- und Beobachtungsservice die generierten Daten verwalten kann. Identifizieren Sie, welche Metriken die Ressourcennutzung widerspiegeln und auf einer einzelnen Plattform erfasst, aggregiert und korreliert werden können. Diese Metriken sollten alle Ihre Workload-Ressourcen, Anwendungen und Services darstellen, sodass Sie einen systemweiten Überblick erhalten und schnell Möglichkeiten zur Leistungsverbesserung und Schwierigkeiten identifizieren können.

Gewünschtes Ergebnis: Alle Metriken in Bezug auf Datenverarbeitungsressourcen werden auf einer einzigen Plattform identifiziert, aggregiert sowie korreliert und die Datenaufbewahrung ist implementiert, um Kosten- und Betriebsziele zu unterstützen.

Gängige Antimuster:

- Sie suchen ausschließlich manuell mithilfe von Protokolldateien nach Metriken.
- Sie veröffentlichen Metriken nur in internen Tools.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware Ihrer Wahl aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.

Vorteile der Einführung dieser bewährten Methode: Um die Leistung der Workloads zu überwachen, müssen Sie mehrere Leistungsmetriken über einen bestimmten Zeitraum aufzeichnen. Mithilfe dieser Metriken können Sie Anomalien bei der Leistung erkennen. Sie helfen auch beim Abgleichen der Leistung mit den Geschäftsmetriken, um sicherzustellen, dass Sie Ihre Workload-Anforderungen erfüllen,

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Identifizieren, sammeln, aggregieren und korrelieren Sie Datenverarbeitungsmetriken. Wenn ein Service wie Amazon CloudWatch verwendet wird, kann die Implementierung schneller erfolgen und ist einfacher zu verwalten. Identifizieren und verfolgen Sie zusätzlich zu den aufgezeichneten Standardmetriken auch weitere Metriken auf Systemebene innerhalb Ihrer Workload. Erfassen Sie Daten zu CPU-Nutzung, Arbeitsspeicher, Datenträger-I/O sowie eingehende und ausgehende Netzwerkmetriken, um Einblick in die Nutzung bzw. in Engpässe zu erhalten. Diese Daten sind von entscheidender Bedeutung, um festzustellen, wie leistungsfähig die Workload ist und wie die Datenverarbeitungslösung genutzt wird. Nutzen Sie diese Kennzahlen im Rahmen eines datengestützten Ansatzes, der Ihnen die aktive Feinabstimmung und Optimierung der vom Workload genutzten Ressourcen ermöglicht.

Implementierungsschritte:

1. Welche Metriken zu Datenverarbeitungslösungen sollten nachverfolgt werden?
    a. EC2-Standardmetriken
    b. Amazon ECS-Standardmetriken
    c. EKS-Standardmetriken
    d. Lambda-Standardmetriken
    e. EC2-Arbeitsspeicher- und -Datenträgermetriken
2. Habe ich derzeit eine genehmigte Protokollierungs- und Überwachungslösung?
    a. Amazon CloudWatch
    b. AWS Distro for OpenTelemetry
    c. Amazon Managed Service for Prometheus
3. Habe ich meine Datenaufbewahrungsrichtlinien identifiziert und konfiguriert, sodass sie meinen Sicherheits- und Betriebszielen entsprechen?
    a. Standard-Datenaufbewahrung für CloudWatch-Metriken
    b. Standard-Datenaufbewahrung für CloudWatch Logs
4. Wie stellen Sie Ihre Metrik- und Protokollaggregationsagenten bereit?
    a. Automatisierung von AWS Systems Manager
    b. OpenTelemetry Collector

Grad des Aufwands für den Implementierungsplan Der Grad des Aufwands ist mittel, um Metriken von allen Datenverarbeitungsressourcen zu identifizieren, nachzuverfolgen, zu erfassen, zu aggregieren und zu korrelieren.

Ressourcen

Ähnliche Dokumente:

- Amazon CloudWatch-Dokumentation
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und On-Premises-Servern mit dem CloudWatch Agent
- Zugriff auf Amazon CloudWatch Logs für AWS Lambda
- CloudWatch Logs mit Container-Instances verwenden
- Veröffentlichen von benutzerdefinierten Metriken
- AWS Answers: Zentralisierte Protokollierung
- CloudWatch-Services, die AWS-Metriken veröffentlichen
- Amazon EKS auf AWS Fargate überwachen

Ähnliche Videos:

- Verwaltung der Anwendungsleistung in AWS
- Erstellen eines Überwachungsplans

Ähnliche Beispiele:

- Level 100: Monitoring with CloudWatch Dashboards (Stufe 100: Überwachung mit Cloudwatch-Dashboards)
- Level 100: Monitoring Windows EC2 instance with CloudWatch Dashboards (Stufe 100: Überwachung einer Windows-EC2-Instance mit Cloudwatch-Dashboards)
- Level 100: Monitoring an Amazon Linux EC2 instance with CloudWatch Dashboards (Stufe 100: Überwachung einer Amazon-Linux-EC2-Instance mit Cloudwatch-Dashboards)

PERF02-BP04 Bestimmen der erforderlichen Konfiguration durch Dimensionieren

Analysieren Sie die verschiedenen Leistungsmerkmale Ihrer Workload und bewerten Sie, wie sich diese auf Arbeitsspeicher, Netzwerk und CPU-Auslastung auswirken. Wählen Sie anhand dieser Daten die für das Workload-Profil am besten geeigneten Ressourcen aus. Beispielsweise könnte eine arbeitsspeicherintensive Workload wie z. B. eine Datenbank am besten von der r-Familie der Instances bedient werden. Eine Bursting-Workload kann jedoch mehr von einem elastischen Containersystem profitieren.

Gängige Antimuster:

- Sie wählen die größte verfügbare Instance für alle Workloads aus.
- Zur einfacheren Verwaltung verwenden Sie für alle Instances einen Typ als Standard.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie mit den Datenverarbeitungsangeboten von AWS vertraut sind, können Sie die richtige Lösung für Ihre verschiedenen Workloads bestimmen. Nachdem Sie die verschiedenen Datenverarbeitungsangebote für die Workload ausgewählt haben, können Sie anhand von schnellen Experimenten mit diesen Angeboten flexibel feststellen, welche davon Ihren Workload-Anforderungen entsprechen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Ändern der Workload-Konfiguration durch richtige Dimensionierung: Bestimmen Sie zur Optimierung von Leistung und Gesamteffizienz, welche Ressourcen Ihre Workload benötigt. Wählen Sie arbeitsspeicheroptimierte Instances für Systeme, die mehr Arbeitsspeicher als CPU benötigen. Verwenden Sie hingegen rechenoptimierte Instances für Komponenten, die eine nicht arbeitsspeicherintensive Datenverarbeitung durchführen. Bei korrekter Dimensionierung bietet die Workload eine optimale Leistung, wobei nur die benötigten Ressourcen verwendet werden.

Ressourcen

Zugehörige Dokumente:

- AWS Compute Optimizer
- Cloud Computing mit AWS
- EC2-Instance-Typen

- ECS-Container: Amazon ECS-Container-Instances
- EKS-Container: EKS-Worker-Knoten
- Funktionen: Lambda-Funktionskonfiguration
- Steuerung des Prozessorzustands für Ihre EC2-Instance

Relevante Videos:

- Amazon EC2-Grundlagen (CMP211-R2)
- Bessere, schnellere und kostengünstigere Datenverarbeitung: Kostenoptimierung bei Amazon EC2 (CMP202-R1)
- Bereitstellen leistungsstarker ML-Inferenzen mit AWS Inferentia (CMP324-R1)
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)
- Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems
- Datenverarbeitungsoptionen auswählen
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)

Zugehörige Beispiele:

- Rightsizing with Compute Optimizer and Memory utilization enabled (Die richtige Dimensionierung ermitteln, wenn Amazon Compute Optimizer und die Arbeitsspeicherauslastung aktiviert sind)
- AWS Compute Optimizer-Demo-Code

PERF02-BP05 Nutzen verfügbarer Elastizität von Ressourcen

Die Cloud bietet Ihnen die Flexibilität, Ressourcen dynamisch durch verschiedene Mechanismen zu erweitern oder zu reduzieren, um einem veränderten Bedarf gerecht zu werden. In Kombination mit Rechenmetriken kann eine Workload automatisch auf Änderungen reagieren und die optimalen Ressourcen nutzen, um die Zielvorgabe zu erreichen.

Die optimale Anpassung des Angebots an die Nachfrage ist die kostengünstigste Variante für einen Workload. Wichtig ist jedoch, ein ausreichendes Angebot einzuplanen, um die Bereitstellungszeit und individuelle Ressourcenfehler abzudecken. Die Nachfrage kann fest oder variabel sein und Metriken sowie eine Automatisierung erfordern, um sicherzustellen, dass durch die Verwaltung keine unverhältnismäßig hohen Kosten entstehen.

In AWS können Sie eine Vielzahl verschiedener Ansätze für die Abstimmung von Angebot und Bedarf verwenden. Im Whitepaper zur Säule „Kostenoptimierung“ wird beschrieben, wie Sie die folgenden Kostenansätze anwenden:

- Bedarfsbasierter Ansatz
- Pufferbasierter Ansatz
- Zeitabhängiger Ansatz

Sie müssen sicherstellen, dass Workload-Bereitstellungen sowohl Hoch- als auch Herunterskalierungsereignisse verarbeiten können. Erstellen Sie Testszenarien für Herunterskalierungen, damit sich die Workload wie erwartet verhält.

Gängige Antimuster:

- Sie reagieren auf Alarme, indem Sie die Kapazität manuell erhöhen.
- Sie belassen die erhöhte Kapazität nach dem Hochskalieren, anstatt wieder herunterzuskalieren.

Vorteile der Einführung dieser bewährten Methode: Durch das Konfigurieren und Testen der Workload-Elastizität können Sie Kosten verringern, Leistungs-Benchmarks erhalten und die Zuverlässigkeit bei sich änderndem Datenverkehr verbessern. Die meisten Instances außerhalb der Produktionsumgebung sollten bei Nichtgebrauch angehalten werden. Obwohl ungenutzte Instances manuell heruntergefahren werden können, ist dies bei einer größeren Anzahl von Instances unpraktisch. Sie können zudem die volumenbasierte Elastizität nutzen. Diese ermöglicht die Optimierung der Leistung und Kosten, indem Sie bei Bedarfsspitzen die Anzahl der Datenverarbeitungs-Instances erhöhen und bei sinkendem Bedarf die Kapazität verringern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Elastizität nutzen: Elastizität ermöglicht das Anpassen der verfügbaren Ressourcen an den Bedarf. Instances, Container und Funktionen bieten Mechanismen für Elastizität, sei es in Kombination mit automatischer Skalierung oder als Merkmal des Service. Stellen Sie mithilfe von Elastizität in Ihrer Architektur sicher, dass Sie über genügend Kapazität zur Erfüllung der Leistungsanforderungen für den jeweiligen Nutzungsumfang verfügen. Sorgen Sie dafür, dass die Metriken zum Hoch- oder Herunterskalieren elastischer Ressourcen für die jeweilige Art der bereitgestellten Workload überprüft werden. Wenn Sie eine Anwendung zur Video-Transcodierung bereitstellen, wird eine CPU-Auslastung von 100 % erwartet, weshalb dies nicht die Hauptmetrik sein sollte. Alternativ können

Sie die Warteschlangenlänge von Transcodierungsaufgaben messen, die auf die Skalierung der Instance-Typen warten. Stellen Sie sicher, dass Workload-Bereitstellungen sowohl mit Hoch- als auch mit Herunterskalierungen umgehen können. Das sichere Herunterskalieren von Workload-Komponenten ist genauso wichtig wie das Hochskalieren von Ressourcen bei entsprechendem Bedarf. Erstellen Sie Testszenarien für Herunterskalierungen, damit sich die Workload wie erwartet verhält.

Ressourcen

Zugehörige Dokumente:

- Cloud Computing mit AWS
- EC2-Instance-Typen
- ECS-Container: Amazon ECS-Container-Instances
- EKS-Container: EKS-Worker-Knoten
- Funktionen: Lambda-Funktionskonfiguration
- Steuerung des Prozessorzustands für Ihre EC2-Instance

Relevante Videos:

- Amazon EC2-Grundlagen (CMP211-R2)
- Bessere, schnellere und kostengünstigere Datenverarbeitung: Kostenoptimierung bei Amazon EC2 (CMP202-R1)
- Bereitstellen leistungsstarker ML-Inferenzen mit AWS Inferentia (CMP324-R1)
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)
- Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems

Zugehörige Beispiele:

- Amazon EC2 Auto Scaling-Gruppenbeispiele
- Amazon EFS-Tutorials

PERF02-BP06 Neue Bewertung von Datenverarbeitungsbedarf anhand von Metriken

Identifizieren Sie anhand von Kennzahlen auf Systemebene das Verhalten und die Anforderungen Ihres Workloads in einem bestimmten Zeitraum. Bewerten Sie die Anforderungen Ihrer Workload,

indem Sie die verfügbaren Ressourcen mit diesen Anforderungen vergleichen. Passen Sie die Datenverarbeitungsumgebung so an, dass sie dem Profil der Workload optimal entspricht. Beispiel: Im Laufe der Zeit stellen Sie möglicherweise fest, dass ein System mehr Arbeitsspeicher benötigt, als anfangs gedacht. Ein Wechsel zu einer anderen Instance-Familie oder -Größe kann die Leistung und Effizienz verbessern.

Gängige Antimuster:

- Sie überwachen nur Metriken auf Systemebene, um Einblicke in Ihre Workload zu gewinnen.
- Sie legen Ihre Rechenbedürfnisse auf Workload-Anforderungen zu Spitzenzeiten aus.
- Ihre Datenverarbeitungslösung ist überdimensioniert, um Ihre Skalierungs- oder Leistungsanforderungen zu erfüllen, wenn der Umstieg zu einer neuen Datenverarbeitungslösung Ihren Workload-Merkmalen entsprechen würde.

Vorteile der Einführung dieser bewährten Methode: Um Leistung und Ressourcenauslastung zu optimieren, benötigen Sie einen Gesamtüberblick über den Betrieb, detaillierte Echtzeitdaten und Referenzdaten aus der Vergangenheit. Sie können automatische Dashboards erstellen, um diese Daten zu visualisieren und Metrikberechnungen durchzuführen. So erhalten Sie Einblicke in Betrieb und Auslastung.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Optimieren von Ressourcen mithilfe eines datengestützten Ansatzes: Eine maximale Leistung und Effizienz erzielen Sie, indem Sie anhand der Daten, die Sie im Laufe der Zeit für Ihre Workload erfasst haben, die Ressourcen genau abstimmen und optimieren. Analysieren Sie, wie Ihr Workload die aktuell verfügbaren Ressourcen nutzt und überlegen Sie, welche Änderungen Sie vornehmen könnten, um die Anforderungen Ihres Workloads besser zu erfüllen. Wenn zu viele Ressourcen genutzt werden, verschlechtert sich die Systemleistung, während eine zu geringe Auslastung zu einer ineffizienten Ressourcennutzung und zu höheren Kosten führt.

Ressourcen

Zugehörige Dokumente:

- Cloud Computing mit AWS
- AWS Compute Optimizer
- Cloud Computing mit AWS

- EC2-Instance-Typen
- ECS-Container: Amazon ECS-Container-Instances
- EKS-Container: EKS-Worker-Knoten
- Funktionen: Lambda-Funktionskonfiguration
- Steuerung des Prozessorzustands für Ihre EC2-Instance

Relevante Videos:

- Amazon EC2-Grundlagen (CMP211-R2)
- Bessere, schnellere und kostengünstigere Datenverarbeitung: Kostenoptimierung bei Amazon EC2 (CMP202-R1)
- Bereitstellen leistungsstarker ML-Inferenzen mit AWS Inferentia (CMP324-R1)
- Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS (CMP323-R1)
- Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro-Systems

Zugehörige Beispiele:

- Rightsizing with Compute Optimizer and Memory utilization enabled (Die richtige Dimensionierung ermitteln, wenn Amazon Compute Optimizer und die Arbeitsspeicherauslastung aktiviert sind)
- AWS Compute Optimizer-Demo-Code

LEIST 3 Was ist bei der Wahl der Speicherlösung zu beachten?

Die optimale Speicherlösung für ein System richtet sich nach der Zugriffsmethode (Block, Datei oder Objekt), den Zugriffsmustern (Zufallsprinzip oder sequenziell), dem erforderlichen Durchsatz, der Zugriffshäufigkeit (online, offline, Archiv), der Aktualisierungshäufigkeit (WORM, dynamisch) sowie den Einschränkungen hinsichtlich Verfügbarkeit und Langlebigkeit. Gut geplante Systeme nutzen mehrere Speicherlösungen und bieten unterschiedliche Möglichkeiten zur Leistungsoptimierung und effizienten Ressourcennutzung.

Bewährte Methoden

- PERF03-BP01 Verstehen von Speichereigenschaften und -anforderungen
- PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen
- PERF03-BP03 Einbeziehen von Zugriffsmustern und Metriken in die Entscheidung

PERF03-BP01 Verstehen von Speichereigenschaften und -anforderungen

Ermitteln und dokumentieren Sie den Speicherbedarf der Workloads und definieren Sie die Speichereigenschaften der einzelnen Standorte. Beispiele für Speichereigenschaften sind: gemeinsamer Zugriff, Dateigröße, Wachstumsrate, Durchsatz, IOPS, Latenz, Zugriffsmuster und Datenpersistenz. Beurteilen Sie anhand dieser Eigenschaften, ob Block-, Datei, Objekt- oder Instance-Speicherservices die effizienteste Lösung für Ihren Speicherbedarf darstellen.

Gewünschtes Ergebnis: Ermitteln und dokumentieren Sie den Speicherbedarf pro Speicheranforderung und bewerten Sie die verfügbaren Speicherlösungen. Unter Berücksichtigung der wichtigsten Speichereigenschaften wird Ihr Team verstehen, wie die ausgewählten Speicherservices für eine Verbesserung der Workload-Leistung sorgen werden. Zu den wesentlichen Kriterien gehören, Datenzugriffsmuster, Wachstumsrate, Skalierungsbedarf und Latenzanforderungen.

Typische Anti-Muster:

- Sie verwenden nur einen Speichertyp, z. B. Amazon Elastic Block Store (Amazon EBS), für alle Workloads.
- Sie gehen davon aus, dass für alle Workloads ähnliche Anforderungen an die Speicherzugriffsleistung gelten.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie die Speicherlösung auf der Grundlage der ermittelten und erforderlichen Eigenschaften auswählen, können Sie damit die Leistung Ihrer Workloads verbessern, die Kosten senken und den betrieblichen Aufwand für die Verwaltung Ihrer Workloads verringern. Die Workload-Leistung wird von der Lösung, der Konfiguration und dem Standort des Speicherservice profitieren.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Identifizieren Sie die wichtigsten Speicherleistungsmetriken Ihrer Workload und implementieren Sie Verbesserungen als Teil eines datengesteuerten Ansatzes mithilfe von Benchmarking oder Lasttests. Ermitteln Sie anhand dieser Daten, an welcher Stelle Ihre Speicherlösung Defizite hat. Prüfen Sie anschließend die Konfigurationsoptionen zur Verbesserung der Lösung. Ermitteln Sie die erwartete Wachstumsrate für Ihre Workload und wählen Sie eine Speicherlösung aus, die diesen Raten gerecht wird. Überprüfen Sie die AWS-Speicherangebote, um die richtige Speicherlösung für Ihre

verschiedenen Workload-Anforderungen zu ermitteln. Durch die Bereitstellung von Speicherlösungen in AWS haben Sie bessere Möglichkeiten, Speicherangebote zu testen und festzustellen, ob sie für Ihre Workload-Anforderungen geeignet sind.

AWS-Service	Schlüsselmerkmale	Häufige Anwendungsfälle
Amazon S3	Beständigkeit von 99,999999 999 %, unbegrenztes Wachstum, Zugriff von überall, mehrere Kostenmodelle auf der Grundlage von Zugriff und Ausfallsicherheit	Cloudnative Anwendung sdaten, Datenarchivierung und Backups, Analysen, Data Lakes, Hosting von statischen Websites, IoT-Daten
Amazon S3 Glacier	Latenz von Sekunden bis Stunden, unbegrenztes Wachstum, geringste Kosten, langfristige Speicherung	Datenarchivierung, Medienarc hive, langfristige Aufbewahr ung von Backups
Amazon EBS	Die Größe des Speichers erfordert Verwaltung und Überwachung, niedrige Latenz, dauerhafte Speicheru ng, Beständigkeit von 99,8 % bis 99,9 %, die meisten Volume-Typen sind nur von einer EC2-Instance aus zugänglich.	COTS-Anwendungen, I/O-intensive Anwendungen, relationale Datenbanken und NoSQL-Datenbanken, Sicherung und Wiederher stellung
EC2-Instance-Speicher	Vorab festgelegte Speicherg röße. geringste Latenz, nicht persistent, nur von einer EC2-Instance aus zugänglich	COTS-Anwendungen, I/O-intensive Anwendungen, In-Memory-Datenspeicher
Amazon EFS	Beständigkeit von 99,999999 999 %, unbegrenztes Wachstum, Zugriff von mehreren Datenverarbeitungs services aus möglich	Modernisierte Anwendung en, die Dateien über mehrere Datenverarbeitungsservices hinweg gemeinsam nutzen, Dateispeicher für die Skalierun

AWS-Service	Schlüsselmerkmale	Häufige Anwendungsfälle
		g von Content-Management-Systemen
Amazon FSx	Unterstützt vier Dateisysteme (NetApp, OpenZFS, Windows File Server und Amazon FSx for Lustre), verfügbarer Speicher für jedes Dateisyst em unterschiedlich, Zugriff von mehreren Datenverarbeitungs services aus möglich	Cloudnative Workloads , Private Cloud Bursting, migrierte Workloads, die ein bestimmtes Dateisystem erfordern, VMC, ERP-Syste me, On-Premises-Dateis peicherung und -Backups
Snow Family	Tragbare Geräte, 256-Bit-V erschlüsselung, NFS-Endpu nkt, On-Board-Computing, TB Speicherplatz	Migration von Daten in die Cloud, Speicherung und Datenverarbeitung unter extremen On-Premises-Bedingungen, Notfallwi ederherstellung, Remote-Da tenerfassung
AWS Storage Gateway	Bietet On-Premises-Zugriff mit niedriger Latenz auf Cloud-gestützten Speicher, vollständ ig verwalteter On-Premises-Cache	Migrationen von On-Premises-Daten in die Cloud, Auffüllen von Cloud Data Lakes aus On-Premises-Quellen, modernisi erte Dateifreigabe

Implementierungsschritte:

1. Nutzen Sie Benchmarking oder Ladetests, um die wichtigsten Merkmale Ihres Speicherbedarfs zu erfassen. Schlüsselmerkmale sind:
    a. Gemeinsam nutzbar (welche Komponenten greifen auf diesen Speicher zu)
    b. Wachstumsrate
    c. Durchsatz
    d. Latenz

e. I/O-Größe

f. Stabilität

g. Zugriffsmuster (Lese- oder Schreibzugriff, Häufigkeit, schwankend oder konsistent)

2. Ermitteln Sie die für Ihre Speichereigenschaften geeignete Art von Speicherlösung.

a. Amazon S3 ist ein Objektspeicherservice mit unbegrenzter Skalierbarkeit, hoher Verfügbarkeit und mehreren Zugriffsoptionen. Für die Übertragung von Objekten in und aus Amazon S3 und den Zugriff auf diese Objekte können Sie einen Service wie z. B. Transfer Acceleration oder Zugriffspunkte nutzen, um Ihren Standort, Ihre Sicherheitsanforderungen und Zugriffsmuster zu unterstützen. Verwenden Sie die Amazon S3-Leistungsrichtlinien, um Ihre Amazon S3-Konfiguration zu optimieren und damit den Anforderungen an Ihre Workload-Leistung gerecht zu werden.

b. Amazon S3 Glacier ist eine Speicherklasse vonAmazon S3 für die Datenarchivierung. Sie haben drei Archivierungslösungen zur Auswahl, die von einem Millisekundenzugriff bis zu einem Zugriff von 5 bis 12 Stunden bei unterschiedlichen Kosten und Sicherheitsoptionen reichen. Amazon S3 Glacier kann Ihnen helfen, die Leistungsanforderungen zu erfüllen, indem ein Datenlebenszyklus implementiert wird, der Ihre geschäftlichen Anforderungen und Dateneigenschaften unterstützt.

c. Amazon Elastic Block Store (Amazon EBS) ist ein hochleistungsfähiger Blockspeicherservice für Amazon Elastic Compute Cloud (Amazon EC2). Sie können unter SSD- oder HDD-basierten Lösungen mit unterschiedlichen Merkmalen auswählen, die IOPS oder Durchsatz priorisieren. EBS-Volumes sind gut geeignet für Hochleistungs-Workloads, primären Speicher für Dateisysteme, Datenbanken oder Anwendungen, die nur auf angehängte Stage-Systeme zugreifen können.

d. Amazon EC2-Instance-Speicher ist ähnlich wie Amazon EBS, da er an eine Amazon EC2-Instance angehängt wird. Allerdings ist Instance-Speicher nur ein temporärer Speicher, der idealerweise als Puffer, Cache oder für andere temporäre Inhalte verwendet werden sollte. Ein Instance-Speicher kann nicht getrennt werden. Wenn die Instance heruntergefahren wird, gehen alle Daten verloren. Instance-Speicher kann für Anwendungsfälle mit hoher I/O-Leistung und niedriger Latenz verwendet werden, bei denen die Daten nicht bestehen bleiben müssen.

e. Amazon Elastic File System (Amazon EFS) ist ein mountfähiges Dateisystem, auf das verschiedene Arten von Datenverarbeitungslösungen zugreifen können. Amazon EFS erweitert und verringert den Speicher automatisch und ist leistungsoptimiert, um durchgängig niedrige Latenzen zu bieten. EFS verfügt über zwei Leistungskonfigurationsmodi:: Allzweck und max. I/O. Der Allzweckmodus weist eine Leselatenz von weniger als einer Millisekunde und eine Schreiblatenz im einstelligen Millisekundenbereich auf. Die „Max. I/O“-Funktion kann Tausende

von Computing-Instances unterstützen, die ein gemeinsames Dateisystem benötigen. Amazon EFS unterstützt zwei Durchsatzmodi: Bursting und Bereitgestellt. Für eine Workload mit schwankendem Zugriffsmuster wird der Bursting-Durchsatzmodus vorteilhaft sein, während eine konstant hohe Workload bei Nutzung des bereitgestellten Durchsatzmodus eine gute Leistung zeigen wird.

f. Amazon FSx basiert auf den neuesten AWS-Datenverarbeitungslösungen und unterstützt vier gängige Dateisysteme: NetApp ONTAP, OpenZFS, Windows File Server und Lustre. Die Latenz, der Durchsatz und die IOPS von Amazon FSx variieren je nach Dateisystem und sollten bei der Auswahl des richtigen Dateisystems für Ihre Workload-Anforderungen berücksichtigt werden.

g. AWS Snow Family sind Speicher- und Datenverarbeitungsgeräte, die eine Online- und Offline-Datenmigration in die Cloud sowie die Datenspeicherung und -verarbeitung On-Premises unterstützen. AWS-Snow-Geräte unterstützen die Erfassung großer Mengen an On-Premises-Daten, die Verarbeitung dieser Daten und die Verschiebung der Daten in die Cloud. Es sind mehrere bewährte Methoden zur Leistungsoptimierung in Bezug auf die Anzahl der Dateien, die Dateigrößen und die Komprimierung dokumentiert.

h. AWS Storage Gateway bietet On-Premises-Anwendungen Zugriff auf cloudbasierten Speicher. AWS Storage Gateway unterstützt mehrere Cloud-Speicherservices, darunter Amazon S3, Amazon S3 Glacier, Amazon FSx und Amazon EBS. Der Service unterstützt verschiedene Protokolle wie z. B. iSCSI, SMB und NFS. Er bietet Leistung mit niedriger Latenz, da häufig abgerufene Daten On-Premises zwischengespeichert werden und nur geänderte und komprimierte Daten an AWS gesendet werden.

3. Nachdem Sie mit Ihrer neuen Speicherlösung experimentiert und die optimale Konfiguration ermittelt haben, planen Sie Ihre Migration und überprüfen Sie Ihre Leistungsmetriken. Dies ist ein kontinuierlicher Prozess, der neu bewertet werden sollte, wenn sich wichtige Merkmale ändern oder es Änderungen in Bezug auf die verfügbaren Services oder Optionen gibt.

Aufwand für den Implementierungsplan: Wenn eine Workload von einer Speicherlösung zu einer anderen verschoben wird, könnte der Faktorwechsel der Anwendung mit einem moderaten Aufwand verbunden sein.

Ressourcen

Zugehörige Dokumente:

- Amazon EBS Volume-Typen

- Amazon EC2 Speicher
- Amazon EFS: Leistung von Amazon EFS
- Leistung von Amazon FSx for Lustre
- Leistung von Amazon FSx for Windows File Server
- Leistung von Amazon FSx for NetApp ONTAP
- Leistung von Amazon FSx for OpenZFS
- Amazon S3 Glacier: Dokumentation zu Amazon S3 Glacier
- Amazon S3: Überlegungen zu Anfragerate und Leistung
- Cloud-Speicher mit AWS
- AWS Snow Family
- EBS-I/O-Merkmale

Zugehörige Videos:

- Ausführliche Beschreibung von Amazon EBS (STG303-R1)
- Optimieren Sie Ihre Speicherleistung mit Amazon S3 (STG343)

Zugehörige Beispiele:

- Amazon EFS-CSI-Treiber
- Amazon EBS-CSI-Treiber
- Amazon EFS-Dienstprogramme
- Amazon EBS – automatische Skalierung
- Amazon S3-Beispiele
- Amazon FSx for Lustre Container Storage Interface (CSI)-Treiber

PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen

Bewerten Sie die verschiedenen Merkmale und Konfigurationsoptionen und ermitteln Sie, welche Auswirkungen sie auf den Speicher haben. Finden Sie heraus, wo und wie Sie bereitgestellte IOPS, SSDs, magnetischen Speicher, Objektspeicher, Archivspeicher oder flüchtigen Speicher idealerweise einsetzen, um Speicherplatz und Leistung Ihrer Workload zu optimieren.

Amazon EBS bietet eine Reihe von Optionen, mit denen Sie die Speicherleistung optimieren und die Workload-Kosten senken können. Dabei gibt es zwei Hauptkategorien: SSD-gestützten Speicher für Transaktions-Workloads wie etwa Datenbanken und Boot-Volumes (Leistung hängt primär von IOPS ab), sowie HDD-gestützten Speicher für durchsatzintensive Workloads wie MapReduce und die Protokollverarbeitung (Leistung hängt primär von MB/s ab).

Zu den SSD-gestützten Volumes zählen extrem leistungsstarke SSDs mit bereitgestellten IOPS für Transaktions-Workloads, bei denen eine geringe Latenz wichtig ist, sowie allgemeine SSDs mit einem guten Preis-Leistungs-Verhältnis, die sich für eine Vielzahl von Transaktionsdaten eignen.

Amazon S3 Transfer Acceleration ermöglicht die schnelle Datenübertragung zwischen Ihrem Client und Ihrem S3-Bucket über große Entfernungen. Transfer Acceleration nutzt global verteilte Amazon CloudFront-Edge-Standorte, um den Netzwerkpfad für die Datenweiterleitung zu optimieren. Für eine Workload in einem S3-Bucket mit umfassenden GET-Anfragen empfiehlt sich die Verwendung von Amazon S3 mit CloudFront. Wenn Sie große Dateien hochladen, sind mehrteilige Uploads von Vorteil. Durch das Hochladen mehrerer Teile können Sie den Netzwerkdurchsatz maximieren.

Amazon Elastic File System (Amazon EFS) bietet ein einfaches, skalierbares, vollständig verwaltetes elastisches NFS-Dateisystem für die Verwendung mit AWS Cloud-Services und On-Premises-Ressourcen. Zur Unterstützung einer Vielzahl von Cloud-Speicher-Workloads bietet Amazon EFS zwei Leistungsmodi: den Allzweck-Leistungsmodus und den Max. E/A-Leistungsmodus. Es stehen zwei Durchsatzmodi für Ihr Dateisystem zur Auswahl: Bursting und Bereitgestellt. Informationen dazu, welche Einstellungen für Ihren Workload verwendet werden sollten, finden Sie im Amazon EFS-Benutzerhandbuch.

Amazon FSx bietet vier Dateisysteme zur Auswahl: Amazon FSx for Windows File Server für Enterprise-Workloads, Amazon FSx for Lustre für Hochleistungs-Workloads, Amazon FSx for NetApp ONTAP für das Dateisystem NetApp ONTAP und Amazon FSx for OpenZFS für Linux-basierte Dateiserver. FSx ist SSD-gestützt und bietet eine schnelle, vorhersehbare, skalierbare und konsistente Leistung. Amazon FSx-Dateisysteme bieten dauerhaft hohe Lese- und Schreibgeschwindigkeiten und konsistenten Datenzugriff mit geringer Latenz. Sie können das Durchsatzniveau auswählen, den Sie benötigen, um den Anforderungen Ihrer Workload zu entsprechen.

Gängige Antimuster:

- Sie verwenden nur einen Speichertyp, z. B. Amazon EBS, für alle Workloads.
- Sie verwenden bereitgestellte IOPS für alle Workloads, ohne reale Tests auf allen Speicherebenen durchzuführen.

- Sie gehen davon aus, dass für alle Workloads ähnliche Anforderungen an die Speicherzugriffsleistung gelten.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie alle Speicherservice-Optionen auswerten, können Sie die Kosten für die Infrastruktur und den Aufwand reduzieren, der zur Aufrechterhaltung Ihrer Workloads erforderlich ist. Dies kann Ihre Markteinführungszeit für die Bereitstellung neuer Services und Funktionen beschleunigen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Bestimmen der Speichermerkmale: Überlegen Sie bei der Wahl einer Speicherlösung, welche Speichereigenschaften für Sie wichtig sind, wie etwa Freigabefähigkeit, Datei- und Cache-Größe, Latenz, Durchsatz und Datenpersistenz. Prüfen Sie anschließend, welcher AWS-Service Ihre Anforderungen am besten erfüllt.

Ressourcen

Zugehörige Dokumente:

- Cloud-Speicher mit AWS
- Amazon EBS Volume-Typen
- Amazon EC2 Speicher
- Amazon EFS: Leistung von Amazon EFS
- Leistung von Amazon FSx for Lustre
- Leistung von Amazon FSx for Windows File Server
- Amazon Glacier: Dokumentation zu Amazon Glacier
- Amazon S3: Überlegungen zu Anfragerate und Leistung
- Cloud-Speicher mit AWS
- Cloud-Speicher mit AWS
- EBS-E/A-Merkmale

Relevante Videos:

- Ausführliche Beschreibung von Amazon EBS (STG303-R1)
- Optimieren Sie Ihre Speicherleistung mit Amazon S3 (STG343)

Zugehörige Beispiele:

- Amazon EFS-CSI-Treiber
- Amazon EBS-CSI-Treiber
- Amazon EFS-Dienstprogramme
- Amazon EBS automatische Skalierung
- Amazon S3-Beispiele

PERF03-BP03 Einbeziehen von Zugriffsmustern und Metriken in die Entscheidung

Wählen Sie Speichersysteme basierend auf den Zugriffsmustern Ihrer Workload aus und konfigurieren Sie sie, indem Sie festlegen, wie die Workload auf Daten zugreift. Erhöhen Sie die Speichereffizienz, indem Sie Objektspeicher statt Blockspeicher auswählen. Konfigurieren Sie die von Ihnen gewählten Speicheroptionen so, dass sie den Datenzugriffsmustern entsprechen.

Die Leistung der Speicherlösung hängt davon ab, wie Sie auf Daten zugreifen. Wählen Sie für maximale Leistung die für Ihre Zugriffsmuster geeignete Speicherlösung, oder passen Sie Ihre Zugriffsmuster an die Speicherlösung an.

Indem Sie ein RAID 0-Array erstellen, können Sie die Leistung eines Dateisystems gegenüber der Bereitstellung eines einzelnen Volumes erhöhen. RAID 0 empfiehlt sich, wenn die E/A-Leistung wichtiger als die Fehlertoleranz ist. Das Array eignet sich beispielsweise für eine intensiv genutzte Datenbank, bei der die Datenreplikation bereits separat eingerichtet ist.

Wählen Sie für Ihren Workload geeignete Speichermetriken für alle Speicheroptionen aus, die für den Workload verwendet werden. Wenn Sie Dateisysteme verwenden, die Burst-Guthaben verwenden, erstellen Sie Alarme, damit Sie informiert werden, wenn Sie sich diesen Guthabenlimits nähern. Sie müssen Speicher-Dashboards erstellen, um den gesamten Workload-Speicherzustand anzuzeigen.

Stellen Sie bei Speichersystemen mit einer festen Größe wie Amazon EBS oder Amazon FSx sicher, dass Sie die Menge des verwendeten Speichers im Vergleich zur Gesamtspeichergröße überwachen und nach Möglichkeit die Speichergröße beim Erreichen eines Schwellenwerts automatisch erhöhen.

Gängige Antimuster:

- Sie gehen davon aus, dass die Speicherleistung ausreichend ist, wenn sich Kunden nicht beschweren.
- Sie verwenden nur eine Speicherebene, vorausgesetzt, dass alle Workloads in diese Ebene passen.

Vorteile der Einführung dieser bewährten Methode: Um Leistung und Ressourcenauslastung zu optimieren, benötigen Sie einen Gesamtüberblick über den Betrieb, detaillierte Echtzeitdaten und Referenzdaten aus der Vergangenheit. Sie können automatische Dashboards und Daten mit einer Granularität von einer Sekunde erstellen, um Metrikberechnungen für Ihre Daten durchzuführen und Einblicke in Betrieb und Auslastung Ihrer Speicheranforderungen zu erhalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Optimieren von Speichernutzung und Zugriffsmustern: Wählen Sie die Speichersysteme je nach Zugriffsmuster der Workload und auf Basis der Merkmale der verfügbaren Speicheroptionen aus. Achten Sie bei der Wahl des Datenspeicherorts darauf, dass Ihre Anforderungen erfüllt und gleichzeitig der Overhead minimiert werden. Ziehen Sie beim Konfigurieren und Interagieren mit Daten, je nach Speichermerkmalen, Leistungsoptimierungen und Zugriffsmuster heran (z. B.Volume Striping oder Datenpartitionierung).

Auswählen geeigneter Metriken für Speicheroptionen: Stellen Sie sicher, dass Sie die entsprechenden Speichermetriken für die Workload auswählen. Jede Speicheroption bietet verschiedene Metriken, um zu verfolgen, wie Ihre Workload im Laufe der Zeit ausgeführt wird. Stellen Sie sicher, dass Sie anhand von Speicherburst-Metriken messen (z. B. Überwachung von Burst-Guthaben für Amazon EFS). Stellen Sie bei Speichersystemen mit fester Größe wie Amazon Elastic Block Store oder Amazon FSx sicher, dass Sie die verwendete Speichermenge im Vergleich zur Gesamtspeichergröße überwachen. Erstellen Sie nach Möglichkeit eine Automatisierung, um die Speichergröße zu erhöhen, wenn Sie einen Schwellenwert erreichen.

Überwachen von Metriken: Mithilfe von Amazon CloudWatch lassen sich Kennzahlen aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können auch benutzerdefinierte Kennzahlen erfassen und in Oberflächen-, Geschäfts- oder abgeleiteten Kennzahlen veröffentlichen. Richten Sie mit CloudWatch oder mit Lösungen von Drittanbietern Alarme ein, die auf das Überschreiten von Schwellenwerten hinweisen.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS Volume-Typen
- Amazon EC2 Speicher
- Amazon EFS: Leistung von Amazon EFS

- Leistung von Amazon FSx for Lustre
- Leistung von Amazon FSx for Windows File Server
- Amazon Glacier: Dokumentation zu Amazon Glacier
- Amazon S3: Überlegungen zu Anfragerate und Leistung
- Cloud-Speicher mit AWS
- EBS-E/A-Merkmale
- Die Leistung von Amazon EBS mithilfe von Amazon CloudWatch überwachen und verstehen

Ähnliche Videos:

- Ausführliche Beschreibung von Amazon EBS (STG303-R1)
- Optimieren Sie Ihre Speicherleistung mit Amazon S3 (STG343)

Ähnliche Beispiele:

- Amazon EFS-CSI-Treiber
- Amazon EBS-CSI-Treiber
- Amazon EFS-Dienstprogramme
- Amazon EBS automatische Skalierung
- Amazon S3-Beispiele

LEIST 4 Was ist bei der Wahl der Datenbanklösung zu beachten?

Welche Datenbanklösung sich am besten für ein System eignet, hängt von der erforderlichen Verfügbarkeit, Konsistenz, Partitionstoleranz, Latenz, Langlebigkeit, Skalierbarkeit und Abfragefähigkeit ab. Viele Systeme nutzen für verschiedene Untersysteme unterschiedliche Datenbanklösungen und unterstützen unterschiedliche Funktionen zur Leistungsoptimierung. Die Wahl der falschen Datenbanklösung und -funktionen kann die Leistungseffizienz eines Systems schmälern.

Bewährte Methoden

- PERF04-BP01 Verstehen von Datenmerkmalen
- PERF04-BP02 Prüfen der verfügbaren Optionen

- PERF04-BP03 Erfassen und Aufzeichnen von Metriken zur Datenbankleistung
- PERF04-BP04 Wählen des Datenspeichers nach Zugriffsmuster
- PERF04-BP05 Optimieren des Datenspeicher nach Zugriffsmuster und Metriken

PERF04-BP01 Verstehen von Datenmerkmalen

Wählen Sie Ihre Datenverwaltungslösungen aus, sodass Sie den Eigenschaften, Zugriffsmustern und Anforderungen Ihrer Workload-Datensätze optimal entsprechen. Beim Auswählen und Implementieren Ihrer Datenverwaltungslösung müssen Sie sicherstellen, dass Ihre Abfrage-, Skalierungs- und Speichermerkmale die Datenanforderungen der Workload unterstützen. Erfahren Sie, wie unterschiedliche Datenbankoptionen Ihren Datenmodellen entsprechen und welche Konfigurationsoptionen am besten für Ihren Anwendungsfall geeignet sind.

AWS bietet zahlreiche speziell entwickelte Datenbankmodule, darunter relationale, Schlüssel-Werte-, Dokument-, In-Memory-, Graph-, Zeitreihen- und Ledger-Datenbanken. Jede Datenverwaltungslösung hat verfügbare Optionen und Konfigurationen, um Ihre Anwendungsfälle und Datenmodelle zu unterstützen. Basierend auf Ihren Datenmerkmalen kann Ihre Workload möglicherweise mehrere unterschiedliche Datenbanklösungen verwenden. Sie können den Umstieg von monolithischen Datenbanken bewerkstelligen, die mit ihrem Einheitsansatz restriktiv sind, indem Sie die beste Datenbanklösung für ein spezifisches Problem auswählen und sich darauf konzentrieren, Daten zu verwalten, um die Bedürfnisse Ihrer Kunden zu erfüllen.

Gewünschtes Ergebnis: Die Datenmerkmale von Workloads sind mit ausreichenden Details dokumentiert, um die Auswahl und Konfiguration von unterstützenden Datenbanklösungen zu ermöglichen und Einblicke in mögliche Alternativen zu bieten.

Gängige Antimuster:

- Möglichkeiten nicht in Betracht ziehen, größere Datensätze, die ähnliche Merkmale aufweisen, in kleinere Datensammlungen aufzuteilen, was dazu führt, dass Chancen verabsäumt werden, um speziell entwickelte Datenbanken zu verwenden, die den Daten- und Wachstumsmerkmalen besser entsprechen.
- Datenzugriffsmuster nicht vorab identifizieren, was später zu kostspieliger und komplexer Nachbearbeitung führt.
- Wachstum einschränken, indem Datenspeicherstrategien verwendet werden, die nicht ausreichend schnell skalieren.
- Einen Datenbanktyp und -anbieter für alle Workloads auswählen.

- An einer Datenbanklösung festhalten, da es interne Erfahrungen und Wissen über eine bestimmte Datenbanklösung gibt.
- Eine Datenbanklösung behalten, weil sie in einer On-Premises-Umgebung gut funktioniert hat.

Vorteile der Einführung dieser bewährten Methode: Sie sollten mit allen AWS-Datenbanklösungen vertraut sein, damit Sie die richtige Datenbanklösung für Ihre verschiedenen Workloads bestimmen können. Nachdem Sie die geeignete Datenbanklösung für Ihren Workload ausgewählt haben, können Sie schnell mit diesen Datenbankangeboten experimentieren, um festzustellen, ob sie Ihren Workload-Anforderungen weiterhin entsprechen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

- Potenzielle Kosteneinsparungen werden möglicherweise nicht erkannt.
- Daten werden möglicherweise nicht im erforderlichen Ausmaß gesichert.
- Der Datenzugriff und die Speicherleistung sind möglicherweise nicht optimal.

Implementierungsleitfaden

Definieren Sie die Datenmerkmale und Zugriffsmuster Ihrer Workload. Überprüfen Sie alle verfügbaren Datenbanklösungen, um herauszufinden, welche Lösung am besten Ihren Datenanforderungen entspricht. Innerhalb einer bestimmten Workload können mehrere Datenbanken ausgewählt werden. Evaluieren Sie jeden Service oder jede Servicegruppe und führen Sie eine individuelle Bewertung durch. Wenn potenzielle alternative Datenverwaltungslösungen für alle oder Teile der Daten identifiziert werden, experimentieren Sie mit unterschiedlichen Implementierungen, die zu Vorteilen hinsichtlich Kosten, Sicherheit, Leistung und Zuverlässigkeit führen können. Aktualisieren Sie die bestehende Dokumentation, falls ein neuer Ansatz zur Datenverwaltung eingeführt wird.

Typ	AWS-Services	Schlüsselmerkmale	Gängige Anwendungsfälle
Relational	Amazon RDS, Amazon Aurora	Referenzielle Integrität, ACID-Transaktionen, Schema-on-Write	ERP, CRM, kommerzielle Standardsoftware
Schlüssel-Werte-Datenbanken	Amazon DynamoDB	Hoher Durchsatz, geringe Latenz,	Einkaufswagen (E-Commerce),

Typ	AWS-Services	Schlüsselmerkmale	Gängige Anwendungsfälle
		beinahe unendliche Skalierbarkeit	Produktkataloge, Chat-Anwendungen
Dokumentdatenbanken	Amazon DocumentDB	JSON-Dokumente speichern und Abfragen zu jedem Attribut durchführen	Content-Management (CMS), Kundenprofile, mobile Anwendungen
In-Memory	Amazon ElastiCache, Amazon MemoryDB	Latenz im Mikrosekundenbereich	Caching, Ranglisten für Spiele
Graphdatenbanken	Amazon Neptune	Höchst relationale Daten, wobei die Beziehung zwischen Daten von Bedeutung ist	Soziale Netzwerke, Personalisierungs-Engines, Betrugserkennung
Zeitreihendatenbanken	Amazon Timestream	Daten, bei denen Zeit wesentlich ist	DevOps, IoT, Überwachung
Wide-Column-Datenbanken	Amazon Keyspaces	Cassandra-Workloads	Instandhaltung von Industrieanlagen, Routenoptimierung
Ledger-Datenbanken	Amazon QLDB	Unveränderliches und kryptografisch überprüfbares Änderungs-Ledger	Systems of Record, Gesundheitswesen, Lieferketten, Finanzinstitute

Implementierungsschritte

1. Wie sind die Daten strukturiert? (z. B. nicht strukturiert, Schlüssel-Wert, halbstrukturiert, relational)
   a. Wenn die Daten nicht strukturiert sind, erwägen Sie einen Objektspeicher wie Amazon S3 oder eine NoSQL-Datenbank wie Amazon DocumentDB.
   b. Erwägen Sie für Schlüssel-Werte-Daten DynamoDB, ElastiCache für Redis oder MemoryDB.

   c. Welche Ebene an Referenzintegrität ist erforderlich, wenn die Daten über eine relationale Struktur verfügen?

      i. Bei Fremdschlüsseleinschränkungen können relationale Datenbanken wie Amazon RDS und Aurora diese Integritätsebene bieten.

      ii. Üblicherweise würden Sie innerhalb eines NoSQL-Datenmodells Ihre Daten in ein einzelnes Dokument oder eine Sammlung von Dokumenten denormalisieren, die in einer einzelnen Anfrage abgerufen werden können, anstatt Daten in Dokumenten oder Tabellen zusammenzufügen.

2. Ist AKID-Compliance (Atomarität, Konsistenz, Isolation, Dauerhaftigkeit) erforderlich?

   a. Wenn mit relationalen Datenbanken zusammenhängende AKID-Eigenschaften erforderlich sind, erwägen Sie eine relationale Datenbank wie Amazon RDS und Aurora.

3. Welches Konsistenzmodell ist erforderlich?

   a. Wenn Ihre Anwendung eventuelle Kohärenz tolerieren kann, ziehen Sie eine NoSQL-Implementierung in Erwägung. Überprüfen Sie die anderen Eigenschaften, um zu bestimmen, welche NoSQL-Datenbank am besten geeignet ist.

   b. Wenn strikte Konsistenz erforderlich ist, können Sie strikt konsistente Lesevorgänge mithilfe von DynamoDB durchführen oder mit einer relationalen Datenbank wie Amazon RDSevaluieren.

4. Welche Abfrage- und Ergebnisformate müssen unterstützt werden? (z. B. SQL, CSV, Parque, Avro, JSON usw.)

5. Welche Datentypen, Feldgrößen und Gesamtmengen sind vorhanden? (z. B. Text, numerische, räumliche, zeitreihenkalkulierte, binäre oder BLOB-Daten)

6. Wie ändern sich die Speicheranforderungen im Laufe der Zeit? Wie beeinflusst dies die Skalierbarkeit?

   a. Serverless-Datenbanken wie DynamoDB und Amazon Quantum Ledger Database skalieren dynamisch auf beinahe unbeschränkten Speicher.

   b. Relationale Datenbanken haben oftmals Obergrenzen bei bereitgestelltem Speicher und müssen mithilfe von Mechanismen wie Sharding horizontal partitioniert werden, sobald sie diese Grenzen erreicht haben.

7. In welcher Proportion stehen Leseabfragen zu Schreibabfragen? Könnte Caching die Leistung verbessern?

   a. Leseintensive Workloads könnten von einer Caching-Ebene profitieren. Diese könnte ElastiCache oder DAX sein, wenn es sich bei der Datenbank um DynamoDB handelt.

   b. Lesevorgänge können auch zu Read Replicas mit relationalen Datenbanken ausgelagert werden, wie Amazon RDSevaluieren.

8. Wird Speicher und Modifizierung (OLTP – Online Transaction Processing) oder Abruf und Berichterstattung (OLAP – Online Analytical Processing) eine höhere Priorität eingeräumt?
   a. Erwägen Sie für Transaktionsverarbeitung mit hohem Durchsatz eine NoSQL-Datenbank wie DynamoDB oder Amazon DocumentDB.
   b. Erwägen Sie für analytische Abfragen eine spaltenbasierte Datenbank wie Amazon Redshift oder das Exportieren von Daten zu Amazon S3 und das Durchführen von Analysen mithilfe von Athena oder QuickSight.
9. Wie sensibel sind die Daten und welches Ausmaß an Schutz und Verschlüsselung erfordern sie?
   a. Alle Amazon RDS- und Aurora-Engines unterstützen Datenverschlüsselung im Ruhezustand mithilfe von AWS KMS. Microsoft SQL Server und Oracle unterstützen auch Transparent Data Encryption (TDE), wenn Amazon RDS verwendet wird.
   b. Sie können für DynamoDB eine differenzierte Zugriffskontrolle mit IAM verwenden, um zu steuern, wer Zugriff auf welche Daten auf Schlüsselebene hat.
10. Welches Ausmaß an Stabilität erfordern die Daten?
   a. Aurora repliziert Ihre Daten automatisch in drei Availability Zones innerhalb von einer Region, was bedeutet, dass Ihre Daten hochbeständig sind und eine geringere Wahrscheinlichkeit von Datenverlust besteht.
   b. DynamoDB wird automatisch in mehreren Availability Zones repliziert und bietet hohe Verfügbarkeit und Datenstabilität.
   c. Amazon S3 bietet eine Stabilitätsgarantie von 99,999999999 %. Viele Datenbankservices wie Amazon RDS und DynamoDB unterstützen das Exportieren von Daten zu Amazon S3 für Langzeitaufbewahrung und Archivierung.
11. Beeinflussen die Anforderungen an die Wiederherstellungsdauer (Recovery Time Objective, RTO) oder den Wiederherstellungszeitpunkt (Recovery Point Objective, RPO) die Lösung?
   a. Amazon RDS, Aurora, DynamoDB, Amazon DocumentDB und Neptune unterstützen alle Point-in-Time-Wiederherstellung und On-Demand-Sicherung und -Wiederherstellung.
   b. Bei Anforderungen für eine hohe Verfügbarkeit können DynamoDB-Tabellen mithilfe der Funktion Globale Tabellen global repliziert werden und Aurora-Cluster können mithilfe der Funktion „Globale Datenbanken“ innerhalb von mehreren Regionen repliziert werden. Zusätzlich können S3-Buckets in AWS-Regionen mithilfe von regionsübergreifender Replikation repliziert werden.

12. Besteht der Wunsch, sich von kommerziellen Datenbank-Engines/Lizenzkosten zu entfernen?
    a. Ziehen Sie Open-Source-Engines wie PostgreSQL und MySQL auf Amazon RDS oder Aurora in Erwägung.
    b. Nutzen Sie AWS DMS und AWS SCT zum Migrieren von kommerziellen Datenbank-Engines zu Open Source-Lösungen.
13. Was ist die Betriebserwartung an die Datenbank? Ist der Umstieg zu verwalteten Services eine Priorität?
    a. Das Verwenden von Amazon RDS anstatt von Amazon EC2 und DynamoDB oder Amazon DocumentDB anstatt eine NoSQL-Datenbank selbst zu hosten kann den Betriebsaufwand verringern.
14. Wie erfolgt derzeit der Zugriff auf die Datenbank? Handelt es sich nur um einen Anwendungszugriff oder gibt es Business-Intelligence (BI)-Benutzer und andere Standardanwendungen?
    a. Wenn Sie von externen Tools abhängig sind, müssen Sie möglicherweise mit der Datenbank, die unterstützt wird, die Kompatibilität aufrecht erhalten. Amazon RDS ist vollständig kompatibel mit den unterschiedlichen Engine-Versionen, die unterstützt werden, einschließlich Microsoft SQL Server, Oracle, MySQL und PostgreSQL.
15. Nachstehend finden Sie eine Liste von möglichen Datenmanagementservices und wo diese am besten verwendet werden können:
    a. In relationalen Datenbanken werden Daten mit vordefinierten Schemata und Beziehungen zwischen ihnen gespeichert. Diese Datenbanken unterstützen AKID-Transaktionen (Atomarität, Konsistenz, Isolation und Dauerhaftigkeit) und gewährleisten die referentielle Integrität sowie eine starke Datenkonsistenz. Bei zahlreichen herkömmlichen Anwendungen, Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) und E-Commerce werden relationale Datenbanken zum Speichern der Daten verwendet. Viele dieser Datenbank-Engines können Sie in Amazon EC2 ausführen oder Sie können einen der von AWS verwalteten Datenbankservices nutzen: Amazon Aurora, Amazon RDSund Amazon Redshiftevaluieren.
    b. Schlüssel-Werte-Datenbanken sind auf gängige Zugriffsmuster ausgelegt, üblicherweise zum Speichern und Abrufen großer Datenmengen. Diese Datenbanken bieten kurze Reaktionszeiten, selbst bei extrem großen Mengen gleichzeitiger Anforderungen. Web-Apps mit hohem Datenverkehr, E-Commerce-Systeme und Gaming-Anwendungen sind typische Anwendungsfälle für Schlüssel-Werte-Datenbanken. In AWS können Sie Amazon DynamoDBverwenden, eine vollständig verwaltete, regionsübergreifende, beständige Multi-

Master-Datenbank mit integrierter Sicherheit, Sicherung und Wiederherstellung sowie In-Memory-Caching für internetfähige Anwendungen.

c. In-Memory-Datenbanken werden für Anwendungen eingesetzt, die einen Echtzeitzugriff auf Daten, die niedrigste Latenz und den höchsten Durchsatz erfordern. Durch das direkte Speichern von Daten im Arbeitsspeicher liefern diese Datenbanken eine Latenz im Mikrosekundenbereich für Anwendungen, wo eine Latenz im Millisekundenbereich nicht ausreicht. Sie können In-Memory-Datenbanken für Anwendungs-Caching, Sitzungsverwaltung, Gaming-Bestenlisten und Geodatenanwendungen verwenden. Amazon ElastiCache ist ein vollständig verwalteter In-Memory-Datenspeicher, der mit Redis oder Memcachedevaluieren. Wenn die Anwendungen eine höhere Stabilität erfordern, bietet Amazon MemoryDB für Redis eine Kombination, da es ein stabiler In-Memory-Datenbankservice für ultraschnelle Leistung ist.

d. Eine Dokumentdatenbank ist darauf ausgelegt, halbstrukturierte Daten als JSON-ähnliche Dokumente zu speichern. Mit diesen Datenbanken können Entwickler Anwendungen wie Content Management, Kataloge und Benutzerprofile schnell erstellen und aktualisieren. Amazon DocumentDB ist ein schneller, skalierbarer, hochverfügbarer und vollständig verwalteter Dokumentdatenbank-Service, der MongoDB-Workloads unterstützt.

e. Ein Wide Column-Speicher ist eine Art NoSQL-Datenbank. Sie verwendet Tabellen, Zeilen und Spalten, aber im Gegensatz zu einer relationalen Datenbank können sich die Namen und das Format der Spalten von Zeile zu Zeile in derselben Tabelle unterscheiden. In der Regel werden Wide Column-Speicher in umfangreichen Branchen-Apps für Gerätewartung, Flottenverwaltung und Routenoptimierung eingesetzt. Amazon Keyspaces (für Apache Cassandra) ist ein skalierbarer, hoch verfügbarer und verwalteter Apache Cassandra-kompatibler Datenbankservice.

f. Graph-Datenbanken sind für Anwendungen gedacht, die in Millionen von Beziehungen zwischen hochgradig vernetzten Diagrammdatensätzen mit Millisekunden-Latenz navigieren und diese abfragen müssen. Viele Unternehmen verwenden Graph-Datenbanken für Betrugserkennung, soziale Netzwerke und Empfehlungs-Engines. Amazon Neptune ist ein schneller, zuverlässiger, vollständig verwalteter Graph-Datenbankservice, der das Erstellen und Ausführen von Anwendungen vereinfacht, die mit hochgradig verbundenen Datensätzen arbeiten.

g. Zeitreihen-Datenbanken erfassen, generieren und gewinnen auf effiziente Weise Einblicke aus Daten, die sich im Laufe der Zeit ändern. IoT-Anwendungen, DevOps und industrielle Telemetrie können Zeitreihen-Datenbanken nutzen. Amazon Timestream ist ein schneller, skalierbarer, vollständig verwalteter Zeitreihen-Datenbankservice für IoT- und

Betriebsanwendungen, der das Speichern und Analysieren von Billionen von Ereignissen pro Tag vereinfacht.

h. Ledger-Datenbanken bieten eine zentrale und vertrauenswürdige Instanz für die Verwaltung einer skalierbaren, unveränderlichen und kryptografisch überprüfbaren Aufzeichnung von Transaktionen für jede Anwendung. Ledger-Datenbanken werden für Datensatzsysteme, Lieferketten, Registrierungen und sogar Banktransaktionen verwendet. Amazon Quantum Ledger Database (Amazon QLDB) ist eine vollständig verwaltete Ledger-Datenbank, die ein transparentes, unveränderliches und kryptografisch überprüfbares Transaktionsprotokoll bereitstellt, das sich im Besitz einer zentralen vertrauenswürdigen Stelle befindet. Amazon QLDB verfolgt jede Änderung der Anwendungsdaten und pflegt einen vollständigen und überprüfbaren Änderungsverlauf.

Grad des Aufwands für den Implementierungsplan: Wenn eine Workload von einer Datenbanklösung zu einer anderen verschoben wird, stellt dies möglicherweise einen hohen Grad des Aufwands beim Faktorwechsel der Daten und Anwendung dar.

Ressourcen

Zugehörige Dokumente:

- Cloud-Datenbanken mit AWS
- AWS-Datenbank-Caching
- Amazon DynamoDB Accelerator
- Bewährte Methoden für Amazon Aurora
- Amazon Redshift-Leistung
- Die besten 10 Leistungstipps für Amazon Athena
- Bewährte Methoden für Amazon Redshift Spectrum
- Bewährte MethodenAmazon DynamoDB
- Wählen Sie zwischen EC2 und Amazon RDS
- Bewährte Methoden für die Implementierung von Amazon ElastiCache

Relevante Videos:

- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Verständliche Beschreibung des Amazon Aurora-Speichers: Funktionsweise (DAT309-R)

- Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster (DAT403-R1)

Zugehörige Beispiele:

- Optimierung von Datenmustern mithilfe von Amazon Redshift Data Sharing
- Datenbankmigrationen
- MS SQL Server – AWS Database Migration Service (DMS)-Replikationsdemo
- Praktischer Workshop für die Datenbankmodernisierung
- Amazon Neptune-Beispiele

PERF04-BP02 Prüfen der verfügbaren Optionen

Verstehen Sie die verfügbaren Datenbankoptionen und wie sie Ihre Leistung optimieren können, bevor Sie Ihre Datenverwaltungslösung auswählen. Identifizieren Sie mithilfe von Lasttests Datenbankmetriken, die für Ihre Workload wichtig sind. Während Sie die Datenbanktoptionen erkunden, sollten Sie unterschiedliche Aspekte in Betracht ziehen, wie Parametergruppen, Speicheroptionen, Arbeitsspeicher, Rechenvorgänge, Read Replica, eventuelle Kohärenz, Verbindungs-Pooling und Caching-Optionen. Experimentieren Sie mit diesen unterschiedlichen Konfigurationsoptionen, um die Metriken zu verbessern.

Gewünschtes Ergebnis: Eine Workload könne eine oder mehrere Datenbanklösungen verwenden, basierend auf Datentypen. Die Funktionen und Vorteile der Datenbank entsprechen optimal den Datenmerkmalen, Zugriffsmustern und Workload-Anforderungen. Zur Optimierung der Leistung und Kosten Ihrer Datenbank müssen Sie die Datenzugriffsmuster auswerten, um die entsprechenden Datenbankoptionen zu bestimmen. Evaluieren Sie akzeptable Abfragezeiten, um sicherzustellen, dass die ausgewählten Datenbankoptionen die Anforderungen erfüllen können.

Gängige Antimuster:

- Sie identifizieren Datenzugriffsmuster nicht.
- Ihnen fehlt das Bewusstsein für die Wahl der Konfigurationsoptionen der Datenverwaltungslösung.
- Sie verlassen sich ausschließlich auf das Vergrößern der Instance-Größe, ohne andere verfügbare Konfigurationsoptionen in Betracht zu ziehen.
- Sie testen die Skalierungsoptionen der ausgewählten Lösung nicht.

Vorteile der Einführung dieser bewährten Methode: Indem Sie Datenbankoptionen erkunden und mit ihnen experimentieren, können Sie möglicherweise Infrastrukturkosten senken, die Leistung und Skalierbarkeit verbessern und den Aufwand zur Verwaltung Ihrer Workloads verringern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

- Die Optimierung für eine Größe, die für alle Datenbanken passt, bedeutet, dass unnötige Kompromisse eingegangen werden müssen.
- Da die Datenbanklösung nicht für die Datenverkehrsmuster konfiguriert ist, entstehen höhere Kosten.
- Skalierungsprobleme können Schwierigkeiten beim Betrieb verursachen.
- Daten werden möglicherweise nicht im erforderlichen Ausmaß gesichert.

Implementierungsleitfaden

Sie müssen die Datenmerkmale Ihrer Workload kennen, damit Sie Ihre Datenbankoptionen konfigurieren können. Führen Sie Lasttests durch, um Ihre wesentlichen Leistungsmetriken und Engpässe zu identifizieren. Verwenden Sie diese Merkmale und Metriken, um Datenbankoptionen zu evaluieren und unterschiedliche Konfigurationen auszuprobieren.

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
Skalieren der Rechenvorgänge	Erhöhen der Instance-Größe, Aurora-Serverless-Instances skalieren automatisch als Reaktion	Automatisches Skalieren der Lese- und Schreibvorgänge mit einem On-Demand	Erhöhen der Instance-Größe	Erhöhen der Instance-Größe, Hinzufügen von Knoten zu einem Cluster	Erhöhen der Instance-Größe	Skaliert automatisch, um sich der Kapazität anzupassen	Automatisches Skalieren der Lese- und Schreibvorgänge mit einem On-Demand	Skaliert automatisch, um sich der Kapazität anzupassen

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
	auf Änderungen der Last	- Kapazitätsmodus oder automatisches Skalieren von bereitgestellter Kapazität für Lese- und Schreibvorgänge im bereitgestellten Kapazitätsmodus					- Kapazitätsmodus oder automatisches Skalieren von bereitgestellter Kapazität für Lese- und Schreibvorgänge im bereitgestellten Kapazitätsmodus	

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoDB	Amazon DocumentDB	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspaces	Amazon QLDB
Skalieren von Schreibvorgängen	Alle Enginges unterstützen Read Replicas. Aurora unterstützt die automatische Skalierung von Read-Replica-Instances	Erhöhen bereitgestellter Kapazitätseinheiten für bereitgestellte Lesevorgänge	Read Replicas	Read Replicas	Read Replicas. Unterstützt die automatische Skalierung von Read-Replica-Instances	Skaliert automatisch	Erhöhen bereitgestellter Kapazitätseinheiten für bereitgestellte Lesevorgänge	Skaliert automatisch zu dokumentierten Gleichzeitigkeitseinschränkungen hoch

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoDB	Amazon DocumentDB	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspaces	Amazon QLDB
Skalieren von Schreibvorgängen	Erhöhen der Instance-Größe, Batching von Schreibvorgängen in der Anwendung oder Hinzufügen einer Warteschlange vor die Datenbank. Horizontales Skalieren von mehreren Instances über Sharding auf Anwendungsebene	Erhöhen bereitgestellter Kapazitätseinheiten für bereitgestellte Schreibvorgänge. Sicherstellen von optimalen Partitionsschlüsseln, um die Drosselung von Schreibvorgängen auf Partitionsebene zu verhindern	Erhöhen der primären Instance-Größe	Verwenden von Redis im Cluster-Modus, um Schreibvorgänge auf Shards zu verteilen	Erhöhen der Instance-Größe	Schreibanforderungen können beim Skalieren gedrosselt werden. Wenn Drosselungsausnahmen auftreten, senden Sie Daten weiterhin mit dem gleichen (oder höheren) Durchsatz, um automatisch zu skalieren. Batch-Schreibvorgä	Erhöhen bereitgestellter Kapazitätseinheiten für bereitgestellte Schreibvorgänge. Sicherstellen von optimalen Partitionsschlüsseln, um die Drosselung von Schreibvorgängen auf Partitionsebene zu verhindern	Skaliert automatisch zu dokumentierten Gleichzeitigkeitseinschränkungen hoch

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon DocumentB	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
						nge, um gleichzeitige Schreibanforderungen zu verringern		
Engine-Konfiguration	Parametergruppen	–	Parametergruppen	Parametergruppen	Parametergruppen	–	–	–

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoDB	Amazon DocumentDB	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspaces	Amazon QLDB
Caching	In-Memory-Caching, über Parametergruppen konfigurierbar. Koppeln Sie mit einem dedizierten Cache wie ElastiCache für Redis, um Anforderungen für häufig genutzte Elemente auszulagern	DAX (DAX) vollständig verwalteter Cache verfügbar	In-Memory-Caching. Koppeln Sie optional mit einem dedizierten Cache wie ElastiCache für Redis, um Anforderungen für häufig genutzte Elemente auszulagern.	Die primäre Funktion ist das Caching	Verwende Sie den Abfrageergebnis-Cache, um die Ergebnisse der Leseabfrage in den Cache zu speichern	Timestream hat zwei Speicherstufen; eine davon ist eine Hochleistungs-In-Memory-Stufe	Stellen Sie einen separaten dedizierten Cache wie ElastiCache für Redis bereit, um Anforderungen für häufig genutzte Elemente auszulagern	–

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
Hohe Verfügbarkeit / Notfallwiederherstellung	Die empfohlene Konfiguration für Produktions-Workloads ist das Ausführen einer Standby-Instance in einer zweiten Availability Zone, um Stabilität innerhalb einer Region zu bieten. Für die Resilienz in	Innerhalb einer Region hoch verfügbar. Tabellen können innerhalb von Regionen mithilfe von globalen DynamoD Tabellen repliziert werden.	Erstellen Sie für die Verfügbarkeit mehrere Instances in Availability Zones. Schnappshüsse können in Regionen und Clustern mithilfe von DMS repliziert werden, um regionsübergreifende Replikation / Notfallwi	Die empfohlene Konfiguration für Produktions-Cluster ist, zumindest einen Knoten in einer sekundären Availability Zone zu erstellen. Der ElastiCache Global Datastore kann verwendet werden, um Cluster in Regionen	Read Replicas in anderen Availability Zones dienen als Failover-Ziele. Snaphots können innerhalb von Regionen geteilt werden und Cluster können mithilfe von Neptune-Streams repliziert werden, um die Daten	Innerhalb einer Region hoch verfügbar, regionsübergreifende Replikation erfordert benutzerdefinierte Anwendungsentwicklung mithilfe des Timestream SDK.	Innerhalb einer Region hoch verfügbar. Regionsübergreifende Replikation erfordert Anwendungslogik oder Tools von Drittanbietern.	Innerhalb einer Region hoch verfügbar. Exportieren Sie zum regionsübergreifenden Replizieren den Inhalt des Amazon-QLDB-Journals zu einem S3-Bucket und konfigurieren Sie den Bucket für eine regionsübergreifen

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
	Regionen kann Aurora Global Database verwende werden		ederherstellung zu bieten	zu replizieren.	zwischen zwei Clustern in zwei unterschiedlichen Regionen zu replizieren.			de Replikation.

Implementierungsschritte

1. Welche Konfigurationsoptionen sind für die ausgewählten Datenbanken verfügbar?
    a. Mithilfe von Parametergruppen für Amazon RDS und Aurora können Sie gemeinsame Einstellungen auf Datenbank-Engine-Ebene anpassen, wie den dem Cache zugewiesenen Arbeitsspeicher oder das Einstellen der Zeitzone der Datenbank.
    b. Bei bereitgestellten Datenbankservices wie Amazon RDS, Aurora, Neptune Amazon DocumentDB und jenen, die auf Amazon EC2 bereitgestellt werden, können Sie den Instance-Typ und den bereitgestellten Speicher ändern sowie Read Replicas hinzufügen.
    c. Mithilfe von DynamoDB können Sie zwei Kapazitätsmodi angeben: On-Demand und bereitgestellt. Sie können zwischen diesen Modi wechseln und die zugewiesene Kapazität im bereitgestellten Modus jederzeit erhöhen, um unterschiedliche Workloads zu bewältigen.
2. Ist die Workload lese- oder schreiblastig?
    a. Welche Lösungen sind zum Auslagern von Lesevorgängen verfügbar (Read Replicas, Caching usw.)?
        i. Bei DynamoDB-Tabellen können Sie Lesevorgänge mithilfe von DAX für Caching auslagern.

   ii. Sie können für relationale Datenbanken einen ElastiCache-for-Redis-Cluster erstellen und Ihre Anwendung so konfigurieren, dass sie zuerst aus dem Cache liest und dann auf die Datenbank zurückfällt, wenn das angeforderte Element nicht vorhanden ist.

   iii. Relationale Datenbanken wie Amazon RDS und Aurora sowie bereitgestellte NoSQL-Datenbanken wie Neptune und Amazon DocumentDB unterstützen alle das Hinzufügen von Read Replicas, um die Lesevorgänge der Workload auszulagern.

   iv. Serverless-Datenbanken wie DynamoDB skalieren automatisch. Stellen Sie sicher, dass Sie ausreichend Read Capacity Units (RCU) bereitstellen, um die Workload zu verarbeiten.

b. Welche Lösungen sind zum Skalieren von Schreibvorgängen verfügbar (Partitionsschlüssel-Sharding, Einsetzen von Warteschlangen usw.)?

   i. Bei relationalen Datenbanken können Sie die Größe der Instance erhöhen, um eine erhöhte Workload zu bewältigen, oder die bereitgestellten IOPs erhöhen, um einen erhöhten Durchsatz in den zugrunde liegenden Speicher zu ermöglichen.

      - Sie können vor Ihrer Datenbank auch eine Warteschlange einrichten, anstatt direkt in die Datenbank zu schreiben. Mithilfe dieses Musters können Sie die Datenerfassung von der Datenbank entkoppeln und die Flow-Rate steuern, sodass die Datenbank nicht überwältigt wird.
      - Das Batching Ihrer Schreibanforderungen, anstatt mehrere kurzlebige Transaktionen zu erstellen, kann Ihnen dabei helfen, den Durchsatz bei relationalen Datenbanken mit hohem Schreibvolumen zu verbessern.

   ii. Serverless-Datenbanken wie DynamoDB können den Schreibdurchsatz automatisch skalieren oder indem die bereitgestellten Kapazitätseinheiten für Schreibvorgänge (Write Capacity Units, WCU) abhängig vom Kapazitätsmodus angepasst werden.

      - Es können trotzdem Fehler mit einer „Hot Partition“ auftreten, wenn Sie die Durchsatzgrenzen für einen bestimmten Partitionsschlüssel erreichen. Dies kann verhindert werden, indem Sie einen Partitionsschlüssel auswählen, der gleichmäßiger verteilt ist, oder indem Sie die Schreibvorgänge des Partitionsschlüssels in Shards aufteilen.

3. Was sind derzeit die erwarteten höchsten Transaktionen pro Sekunde (TPS)? Testen Sie mithilfe dieser Datenverkehrsmenge und dieser Menge +X%, um die Skalierungsmerkmale zu verstehen.

   a. Native Tools wie pg_bench for PostgreSQL können eingesetzt werden, um die Datenbank einem Stresstest zu unterziehen und Engpässe sowie Skalierungsmerkmale zu verstehen.

   b. Produktionsdatenverkehr sollte erfasst werden, sodass er wiedergegeben werden kann, um zusätzlich zu künstlichen Workloads auch echte Bedingungen zu simulieren.

4. Wenn Sie Serverless-Datenverarbeitung oder elastisch skalierbare Datenverarbeitung verwenden, testen Sie die Auswirkungen, wenn diese auf der Datenbank skaliert wird. Führen Sie Verbindungsverwaltung oder -Pooling ein, falls zutreffend, um die Auswirkungen auf die Datenbank zu verringern.
   a. RDS Proxy kann mit Amazon RDS und Aurora verwendet werden, um Verbindungen mit der Datenbank zu verwalten.
   b. Serverless-Datenbanken wie DynamoDB haben keine ihnen zugewiesenen Verbindungen, aber ziehen Sie die bereitgestellte Kapazität sowie automatische Skalierungsrichtlinien in Betracht, um Datenverkehrsspitzen zu bewältigen.
5. Ist die Last vorhersehbar, gibt es Lastspitzen und Inaktivitätsphasen?
   a. Wenn es Inaktivitätsphasen gibt, erwägen Sie während dieser Zeitspanne das Herunterskalieren der bereitgestellten Kapazität oder Instance-Größe. Aurora Serverless V2 skaliert auf Basis der Last automatisch hoch oder herunter.
   b. Bei Instances außerhalb der Produktionsumgebung erwägen Sie das Pausieren oder Stoppen dieser Instances in arbeitsfreien Zeiten.
6. Müssen Sie Ihre Datenmodelle basierend auf Zugriffsmustern und Datenmerkmalen segmentieren und verteilen?
   a. Erwägen Sie die Verwendung von AWS DMS oder AWS SCT, um Ihre Daten zu anderen Services zu verschieben.

Grad des Aufwands für den Implementierungsplan:

Sie müssen Ihre aktuellen Dateneigenschaften und -metriken kennen, um diese bewährten Methoden einzurichten. Das Erfassen dieser Metriken, Festlegen einer Basislinie und Verwenden von Metriken zum Ermitteln der idealen Datenbankkonfiguration stellt einen niedrigen bis mittleren Grad des Aufwands dar. Die Validierung erfolgt am besten über Lasttests und Experimentieren.

Ressourcen

Zugehörige Dokumente:

- Cloud-Datenbanken mit AWS
- AWS-Datenbank-Caching
- Amazon DynamoDB Accelerator
- Bewährte Methoden für Amazon Aurora
- Amazon Redshift-Leistung

- Die besten 10 Leistungstipps für Amazon Athena
- Bewährte Methoden für Amazon Redshift Spectrum
- Bewährte MethodenAmazon DynamoDB

Relevante Videos:

- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Verständliche Beschreibung des Amazon Aurora-Speichers: Funktionsweise (DAT309-R)
- Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster (DAT403-R1)

Zugehörige Beispiele:

- Amazon DynamoDB-Beispiele
- Beispiele von AWS-Datenbankmigration
- Workshop für die Datenbankmodernisierung
- Arbeiten mit Parametern auf Ihrem Amazon RDS für Postgress DB

PERF04-BP03 Erfassen und Aufzeichnen von Metriken zur Datenbankleistung

Es ist wichtig, relevante Metriken nachzuverfolgen, um zu verstehen, welche Leistung Ihre Datenverwaltungssysteme erbringen. Mithilfe dieser Metriken können Sie Ihre Datenverwaltungsressourcen optimieren, um sicherzustellen, dass Ihre Workload-Anforderungen erfüllt werden, und um eine klare Übersicht über die Workload-Leistung zu erhalten. Nutzen Sie Tools, Bibliotheken und Systeme zum Aufzeichnen von Messungen zur Datenbankleistung.

Diese Metriken beziehen sich auf das System, auf dem die Datenbank gehostet wird (beispielsweise CPU, Speicher, Arbeitsspeicher, IOPS), und es gibt Metriken für den Zugriff auf die eigentlichen Daten (beispielsweise Transaktionen pro Sekunde, Abfrageraten, Reaktionszeiten, Fehler). Support- oder Betriebsmitarbeiter sollten auf diese Metriken zugreifen können und über ausreichend historische Datensätze verfügen, um Tendenzen, Anomalien und Engpässe identifizieren zu können.

Gewünschtes Ergebnis: Um die Leistung Ihrer Datenbank-Workloads zu überwachen, müssen Sie mehrere Leistungsmetriken über einen bestimmten Zeitraum aufzeichnen. Auf diese Weise

können Sie Anomalien erkennen und die Leistung anhand von Geschäftsmetriken messen, um sicherzustellen, dass Sie die Anforderungen Ihrer Workload erfüllen.

Gängige Antimuster:

- Sie suchen ausschließlich manuell mithilfe von Protokolldateien nach Metriken.
- Sie veröffentlichen Metriken nur in internen Tools, die von Ihrem Team verwendet werden, und Sie haben kein umfassendes Bild Ihrer Workload.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware Ihrer Wahl aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.
- Sie überwachen Metriken nur auf Systemebene und erfassen keine Datenzugriffs- und Nutzungsmetriken.

Vorteile der Einführung dieser bewährten Methode: Das Einrichten einer Leistungsbasislinie hilft dabei, normales Verhalten und die Anforderungen von Workloads zu verstehen. Abnorme Muster können schneller identifiziert und behoben werden, was die Leistung und Zuverlässigkeit der Datenbank erhöht. Die Datenbankkapazität kann konfiguriert werden, um die optimalen Kosten ohne Leistungseinschränkung sicherzustellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

- Wenn zwischen normalen und abnormalen Leistungsebenen nicht unterschieden wird, kann dies Schwierigkeiten bei der Fehlererkennung und Entscheidungsfindung verursachen.
- Potenzielle Kosteneinsparungen werden möglicherweise nicht erkannt.
- Wachstumsmuster werden nicht erkannt, was zur Verringerung von Zuverlässigkeit oder Leistung führen kann.

Implementierungsleitfaden

Identifizieren, sammeln, aggregieren und korrelieren Sie Datenbankmetriken. Metriken sollten das zugrunde liegende System, das die Datenbank unterstützt, sowie die Datenbankmetriken enthalten. Die Metriken des zugrunde liegenden Systems können die CPU-Auslastung, den Arbeitsspeicher, den verfügbaren Festplattenspeicher, Festplatten-E/A und Metriken zum eingehenden und ausgehenden Netzwerkdatenverkehr umfassen, während die Datenbankmetriken die Transaktionen pro Sekunde, die häufigsten Abfragen, die durchschnittlichen Abfrageraten, Antwortzeiten, die Indexauslastung, Tabellenschlösser, Abfragezeitüberschreitungen und die Anzahl offener

Verbindungen enthält. Diese Daten sind von entscheidender Bedeutung, um festzustellen, wie leistungsfähig die Workload ist und wie die Datenbanklösung genutzt wird. Nutzen Sie diese Kennzahlen im Rahmen eines datengestützten Ansatzes, der Ihnen die Feinabstimmung und Optimierung der vom Workload genutzten Ressourcen ermöglicht.

Implementierungsschritte:

1. Welche Datenbankmetriken sollten verfolgt werden?
    a. Überwachungmetriken für Amazon RDS
    b. Überwachung mit Leistungserkenntnissen
    c. Erweiterte Überwachung
    d. DynamoDB-Metriken
    e. Überwachung von DynamoDB DAX
    f. Überwachung von MemoryDB
    g. Überwachung von Amazon Redshift
    h. Zeitreihenmetriken und -dimensionen
    i. Cluster-Metriken für Aurora
    j. Überwachung von Amazon Keyspaces
    k. Überwachung von Amazon Neptune
2. Würde die Datenbanküberwachung von einer Machine-Learning-Lösung profitieren, die Betriebsanomalien und Leistungsprobleme erkennt?
    a. Amazon DevOps Guru for Amazon RDS ermöglicht einen Einblick in Leistungsprobleme und bietet Empfehlungen für Korrekturmaßnahmen.
3. Benötigen Sie Informationen über die SQL-Nutzung auf Anwendungsebene?
    a. AWS X-Ray kann in der Anwendung verwendet werden, um Erkenntnisse zu gewinnen und alle Datenpunkte für eine Abfrage zusammenzufassen.
4. Haben Sie derzeit eine genehmigte Protokollierungs- und Überwachungslösung?
    a. Mithilfe von Amazon CloudWatch lassen sich Kennzahlen aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können auch benutzerdefinierte Kennzahlen erfassen und in Oberflächen-, Geschäfts- oder abgeleiteten Kennzahlen veröffentlichen. Richten Sie mit CloudWatch oder mit Lösungen von Drittanbietern Alarme ein, die auf das Überschreiten von Schwellenwerten hinweisen.
5. Haben Ihre Datenaufbewahrungsrichtlinien identifiziert und konfiguriert, sodass sie Ihren Sicherheits- und Betriebszielen entsprechen?

a. Standard-Datenaufbewahrung für CloudWatch-Metriken
b. Standard-Datenaufbewahrung für CloudWatch Logs

Grad des Aufwands für den Implementierungsplan: Der Grad des Aufwands ist mittel, um Metriken von allen Datenbankressourcen zu identifizieren, nachzuverfolgen, zu erfassen, zu aggregieren und zu korrelieren.

Ressourcen

Ähnliche Dokumente:

- AWS-Datenbank-Caching
- Die besten 10 Leistungstipps für Amazon Athena
- Bewährte Methoden für Amazon Aurora
- Amazon DynamoDB Accelerator
- Bewährte MethodenAmazon DynamoDB
- Bewährte Methoden für Amazon Redshift Spectrum
- Amazon Redshift-Leistung
- Cloud-Datenbanken mit AWS
- Amazon RDS-Leistungserkenntnisse

Ähnliche Videos:

- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Verständliche Beschreibung des Amazon Aurora-Speichers: Funktionsweise (DAT309-R)
- Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster (DAT403-R1)

Ähnliche Beispiele:

- Level 100: Monitoring with CloudWatch Dashboards (Stufe 100: Überwachung mit Cloudwatch-Dashboards)
- AWS Dataset Ingestion Metrics Collection Framework (Framework zur AWS-Datenerfassung und Sammlung von Metriken)
- Amazon RDS Monitoring Workshop (Workshop zur Überwachung von Amazon RDS)

PERF04-BP04 Wählen des Datenspeichers nach Zugriffsmuster

Bestimmen Sie anhand der Zugriffsmuster des Workloads, welche Services und Technologien sich anbieten. Zusätzlich zu nicht-funktionalen Anforderungen wie Leistung und Skalierung, beeinflussen Zugriffsmuster stark die Wahl der Datenbank- und Speicherlösungen. Die erste Dimension ist der Bedarf nach Transaktionen, ACID-Compliance und konsistenten Lesevorgängen. Nicht jede Datenbank unterstützt dies und die meisten NoSQL-Datenbanken bieten ein Modell für die eventuelle Konsistenz Die zweite wichtige Dimension wäre die Verteilung von Schreib- und Lesevorgängen über Zeit und Raum. Global verteile Anwendungen müssen Datenverkehrsmuster, Latenz und Zugriffsanforderungen in Betracht ziehen, um die optimale Speicherlösung zu ermitteln. Der dritte auszuwählende wesentliche Aspekt ist die Flexibilität des Abfragemusters, zufällige Zugriffsmuster und einmalige Abfragen. Überlegungen zu hochspezialisierten Abfragefunktionen für die Verarbeitung von Text und natürlicher Sprache, Zeitreihendatenbanken und Diagrammen sollten ebenfalls in Betracht gezogen werden.

Gewünschtes Ergebnis: Der Datenspeicher wurde auf Basis von identifizierten und dokumentieren Datenzugriffsmustern ausgewählt. Dies kann die gängigsten Lese-, Schreib und Löschabfragen, die Notwendigkeit für Ad-hoc-Kalkulationen und -Aggregationen, die Komplexität von Daten, die Abhängigkeiten zwischen Daten und die benötigten Kohärenzanforderungen umfassen.

Gängige Antimuster:

- Sie wählen nur einen Datenbankanbieter aus, um die Betriebsverwaltung zu vereinfachen.
- Sie gehen davon aus, dass Datenzugriffsmuster im Laufe der Zeit konsistent bleiben.
- Sie implementieren komplexe Transaktionen, Rollback und Konsistenzlogik in der Anwendung.
- Die Datenbank ist konfiguriert, um potenzielle Datenverkehrsspitzen zu unterstützen, was dazu führt, dass die Datenbankressourcen die meiste Zeit nicht genutzt werden.
- Es wird eine geteilte Datenbank für Transaktions- und analytische Anwendungen verwendet.

Vorteile der Einführung dieser bewährten Methode: Das Auswählen und Optimieren Ihres Datenspeichers auf Basis von Zugriffsmustern hilft beim Verringern der Entwicklungskomplexität und Optimieren Ihrer Leistungsmöglichkeiten. Verstehen, wann Read Replicas, globale Tabellen, Datenpartitionen und Caching verwendet werden sollen und wie dies Ihnen dabei hilft, den Betriebsaufwand zu verringern und basierend auf Ihren Workload-Anforderungen zu skalieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Identifizieren und evaluieren Sie Ihre Datenzugriffsmuster, um die richtige Speicherkonfiguration auszuwählen. Bei jeder Datenbanklösung gibt es Optionen, um Ihre Speicherlösung zu konfigurieren und zu optimieren. Verwenden Sie die erfassten Metriken und Protokolle und experimentieren Sie mit Optionen, um die optimale Konfiguration zu finden. Verwenden Sie die nachfolgende Tabelle, um Speicheroptionen je nach Datenbankservice anzusehen.

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
Skalierung von Speicher	Eine Option zur automatischen Skalierung des Speichers ist verfügbar, um den bereitgestellten Speicher automatisch zu skalieren. IOPS können ebenfalls unabhängig vom bereitges	Skaliert automatisch. Tabellen haben keine Größenein schränkung.	Eine Option zur automatischen Skalierung des Speichers ist verfügbar, um den bereitgestellten Speicher automatisch zu skalieren	In-Memory-Speicher, der an den Instance-Typ oder die Instance-Anzahl gebunden ist	Eine Option zur automatischen Skalierung des Speichers ist verfügbar, um den bereitgestellten Speicher automatisch zu skalieren	Konfigurieren einer Aufbewahrungsfrist in Tagen für In-Memory- und Magnetebnen	Skaliert Tabellenspeicher automatisch hoch und herunter	Skaliert automatisch. Tabellen haben keine Größenein schränkung.

AWS-Services	Amazon RDS, Amazon Aurora	Amazon DynamoD	Amazon Document B	Amazon ElastiCache	Amazon Neptune	Amazon Timestream	Amazon Keyspace	Amazon QLDB
	tellten Speicher skaliert werden, wenn bereitgestellte IOPS-Speichertypen eingesetzt werden.							

Implementierungsschritte:

1. Identifizieren und dokumentieren Sie das antizipierte Wachstum von Daten und Datenverkehr.
    a. Amazon RDS und Aurora unterstützen automatische Skalierung bis hin zu dokumentierten Grenzen. Erwägen Sie darüber hinaus, ältere Daten für die Archivierung an Amazon S3 zu übertragen, historische Daten für Analyseverfahren zu aggregieren oder über Sharding horizontal zu skalieren.
    b. DynamoDB und Amazon S3 skalieren automatisch auf beinahe unbegrenztes Speichervolumen.
    c. Amazon RDS-Instances and Datenbanken, die auf EC2 ausgeführt werden, können manuell in ihrer Größe angepasst werden und zu EC2-Instances können später neue EBS-Datenträger für zusätzlichen Speicher hinzugefügt werden.
    d. Instance-Typen können auf Basis von Aktivitätsänderungen geändert werden. Sie können beispielsweise mit einer kleineren Instance starten, während Sie Tests durchführen, und die Instance dann skalieren, wenn allmählich Produktionsdatenverkehr im Service eingeht. Aurora Serverless V2 skaliert automatisch als Antwort auf Änderungen bei der Last.

1. Dokumentieren Sie Anforderungen in Bezug auf normale und Spitzenleistung (Transaktionen pro Sekunde, TPS, und Abfragen pro Sekunde, QPS) sowie Kohärenz (ACID und eventuelle Kohärenz).
2. Dokumentieren Sie Bereitstellungsaspekte der Lösung und die Zugriffsanforderungen der Datenbank (global, Multi-AZ, Read Replication, mehrere Schreibknoten).

Grad des Aufwands für den Implementierungsplan: Wenn Sie keine Protokolle oder Metriken für Ihre Datenverwaltungslösung haben, müssen Sie dies zuerst abschließen, bevor Sie Ihre Datenzugriffsmuster identifizieren und dokumentieren. Sobald Sie Ihr Datenzugriffsmuster verstanden haben, stellen das Auswählen und Konfigurieren Ihres Datenspeichers einen niedrigen Grad des Aufwands dar.

Ressourcen

Ähnliche Dokumente:

- AWS-Datenbank-Caching
- Die besten 10 Leistungstipps für Amazon Athena
- Bewährte Methoden für Amazon Aurora
- Amazon DynamoDB Accelerator
- Bewährte MethodenAmazon DynamoDB
- Bewährte Methoden für Amazon Redshift Spectrum
- Amazon Redshift-Leistung
- Cloud-Datenbanken mit AWS
- Amazon RDS-Speichertypen

Ähnliche Videos:

- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Verständliche Beschreibung des Amazon Aurora-Speichers: Funktionsweise (DAT309-R)
- Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster (DAT403-R1)

Ähnliche Beispiele:

- Experimentieren mit und Testen von verteilten Lasttests auf AWS

PERF04-BP05 Optimieren des Datenspeicher nach Zugriffsmuster und Metriken

Optimieren Sie anhand der Leistungsmerkmale und Zugriffsmuster die Art und Weise, in der Daten gespeichert oder abgefragt werden. So lässt sich die bestmögliche Leistung erzielen. Messen Sie, wie sich Optimierungen, z. B. Indizierung, Schlüsselverteilung, Data Warehouse Design oder Caching-Strategien, auf die Systemleistung oder die allgemeine Effizienz auswirken.

Gängige Antimuster:

- Sie suchen ausschließlich manuell mithilfe von Protokolldateien nach Metriken.
- Sie veröffentlichen Metriken nur in internen Tools.

Vorteile der Einführung dieser bewährten Methode: Um sicherzustellen, dass Sie die für die Workload erforderlichen Metriken erfüllen, müssen Sie die Datenbank-Leistungsmetriken für Lese- und Schreibvorgänge überwachen. Sie können diese Daten verwenden, um neue Optimierungen für Lese- und Schreibvorgänge zur Datenspeicherebene hinzuzufügen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Speicher basierend auf Kennzahlen und Mustern optimieren: Verwenden Sie gemeldete Metriken, um Bereiche in Ihrer Workload zu identifizieren und Ihre Datenbankkomponenten zu optimieren. Für jedes Datenbanksystem müssen eigene Leistungsmerkmale in Betracht gezogen werden, etwa das Verfahren, mit dem Daten indiziert, in den Cache gelesen oder auf mehrere Systeme verteilt werden. Messen Sie die Auswirkungen Ihrer Optimierungen.

Ressourcen

Zugehörige Dokumente:

- AWS-Datenbank-Caching
- Die besten 10 Leistungstipps für Amazon Athena
- Bewährte Methoden für Amazon Aurora
- Amazon DynamoDB Accelerator
- Bewährte MethodenAmazon DynamoDB
- Bewährte Methoden für Amazon Redshift Spectrum
- Amazon Redshift-Leistung

- Cloud-Datenbanken mit AWS
- Analysieren von Leistungsanomalien mit DevOps Guru für RDS
- Lese-/Schreibmodus für DynamoDB

Relevante Videos:

- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Verständliche Beschreibung des Amazon Aurora-Speichers: Funktionsweise (DAT309-R)
- Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster (DAT403-R1)

Zugehörige Beispiele:

- Praktische Übungen für Amazon DynamoDB

LEIST 5 Was ist beim Konfigurieren der Netzwerklösung zu beachten?

Welche Netzwerklösung für eine Workload optimal ist, richtet sich nach der Latenz, dem erforderlichen Durchsatz, dem Jitter und der Bandbreite. Die Standortoptionen sind von den physischen Einschränkungen abhängig, z. B. von Benutzer- oder lokalen Ressourcen. Diese Einschränkungen können durch Edge-Standorte oder die Ressourcenplatzierung wettgemacht werden.

Bewährte Methoden

- PERF05-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung
- PERF05-BP02 Evaluieren verfügbarer Netzwerkfunktionen
- PERF05-BP03 Auswählen einer richtig ausgelegten dedizierten Konnektivität oder eines VPN für Hybrid-Workloads:
- PERF05-BP04 Nutzen von Lastausgleich und Verschlüsselungsauslagerung
- PERF05-BP05 Auswählen leistungsfördernder Netzwerkprotokolle
- PERF05-BP06 Auswählen des Workload-Standortes entsprechend den Netzwerkanforderungen
- PERF05-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken

PERF05-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung

Analysieren Sie, wie sich Netzwerkentscheidungen auf die Leistung des Workloads auswirken. Das Netzwerk ist für die Verbindung zwischen Anwendungskomponenten, Cloud-Services, Edge-Netzwerken und On-Premises-Daten verantwortlich und kann daher die Workload-Leistung wesentlich beeinflussen. Die Benutzererfahrung wird nicht nur durch die Workload-Leistung, sondern auch durch die Netzwerklatenz, die Bandbreite, Protokolle, den Standort, Netzwerküberlastungen, Jitter, den Durchsatz und Routing-Regeln beeinträchtigt.

Gewünschtes Ergebnis: Sie haben eine dokumentierte Liste an Netzwerkanforderungen der Workload, einschließlich Latenz, Paketgröße, Routingregeln, Protokolle und unterstützender Datenverkehrsmuster. Sie überprüfen alle verfügbaren Netzwerklösungen und identifizieren, welcher Dienst den Netzwerkmerkmalen Ihrer Workload entspricht. Da cloudbasierte Netzwerke schnell geändert werden können, müssen Sie Ihre Netzwerkarchitektur im Laufe der Zeit weiterentwickeln, um die effiziente Leistung zu verbessern.

Gängige Antimuster:

- Jeglicher Datenverkehr fließt durch Ihre bestehenden Rechenzentren.
- Sie erstellen große Direct-Connect-Sitzungen, ohne die tatsächlichen Nutzungsanforderungen zu verstehen.
- Sie berücksichtigen beim Definieren Ihrer Netzwerklösungen die Workload-Eigenschaften und den Verschlüsselungsaufwand nicht.
- Sie verwenden On-Premises-Konzepte und -Strategien für Netzwerklösungen in der Cloud.

Vorteile der Einführung dieser bewährten Methode: Indem Sie verstehen, wie das Netzwerk die Workload-Leistung beeinflusst, können Sie potenzielle Engpässe erkennen, die Benutzererfahrung verbessern, die Zuverlässigkeit erhöhen und den Betriebsaufwand verringern, während sich die Workload verändert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Identifizieren Sie wichtige Metriken der Netzwerkleistung für Ihre Workload und erfassen Sie ihre Netzwerkeigenschaften. Definieren und dokumentieren Sie Anforderungen im Rahmen eines datengestützten Ansatzes unter Einsatz von Benchmarking oder Lasttests. Ermitteln Sie anhand dieser Daten, an welcher Stelle die Netzwerklösung Defizite hat. Prüfen Sie anschließend die

Konfigurationsoptionen, mit denen die der Workload verbessert werden könnte. Verstehen Sie die verfügbaren cloudnativen Netzwerkfunktionen und -optionen und wie diese Ihre Workload-Leistung basierend auf den Anforderungen beeinflussen können. Jede Netzwerkfunktion hat Vor- und Nachteile und kann konfiguriert werden, um Ihren Workload-Merkmalen zu entsprechen und basierend auf Ihren Anforderungen zu skalieren.

Implementierungsschritte:

1. Definieren und dokumentieren Sie die Anforderungen an die Netzwerkleistung:
    a. Schließen Sie Metriken wie Netzwerklatenz, Bandbreite, Protokolle, Standorte, Datenverkehrsmuster (Spitzen und Frequenz), Durchsatz, Verschlüsselung, Überprüfung und Routingregeln mit ein.
2. Erfassen Sie die Merkmale Ihres grundlegenden Netzwerks:
    a. VPC Flow Logs
    b. Merkmale des AWS Transit Gateway
    c. AWS PrivateLink-Metriken
3. Erfassen Sie die Merkmale Ihres Anwendungsnetzwerks:
    a. Elastic Network Adapter
    b. AWS-App Mesh-Metriken
    c. Amazon API Gateway-Metriken
4. Erfassen Sie die Merkmale Ihres Edge-Netzwerks:
    a. Amazon CloudFront-Metriken
    b. Amazon Route 53-Metriken
    c. AWS-Global Accelerator-Metriken
5. Erfassen Sie die Merkmale Ihres Hybridnetzwerks:
    a. Direct-Connect-Metriken
    b. AWS-Site-to-Site-VPN-Metriken
    c. AWS-Client-VPN-Metriken
    d. AWS Cloud-WAN-Metriken
6. Erfassen Sie die Merkmale Ihres Sicherheitsnetzwerks:
    a. AWS Shield, WAF und Netzwerk-Firewall-Metriken
7. Erfassen Sie End-to-End-Leistungsmetriken mit Tools zur Nachverfolgung:
    a. AWS X-Ray

   b. Amazon CloudWatch RUM
8. Benchmarks für die Netzwerkleistung festlegen und testen:
   a. Benchmark- Netzwerkdurchsatz: Einige Faktoren, die EC2-Netzwerkleistung beeinflussen können, wenn sich die Instances in der gleichen VPC befinden. Messen Sie die Netzwerkbandbreite zwischen EC2-Linux-Instances in der gleichen VPC.
   b. Führen Sie Lasttests durch, um mit Netzwerklösungen und -optionen zu experimentieren

Grad des Aufwands für den Implementierungsplan: Der Grad des Aufwands ist mittel, um die Netzwerkanforderungen Ihrer Workload, die Optionen und die verfügbaren Lösungen zu dokumentieren.

Ressourcen

Ähnliche Dokumente:

- Application Load Balancer
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances
- Network Load Balancer
- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs

Ähnliche Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)
- Improve Global Network Performance for Applications (Verbessern der Leistung von globalen Netzwerken für Anwendungen)

- EC2 Instances and Performance Optimization Best Practices (Bewährte Methoden für EC2-Instances und Leistungsoptimierung)
- Optimizing Network Performance for Amazon EC2-Instances (Optimieren der Netzwerkleistung für EC2-Instances)
- Networking best practices and tips with the Well-Architected Framework (Bewährte Methoden für Netzwerke und Tipps für das Well-Architected Framework)
- AWS networking best practices in large-scale migrations (Bewährte Methoden für AWS-Netzwerke in umfangreichen Migrationen)

Ähnliche Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP02 Evaluieren verfügbarer Netzwerkfunktionen

Prüfen Sie die Netzwerkfunktionen in der Cloud, mit denen die Leistung unter Umständen verbessert werden kann. Messen Sie die Auswirkungen der Funktionen anhand von Tests, Metriken und Analysen. Nutzen Sie beispielsweise die verfügbaren Funktionen auf Netzwerkebene, um die Latenz, den Paketverlust oder den Jitter zu reduzieren.

Viele Services werden zur Verbesserung der Leistung entwickelt, andere bieten Funktionen zur Optimierung der Netzwerkleistung. Services wie AWS, Global Accelerator und Amazon CloudFront dienen der Leistungsverbesserung, während die meisten anderen Services über Produktfunktionen zur Optimierung des Netzwerkdatenverkehrs verfügen. Sehen Sie sich zur Verbesserung Ihrer Workload-Leistung Servicefunktionen wie EC2-Instance-Netzwerkfunktionen, erweiterte Netzwerk-Instance-Typen, für Amazon EBS optimierte Instances, Amazon S3 Transfer Acceleration sowie CloudFront an.

Gewünschtes Ergebnis: Sie haben den Bestand an Komponenten in Ihrer Workload dokumentiert und ermittelt, welche Netzwerkkonfigurationen für die einzelnen Komponenten Ihnen helfen werden, Ihre Leistungsanforderungen zu erfüllen. Nach der Evaluierung der Netzwerkfunktionen haben Sie experimentiert und die Leistungsmetriken gemessen, um herauszufinden, wie Sie die Ihnen zur Verfügung stehenden Funktionen nutzen können.

Typische Anti-Muster:

- Sie bringen alle Ihre Workloads in eine Ihrem Hauptsitz am nächsten liegende AWS-Region und nicht in eine AWS-Region in der Nähe Ihrer Endbenutzer.
- Sie versäumen es, ein Benchmarking Ihrer Workload-Leistung durchzuführen und Ihre Workload-Leistung kontinuierlich anhand dieser Benchmark zu bewerten.
- Sie prüfen die Servicekonfigurationen nicht auf Optionen zur Leistungsverbesserung.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie alle Servicefunktionen und Optionen evaluieren, kann dies die Workload-Leistung verbessern, die Infrastrukturkosten senken, den Verwaltungsaufwand für die Workload reduzieren und die allgemeine Sicherheit erhöhen. Dank der weltweiten Abdeckung von AWS können Sie Ihren Kunden stets das bestmögliche Netzwerkerlebnis bieten.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Sehen Sie sich die verfügbaren Konfigurationsoptionen für das Netzwerk an und finden Sie heraus, wie sich diese auf Ihre Workload auswirken. Für die Leistungsoptimierung ist es entscheidend, zu verstehen, wie diese Optionen mit Ihrer Architektur interagieren und welche Auswirkungen sie auf die gemessene Leistung und die von den Benutzern wahrgenommene Leistung haben.

Implementierungsschritte:

1. Erstellen Sie eine Liste der Workload-Komponenten.
   a. Erstellen, verwalten und überwachen Sie das Netzwerk Ihres Unternehmens mithilfe von AWS Cloud WAN.
   b. Erhalten Sie Einblicke in Ihr Netzwerk unter Verwendung von Network Manager. Verwenden Sie ein vorhandenes Konfigurationsmanagementdatenbank-Tool (CMDB-Tool) oder eine Tool wie AWS Config, um eine Bestandsaufnahme Ihrer Workload und deren Konfiguration zu erstellen.
2. Wenn es sich um einen bestehenden Workload handelt, ermitteln und dokumentieren Sie die Benchmark für Ihre Leistungsmetriken. Konzentrieren Sie sich dabei auf Engpässe und Bereiche mit Verbesserungspotenzial. Leistungsbezogene Netzwerkmetriken werden je nach geschäftlichen Anforderungen und Workload-Merkmalen für die einzelnen Workloads unterschiedlich sein. Für den Anfang könnte die Prüfung folgender Metriken für Ihre Workload wichtig sein: Bandbreite, Latenz, Paketverlust, Jitter und erneute Übertragungen.
3. Bei einer neuen Workload sollten Sie Lasttests durchführen, um Leistungsengpässe zu identifizieren.

4. Prüfen Sie für die ermittelten Leistungsengpässe die Konfigurationsoptionen Ihrer Lösungen, um Möglichkeiten zur Leistungsverbesserung zu finden.
5. Wenn Sie Ihren Netzwerkpfad oder Ihre Netzwerkrouten nicht kennen, verwenden Sie Network Access Analyzer, um sie zu ermitteln.
6. Prüfen Sie Ihre Netzwerkprotokolle, um die Latenz weiter zu reduzieren.
   - PERF05-BP05 Auswählen leistungsfördernder Netzwerkprotokolle
7. Wenn Sie ein AWS Site-to-Site VPN über mehrere Standorte hinweg verwenden, um eine Verbindung zu einer AWS-Region herzustellen, prüfen Sie beschleunigte Site-to-Site VPN-Verbindungen auf Möglichkeiten zur Verbesserung der Netzwerkleistung.
8. Wenn Ihr Workload-Datenverkehr über mehrere Konten verteilt ist, evaluieren Sie Ihre Netzwerktopologie und Ihre Services, um die Latenz zu verringern.
   - Bewerten Sie Ihre betrieblichen und leistungsbezogenen Kompromisse zwischen VPC Peering und AWS Transit Gateway bei Verbindung mehrerer Konten. AWS Transit Gateway unterstützt die Skalierung eines AWS-Site-to-Site-VPN-Durchsatzes über eine einzelne IPsec-Höchstgrenze hinaus durch die Verwendung von Multi-Path. Der Datenverkehr zwischen einer Amazon VPC und AWS Transit Gateway bleibt im privaten AWS-Netzwerk und erfolgt nicht über das Internet. AWS Transit Gateway vereinfacht die Verbindung zwischen allen Ihren VPCs, die Tausende von AWS-Konten umfassen und in On-Premises-Netzwerke hineinreichen können. Teilen Sie Ihr AWS Transit Gateway zwischen mehreren Konten mit Resource Access Manager. Wenn Sie einen Einblick in Ihren globalen Netzwerkdatenverkehr erhalten möchten, verwenden Sie Network Manager, um einen zentralen Überblick über Ihre Netzwerkmetriken zu erhalten.
9. Prüfen Sie die Standorte Ihrer Benutzer und minimieren Sie die Distanz zwischen Ihren Benutzern und der Workload.
   a. AWS Global Accelerator ist ein Netzwerkservice, der die Leistung des Benutzerdatenverkehrs unter Verwendung der globalen Netzwerkinfrastruktur von Amazon Web Services um bis zu 60 % verbessert. Bei einer Überlastung des Internets optimiert AWS Global Accelerator den Weg zu Ihrer Anwendung, um Paketverluste, Jitter und Latenz konsistent niedrig zu halten. Der Service bietet auch statische IP-Adressen, die die Verschiebung von Endpunkten zwischen Availability Zones oder AWS-Regionen erleichtern, ohne dass Ihre DNS-Konfiguration aktualisiert werden muss oder kundenorientierte Anwendungen geändert werden müssen.
   b. Amazon CloudFront kann die Leistung Ihrer Workload-Inhaltsbereitstellung und die Latenz global verbessern. CloudFront verfügt über 410 weltweit verteilte Points of Presence, die Ihre Inhalte zwischenspeichern und die Latenzzeit für den Endbenutzer verringern können.

c. Amazon Route 53 bietet Optionen für latenzbasiertes Routing, Geolocation-Routing, Routing auf der Grundlage der geografischen Näheund IP-basiertes Routing und trägt damit zur Leistungsverbesserung der Workload für eine globale Zielgruppe bei. Ermitteln Sie, welche Routing-Option Ihre Workload-Leistung optimieren würde. Prüfen Sie dazu Ihren Workload-Datenverkehr und den Benutzerstandort.

10. Evaluieren Sie weitere Amazon S3-Funktionen zur Verbesserung der Speicher-IOPS.

a. Amazon S3 Transfer Acceleration ist eine Funktion, mit deren Hilfe externe Benutzer beim Hochladen von Daten in Amazon S3 von den Netzwerkoptimierungen von CloudFront profitieren können. Dies erleichtert die Übertragung großer Datenmengen von Remote-Standorten ohne spezielle Konnektivität zur AWS Cloud.

b. Multi-Region-Zugriffspunkte in Amazon S3 replizieren Inhalte in mehreren Regionen und vereinfachen die Workload durch die Bereitstellung eines Zugriffspunkts. Bei Verwendung eines Multi-Region-Zugriffspunkts können Sie Daten anfordern oder in Amazon S3 schreiben, wobei der Service den Bucket mit der geringsten Latenz ermittelt.

11. Prüfen Sie die Netzwerkbandbreite Ihrer Computing-Ressource.

a. Die von EC2-Instances, Containern und Lambda-Funktionen verwendeten Elastic-Network-Schnittstellen (ENA) sind pro Fluss begrenzt. Prüfen Sie Ihre Platzierungsgruppen, um Ihren EC2-Netzwerkdurchsatz zu optimieren.. Um Engpässe auf Pro-Fluss-Basis zu vermeiden, sollten Sie Ihre Anwendung so gestalten, dass mehrere Flüsse verwendet werden. Um Ihre datenverarbeitungsbezogenen Netzwerkmetriken zu überwachen und Einblicke in diese Metriken zu erhalten, verwenden Sie CloudWatch Metrics und `ethtool`. `ethtool` ist im ENA-Treiber enthalten und stellt zusätzliche netzwerkbezogene Metriken zur Verfügung, die als benutzerdefinierte Metriken in CloudWatch veröffentlicht werden können.

b. Neuere EC2-Instances können von Enhanced Networking profitieren. EC2-Instances der N-Seriewie z. B. `M5n` und `M5dn`nutzen die vierte Generation benutzerdefinierter Nitro-Karten, um einen Netzwerkdurchsatz von bis zu 100 Gbit/s zu einer einzelnen Instance zu bieten. Diese Instances bieten das Vierfache an Netzwerkbandbreite und Paketverarbeitung im Vergleich zu den einfachen `M5`- Instances und sind damit ideal für netzwerkintensive Anwendungen.

c. Amazon Elastic Network Adapters (ENA) ermöglichen eine weitere Optimierung, da sie einen besseren Durchsatz für Ihre Instances innerhalb einer Cluster-Placement-Gruppe bieten.

d. Elastic Fabric Adapter (EFA) ist eine Netzwerkschnittstelle für Amazon EC2-Instances, mit der Sie Workloads, die ein hohes Maß an Kommunikation zwischen Knoten erfordern, in AWS bedarfsgesteuert ausführen können. Bei EFA kann für HPC-Anwendungen (High Performance Computing) mit Message Passing Interface (MPI) und für ML-Anwendungen (Machine Learning)

mit NVIDIA Collective Communications Library (NCCL) eine Skalierung auf Tausende von CPUs oder GPUs durchgeführt werden.

e. Amazon EBS-optimierte Instances verwenden einen optimierten Konfigurations-Stack und stellen zusätzliche dedizierte Kapazität zur Erhöhung der Amazon EBS-I/O bereit. Sie können damit die Leistung Ihrer EBS-Volumes maximieren, indem Sie Konflikte zwischen Amazon Amazon EBS-I/O und anderem Datenverkehr von Ihrer Instance minimieren.

Aufwand für den Implementierungsplan:

Um diese bewährte Methode einzuführen, müssen Sie die Optionen Ihrer aktuellen Workload-Komponenten kennen, die sich auf die Netzwerkleistung auswirken. Das Zusammentragen der Komponenten, die Bewertung der Optionen zur Netzwerkverbesserung, das Experimentieren, die Umsetzung und die Dokumentation dieser Verbesserung erfordern einen geringen bis moderaten Aufwand.

Ressourcen

Zugehörige Dokumente:

- Amazon EBS – Optimierte Instances
- Application Load Balancer
- Netzwerkbandbreite der Amazon EC2-Instance
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances
- Network Load Balancer
- Netzwerkprodukte mit AWS
- AWS Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs
- Entwicklung einer Cloud-CMDB
- Skalieren des VPN-Durchsatzes mithilfe von AWS Transit Gateway

Zugehörige Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)
- AWS Global Accelerator

Zugehörige Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP03 Auswählen einer richtig ausgelegten dedizierten Konnektivität oder eines VPN für Hybrid-Workloads:

Wenn zum Verbinden von On-Premises- und Cloud-Ressourcen in AWS ein gemeinsames Netzwerk erforderlich ist, müssen Sie sicherstellen, dass Sie über die entsprechende Bandbreite verfügen, damit Ihre Leistungsanforderungen erfüllt werden. Schätzen Sie, welche Anforderungen an die Bandbreite und Latenz für Ihre Hybrid-Workload bestehen. Diese Werte dienen als Grundlage für die Größenanpassung für AWS Direct Connect oder Ihre VPN-Endpunkte.

Gewünschtes Ergebnis: Bei der Bereitstellung einer Workload, die eine hybride Netzwerkkonnektivität erfordert, gibt es dafür mehrere Konfigurationsoptionen, z. B. verwaltete oder nicht verwaltete VPNs oder Direct Connect. Wählen Sie den geeigneten Verbindungstyp für jede Workload aus, während Sie sicherstellen, dass zwischen Ihrem Standort und der Cloud ausreichend Bandbreite verfügbar ist und die Verschlüsselungsanforderungen erfüllt werden.

Gängige Antimuster:

- Sie evaluieren nur VPN-Lösungen für Ihre Netzwerkverschlüsselungsanforderungen.
- Sie evaluieren keine Optionen für Sicherung oder parallele Verbindungen.
- Sie verwenden Standardkonfigurationen für Router, Tunnel und BGP-Sessions.
- Sie verstehen nicht alle Workload-Anforderungen oder können sie nicht identifizieren (Verschlüsselung, Protokoll, Bandbreite und Datenverkehrsanforderungen).

Vorteile der Einführung dieser bewährten Methode: Das Auswählen und Konfigurieren von hybriden Netzwerklösungen mit angemessener Größe erhöht die Zuverlässigkeit Ihrer Workload und maximiert die Leistungsmöglichkeiten. Indem Sie die Workload-Anforderungen identifizieren, im Voraus planen

und hybride Lösungen evaluieren, verringern Sie teure physische Netzwerkänderungen sowie den Betriebsaufwand und beschleunigen die Markteinführung.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Entwickeln einer hybriden Netzwerkarchitektur entsprechend den Bandbreitenanforderungen: Schätzen Sie die Anforderungen an Bandbreite und Latenz für Ihre Hybridanwendungen ab. Je nach Bandbreitenanforderungen reicht eine einzelne VPN- oder Direct Connect-Verbindung möglicherweise nicht aus. In diesem Fall müssen Sie eine Hybridlösung entwickeln, um den Lastausgleich des Datenverkehrs zwischen mehreren Verbindungen zu ermöglichen. Direct Connect kann erforderlich sein, was aufgrund der privaten Netzwerkverbindung eine vorhersagbarere und konsistentere Leistung bietet. Es eignet sich hervorragend für Produktions-Workloads, für die eine konsistente Latenz und Jitter nahe null erforderlich sind.

AWS Direct Connect ermöglicht eine dedizierte Konnektivität mit der AWS-Umgebung im Bereich zwischen 50 Mbit/s und 10 Gbit/s. Auf diese Weise können Sie die Latenz verwalten und steuern und die erforderliche Bandbreite bereitstellen, um für Ihre Workload leistungsstarke Verbindungen mit anderen Umgebungen zu ermöglichen. Durch die Nutzung eines AWS Direct Connect-Partners erhalten Sie End-to-End-Konnektivität für mehrere Umgebungen und somit ein erweitertes Netzwerk mit konsistenter Leistung.

Das Site-to-Site-VPN von AWS ist ein verwalteter VPN-Service für VPCs. Wenn eine VPN-Verbindung erstellt wird, werden von AWS Tunnel zu zwei verschiedenen VPN-Endpunkten bereitgestellt. Mit AWS Transit Gateway können Sie die Konnektivität zwischen mehreren VPCs vereinfachen und auch eine Verbindung mit einem beliebigen VPC herstellen, der über eine einzelne VPN-Verbindung mit AWS Transit Gateway verknüpft ist. Darüber hinaus können Sie mit AWS Transit Gateway auch eine Skalierung über das IPsec-VPN-Durchsatzlimit von 1,25 Gbit/s hinaus durchführen, indem Sie die ECMP-Routingunterstützung (Equal Cost Multi-Path, ECMP) über mehrere VPN-Tunnel ermöglichen.

Grad des Aufwands für den Implementierungsplan: Der Grad des Aufwands ist hoch, um Workload-Bedürfnisse für hybride Netzwerke zu evaluieren und hybride Netzwerklösungen zu implementieren.

Ressourcen

Ähnliche Dokumente:

- Network Load Balancer

- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs
- Site-to-Site-VPN von AWS
- Bauen einer skalierbaren und sicheren Multi-VPC-AWS-Netzwerkinfrastruktur
- Direct Connect
- Client VPN

Ähnliche Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)
- AWS Global Accelerator
- Direct Connect
- Transit Gateway Connect
- VPN-Lösungen
- Sicherheit mit VPN-Lösungen

Ähnliche Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP04 Nutzen von Lastausgleich und Verschlüsselungsauslagerung

Verteilen Sie den Datenverkehr auf mehrere Ressourcen oder Services, um von der Elastizität der Cloud zu profitieren. Sie können den Lastausgleich auch nutzen, um die Terminierung von Verschlüsselung auszulagern. So lässt sich die Leistung optimieren und Datenverkehr effektiv weiterleiten.

Beim Implementieren einer Scale-Out-Architektur, wo Sie mehrere Instances für Serviceinhalte verwenden möchten, können Sie innerhalb Ihrer Amazon VPC Load Balancer nutzen. AWS bietet

mehrere Modelle für Ihre Anwendungen im ELB-Service. Application Load Balancer eignet sich am besten für die Lastverteilung von HTTP- und HTTPS-Datenverkehr und bietet erweitertes Routing von Anfragen, das auf die Lieferung von modernen Anwendungsarchitekturen abzielt, einschließlich Microservices und Container.

Network Load Balancer eignet sich optimale für die Lastverteilung von TCP-Datenverkehr, wenn eine hohe Leistung erforderlich ist. Hiermit lassen sich mit konstant geringer Latenz Millionen Anforderungen pro Sekunde und plötzliche Datenverkehrsspitzen oder schwankende Datenverkehrsmuster verarbeiten.

Elastic Load Balancing ermöglicht die integrierte Zertifikatverwaltung und SSL/TLS-Entschlüsselung. Auf diese Weise können Sie die SSL-Einstellungen des Load Balancers flexibel zentral verwalten und CPU-intensive Arbeitsschritte für Ihren Workload auslagern.

Gängige Antimuster:

- Sie leiten den gesamten Internetverkehr über vorhandene Load Balancer weiter.
- Sie nutzen einen generischen TCP-Lastausgleich und lassen die SSL-Verschlüsselung von den einzelnen Rechenknoten verarbeiten.

Vorteile der Einführung dieser bewährten Methode: Ein Load Balancer verarbeitet die variierende Last des Anwendungsdatenverkehrs in einer einzigen oder in mehreren Availability Zones. Load Balancer zeichnen sich durch die hohe Verfügbarkeit, die automatische Skalierung und die robuste Sicherheit aus, mit der Anwendungen fehlertolerant gestaltet werden können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Nutzen geeigneter Load Balancer für die Workload: Wählen Sie den geeigneten Load Balancer für Ihre Workload aus. Wenn Sie die Last von HTTP-Anfragen verteilen müssen, empfehlen wir einen Application Load Balancer. Für Netzwerk- und Transportprotokoll-Load-Balancing (Ebene 4 – TCP, UDP) sowie für Anwendungen mit höchster Leistung und geringer Latenz empfehlen wir den Network Load Balancer, Application Load Balancers-Support für HTTPS und Network-Load-Balancer-Support für TLS-Verschlüsselungs-Offloading.

Aktivieren der Auslagerung der HTTPS- oder TLS-Verschlüsselung: Elastic Load Balancing umfasst integrierte Zertifikatverwaltung, Benutzerauthentifizierung und SSL/TLS-Verschlüsselung. Es bietet flexible Möglichkeiten, TLS-Einstellungen zentral zu verwalten und CPU-intensive Workloads aus den

Anwendungen auszulagern. Verschlüsseln Sie den gesamten HTTPS-Datenverkehr im Rahmen der Load-Balancer-Bereitstellung.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS – Optimierte Instances
- Application Load Balancer
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances
- Network Load Balancer
- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs

Ähnliche Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)

Ähnliche Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP05 Auswählen leistungsfördernder Netzwerkprotokolle

Treffen Sie Entscheidungen über Protokolle für die Kommunikation zwischen Systemen und Netzwerken auf Grundlage der Auswirkungen, die sich für die Leistung der Workload ergeben.

In Bezug auf die Erzielung eines höheren Durchsatzes besteht eine Beziehung zwischen der Latenz und der Bandbreite. Wenn für Ihre Dateiübertragung TCP genutzt wird, führt eine höhere Latenz zu einer Reduzierung des allgemeinen Durchsatzes. Es gibt Möglichkeiten, dies per TCP-Optimierung und mit verbesserten Übertragungsprotokollen zu beheben. Bei einigen Ansätzen wird UDP verwendet.

Gängige Antimuster:

- Sie verwenden TCP unabhängig von den Leistungsanforderungen für alle Workloads.

Vorteile der Einführung dieser bewährten Methode: Die Auswahl des richtigen Protokolls für die Kommunikation zwischen Workload-Komponenten gewährleistet die bestmögliche Leistung für die jeweilige Workload. Das verbindungslose UDP ermöglicht zwar eine hohe Geschwindigkeit, bietet aber weder eine erneute Übertragung noch hohe Zuverlässigkeit. TCP ist ein Protokoll mit vollem Funktionsumfang, bringt jedoch einen größeren Overhead für die Verarbeitung der Pakete mit sich.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Optimieren des Netzwerkverkehrs: Wählen Sie das geeignete Protokoll aus, um die Leistung Ihrer Workload zu optimieren. In Bezug auf die Erzielung eines höheren Durchsatzes besteht eine Beziehung zwischen der Latenz und der Bandbreite. Wenn für die Dateiübertragung TCP genutzt wird, führt eine höhere Latenz zu einer Reduzierung des allgemeinen Durchsatzes. Es gibt Möglichkeiten, die Latenz per TCP-Optimierung und mit verbesserten Übertragungsprotokollen zu verkürzen. Bei einigen davon kommt UDP zum Einsatz.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS – Optimierte Instances
- Application Load Balancer
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances

- Network Load Balancer
- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs

Ähnliche Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)

Ähnliche Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP06 Auswählen des Workload-Standortes entsprechend den Netzwerkanforderungen

Verwenden Sie die verfügbaren Optionen für Cloud-Standorte, um die Netzwerklatenz zu verringern oder den Durchsatz zu verbessern. Verwenden Sie AWS-Regionen, Availability Zones, Platzierungsgruppen und Edge-Standorte, z. B. AWS Outposts, AWS Local Zones und AWS Wavelength, um eine Reduzierung der Netzwerklatenz bzw. eine Verbesserung des Durchsatzes zu erreichen.

Im Zentrum der AWS Cloud-Infrastruktur stehen Regionen und Availability Zones. Eine Region ist ein physischer Standort mit mehreren Availability Zones.

Availability Zones bestehen aus mindestens einem eigenständigen Rechenzentrum mit einer redundanten Stromversorgung, einem Netzwerk sowie Konnektivität. Sie sind jeweils in getrennten Einrichtungen untergebracht. Mithilfe der Availability Zones können Sie Produktionsanwendungen und Datenbanken betreiben, die verfügbarer, fehlertoleranter und skalierbarer als bei der Nutzung von nur einem Rechenzentrum sind.

Die Auswahl der passenden Regionen hängt von den folgenden wesentlichen Faktoren ab:

- Dem Standort Ihrer Benutzer: Je weniger weit die Region von den Benutzern Ihres Workloads entfernt ist, desto geringer ist die Latenz bei der Verwendung.
- Dem Standort Ihrer Daten: Bei datenintensiven Anwendungen entsteht der größte Latenzengpass bei der Datenübertragung. Anwendungscode sollte möglichst nah bei den Daten ausgeführt werden.
- Weitere Einschränkungen: Berücksichtigen Sie auch Einschränkungen wie die Sicherheit und Compliance.

Amazon EC2 verfügt über Platzierungsgruppen für das Netzwerk. Eine Platzierungsgruppe ist eine logische Gruppierung von Instances, um die Latenz zu verringern oder die Zuverlässigkeit zu erhöhen. Die Verwendung von Platzierungsgruppen mit unterstützten Instance-Typen und einem Elastic Network Adapter (ENA) ermöglicht die Verarbeitung von Workloads in einem Netzwerk mit 25 Gbit/s und geringer Latenz. Platzierungsgruppen werden für Workloads empfohlen, für die eine niedrige Netzwerklatenz bzw. ein hoher Durchsatz von Vorteil sind. Mit Platzierungsgruppen kann die Stabilität der Netzwerkkommunikation verbessert werden.

Services, bei denen eine geringe Latenz wichtig ist, werden am Edge mithilfe eines globalen Netzwerks aus Edge-Standorten bereitgestellt. Diese Edge-Standorte verfügen in der Regel über Services wie ein Content Delivery Network (CDN) und Domain Name System (DNS). Durch die Platzierung am Edge können die Workloads mit geringer Latenz auf Anforderungen zu Inhalten oder zur DNS-Auflösung reagieren. Es sind auch geografische Services wie das Geo-Targeting von Inhalten (Bereitstellung unterschiedlicher Inhalte gemäß dem Standort von Endbenutzern) oder die latenzbasierte Weiterleitung von Endbenutzern zur nächsten Region (minimale Latenz) verfügbar.

Amazon CloudFront ist ein globales CDN, mit dem sich sowohl statische Inhalte wie Bilder, Skripts und Videos als auch dynamische Inhalte wie APIs oder Webanwendungen beschleunigen lassen. Als Basis dient ein globales Netzwerk aus Edge-Standorten, an denen die Inhalte zwischengespeichert werden. Sie stellen eine leistungsfähige Netzwerkkonnektivität für Ihre Benutzer sicher. CloudFront beschleunigt auch zahlreiche weitere Funktionen wie das Hochladen von Inhalten und dynamische Anwendungen. Sie können damit die Leistung aller Anwendungen steigern, die Datenverkehr über das Internet verursachen. Lambda@Edge ist eine Funktion von Amazon CloudFront, mit der Sie Code näher an den Benutzern Ihres Workloads ausführen können, um die Leistung zu verbessern und die Latenz zu verringern.

Amazon Route 53 ist ein hochverfügbarer und skalierbarer Cloud-DNS-Webservice. Der Service ist für Entwickler und Unternehmen eine äußerst zuverlässige und kostengünstige Möglichkeit, Endbenutzer an Internetanwendungen weiterzuleiten. Hierzu werden Namen wie „www.beispiel.de“

in numerische IP-Adressen wie 192.168.2.1 übersetzt, die von Computern untereinander für den Verbindungsaufbau verwendet werden. Route 53 ist uneingeschränkt mit IPv6 kompatibel.

AWS Outposts wurde für Workloads entwickelt, die aufgrund von Latenzanforderungen lokal verarbeitet werden müssen und die Sie nahtlos mit Ihren restlichen Workloads in AWS ausführen möchten. Bei AWS Outposts handelt es sich um vollständig verwaltete und konfigurierbare Datenverarbeitungs- und Speicher-Racks, die auf von AWS entwickelter Hardware basieren. Hiermit können Sie die Datenverarbeitung und Speicherung lokal durchführen und gleichzeitig nahtlose Verbindungen mit den vielen Services von AWS in der Cloud herstellen.

AWS Local Zones wurde für die Ausführung von Workloads entwickelt, für die eine Latenz im einstelligen Millisekundenbereich benötigt wird, z. B. Video-Rendering und virtuelle Desktop-Anwendungen mit hohen Grafikanforderungen. Mit Local Zones können Sie von allen Vorteilen profitieren, die sich durch die Platzierung der Datenverarbeitungs- und Speicherressourcen in der Nähe Ihrer Endbenutzer ergeben.

AWS Wavelength wurde für die Bereitstellung von Anwendungen mit extrem niedriger Latenz für 5G-Geräte entwickelt, indem die Infrastruktur, Services, APIs und Tools von AWS auf 5G-Netze erweitert wurden. Bei Wavelength wird die Speicherung und Datenverarbeitung in die 5G-Netze von Telekommunikationsanbietern eingebettet, um die Verarbeitung Ihrer 5G-Workload zu verbessern, wenn dafür eine Latenz im einstelligen Millisekundenbereich erforderlich ist. Beispiele hierfür sind IoT-Geräte, Game-Streaming, autonomes Fahren und die Produktion von Live-Medien.

Verwenden Sie Edge-Services, um die Latenz zu reduzieren und das Caching von Inhalten zu ermöglichen. Stellen Sie sicher, dass Sie die Cache-Steuerung für DNS und HTTP/HTTPS richtig konfiguriert haben, um aus diesen Ansätzen den größtmöglichen Nutzen zu ziehen.

Gängige Antimuster:

- Sie konsolidieren alle Workload-Ressourcen an einem geografischen Standort.
- Sie wählen die Region aus, die sich Ihrem Standort, aber nicht dem Workload-Endbenutzer am nächsten befindet.

Vorteile der Einführung dieser bewährten Methode: Sie müssen sicherstellen, dass Ihr Netzwerk überall dort verfügbar ist, wo Sie Kunden erreichen möchten. Durch Einsatz des privaten weltweiten Netzwerkes von AWS wird die kürzestmögliche Latenz für Ihre Kunden gewährleistet. Dazu werden die Workloads an den Orten bereitgestellt, die den Kunden am nächsten liegen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Reduzieren der Latenz durch Auswahl der richtigen Standorte: Geben Sie an, wo sich die Benutzer und Daten befinden. Nutzen Sie AWS-Regionen, Availability Zones, Platzierungsgruppen und Edge-Standorte, um Latenz zu reduzieren.

Ressourcen

Zugehörige Dokumente:

- Amazon EBS – Optimierte Instances
- Application Load Balancer
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances
- Network Load Balancer
- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs

Relevante Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)

Zugehörige Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken

PERF05-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken

Treffen Sie anhand der erfassten und analysierten Daten fundierte Entscheidungen zum Optimieren Ihrer Netzwerkkonfiguration. Messen Sie die Auswirkungen dieser Änderungen und treffen Sie künftige Entscheidungen auf Grundlage dieser Ergebnisse.

Aktivieren Sie VPC-Flussprotokolle (VPC Flow Logs) für alle VPC-Netzwerke, die von Ihrem Workload verwendet werden. VPC Flow Logs sind eine Funktion, mit der Sie Informationen zum ein- und ausgehenden IP-Datenverkehr an den Netzwerkschnittstellen Ihrer VPC erfassen können. VPC Flow Logs dient Ihnen als Hilfe bei verschiedenen Aufgaben, z. B. bei der Fehlerbehebung, wenn Datenverkehr eine Instance nicht erreicht. Dies ist wiederum beim Diagnostizieren von zu strikten Sicherheitsgruppenregeln hilfreich. Sie können Flow Logs als Sicherheitstool zum Überwachen des bei Ihrer Instance eingehenden Datenverkehrs, zum Erstellen von Profilen des Netzwerkverkehrs und zum Ermitteln von anomalem Verhalten im Datenverkehr verwenden.

Verwenden Sie Netzwerkmetriken, um Änderungen an der Netzwerkkonfiguration vorzunehmen, wenn sich der Workload ändert. Da Cloud-basierte Netzwerke schnell geändert werden können, müssen Sie Ihre Netzwerkarchitektur im Laufe der Zeit weiterentwickeln, um weiterhin eine effiziente Leistung zu erzielen.

Gängige Antimuster:

- Sie gehen davon aus, dass alle leistungsbezogenen Probleme auf Anwendungen zurückzuführen sind.
- Sie testen die Netzwerkleistung ausschließlich an einem Standort nahe der Stelle, an der Sie die Workload bereitgestellt haben.

Vorteile der Einführung dieser bewährten Methode:Damit Sie die für die Workload erforderlichen Metriken tatsächlich erfüllen, müssen Sie die Netzwerk-Leistungsmetriken überwachen. Sie können Informationen über den IP-Datenverkehr erfassen, der über Netzwerkschnittstellen in der VPC ein- und ausgeht. Anhand dieser Daten können Sie dann neue Optimierungen hinzufügen oder die Workload in neuen geografischen Regionen bereitstellen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Aktivieren von VPC Flow Logs: Mit VPC Flow Logs können Sie Informationen über den IP-Datenverkehr erfassen, der über die Netzwerkschnittstellen in Ihrer VPC ein- und ausgeht. VPC Flow Logs unterstützt Sie bei verschiedenen Aufgaben, z. B. bei der Fehlerbehebung, wenn Datenverkehr

eine Instance nicht erreicht. Das kann Ihnen beim Diagnostizieren von zu strikten Regeln für Sicherheitsgruppen helfen. Sie können Flow Logs als Sicherheitstool zum Überwachen des bei Ihrer Instance eingehenden Datenverkehrs, zum Erstellen von Profilen des Netzwerkverkehrs und zum Ermitteln von anomalem Verhalten im Datenverkehr verwenden.

Aktivieren der geeigneten Metriken für Netzwerkoptionen: Wählen Sie unbedingt die geeigneten Netzwerkmetriken für Ihre Workload aus. Sie können Metriken für VPC-NAT-Gateways, Transit-Gateways und VPN-Tunnel aktivieren.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS – Optimierte Instances
- Application Load Balancer
- EC2: Enhanced Networking unter Linux
- EC2: Enhanced Networking unter Windows
- EC2: Platzierungsgruppen
- Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter (ENA) in Linux-Instances
- Network Load Balancer
- Netzwerkprodukte mit AWS
- Transit Gateway
- Umstellung auf latenzbasiertes Routing in Amazon Route 53
- VPC-Endpunkte
- VPC Flow Logs
- Monitoring your global and core networks with Amazon Cloudwatch metrics (Überwachen von globalen und Kernnetzwerken mit Amazon-Cloudwatch-Metriken)
- Continuously monitor network traffic and resources (Kontinuierliches Überwachen von Netzwerkdatenverkehr und -ressourcen)

Ähnliche Videos:

- Konnektivität mit AWS und AWS-Hybrid-Netzwerkarchitekturen (NET317-R1)
- Optimieren der Netzwerkleistung für Amazon EC2-Instances (CMP308-R1)

- Monitoring and troubleshooting network traffic (Überwachen des Netzwerkdatenverkehrs und Fehlerbehebung)
- Simplify Traffic Monitoring and Visibility with Amazon VPC Traffic Mirroring (Vereinfachen der Datenverkehrsüberwachung und Sichtbarkeit mit Amazon VPC Traffic Mirroring)

Ähnliche Beispiele:

- AWS Transit Gateway und skalierbare Sicherheitslösungen
- Workshops zu AWS-Netzwerken
- Überwachung des AWS-Netzwerks

Prüfen Sie die Angaben.

Frage

- LEIST 6 Wie profitiert Ihr Workload von neuen Releases?

LEIST 6 Wie profitiert Ihr Workload von neuen Releases?

Bei der Architektur von Workloads sind die Wahlmöglichkeiten begrenzt. Im Laufe der Zeit werden jedoch immer wieder neue Technologien und Ansätze zur Leistungsoptimierung von Workloads entwickelt.

Bewährte Methoden

- PERF06-BP01 Erhalten aktueller Informationen zu neuen Ressourcen und Services
- PERF06-BP02 Definieren eines Prozesses zum Verbessern der Workload-Leistung
- PERF06-BP03 Allmähliches Anpassen der Workload-Leistung

PERF06-BP01 Erhalten aktueller Informationen zu neuen Ressourcen und Services

Evaluieren Sie Möglichkeiten zur Verbesserung der Leistung, wenn neue Services, Entwurfsmuster und Produktangebote verfügbar sind. Ermitteln Sie anhand von Bewertungen, internen Diskussionen oder externen Analysen, wie sich diese neuen Optionen positiv auf die Leistung oder Effizienz der Workload auswirken können.

Definieren Sie einen Prozess zum Bewerten von Updates, neuen Funktionen und Services, die für Ihren Workload relevant sind. Erstellen Sie beispielsweise Machbarkeitsstudien, die auf

neuen Technologien aufbauen, oder beraten Sie sich mit einer internen Gruppe. Führen Sie beim Ausprobieren neuer Ideen oder Services Leistungstests durch, um die Auswirkungen auf die Leistung des Workloads zu messen. Nutzen Sie Infrastructure as Code (IaC) und eine DevOps-Kultur, um neue Ideen oder Technologien häufig bei minimalen Kosten und Risiken zu testen.

Gewünschtes Ergebnis: Sie haben das Inventar der Komponenten, Ihr Entwurfsmuster und die Eigenschaften Ihres Workloads dokumentiert. Anhand dieser Dokumentation erstellen Sie eine Liste von Abonnements zur Benachrichtigung Ihres Teams über Service-Updates, Funktionen und neue Produkte. Sie haben Komponentenbeteiligte identifiziert, die die neuen Versionen evaluieren und eine Empfehlung für geschäftliche Auswirkungen und Prioritäten geben werden.

Typische Anti-Muster:

- Sie überprüfen neue Optionen und Services nur dann, wenn Ihr Workload nicht Ihren Leistungsanforderungen entspricht.
- Sie gehen davon aus, dass alle neuen Produktangebote für Ihren Workload nicht nützlich sind.
- Sie entscheiden sich bei Verbesserungen des Workloads immer für die eigene Erstellung gegenüber dem Kauf.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie neue Services oder Produktangebote in Betracht ziehen, können Sie die Leistung und die Effizienz Ihres Workloads verbessern, die Kosten für Ihre Infrastruktur senken und den Aufwand für die Verwaltung Ihrer Services verringern.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Definieren Sie einen Prozess zum Bewerten von Updates, neuen Funktionen und Services von AWS. Erstellen Sie beispielsweise Machbarkeitsstudien, die auf neuen Technologien aufbauen. Führen Sie beim Ausprobieren neuer Ideen oder Services Leistungstests durch, um die Auswirkungen auf die Effizienz oder Leistung des Workloads zu messen. Dank der flexiblen Möglichkeiten innerhalb von AWS können Sie regelmäßig neue Ideen oder Technologien testen und dabei Kosten und Risiken auf ein Minimum reduzieren.

Implementierungsschritte

1. Dokumentieren Sie Ihre Workload-Lösungen. Verwenden Sie Ihre Configuration Management Database (CMDB)-Lösung zur Dokumentation Ihres Inventars und zur Kategorisierung Ihrer

Services und Abhängigkeiten. Verwenden Sie Tools wie AWS Config zum Erstellen einer Liste aller Services in AWS, die von Ihrem Workload genutzt werden.

2. Wenden Sie eine Markierungsstrategie an, um die Eigentümer für alle Workload-Komponenten und -kategorien zu dokumentieren. Wenn Sie beispielsweise derzeit Amazon RDS als Datenbanklösung verwenden, weisen Sie Ihren Datenbankadministrator (DBA) als Eigentümer für die Evaluierung und Untersuchung neuer Services und Updates zu und dokumentieren Sie dies.
3. Identifizieren Sie Quellen für Neuigkeiten und Updates im Zusammenhang mit Ihren Workload-Komponenten. Im vorher erwähnten Beispiel zu Amazon RDS sollte der Kategorieeigentümer den Blog Neuigkeiten im AWS-Blog für die Produkte abonnieren, die seiner Workload-Komponente entsprechen. Sie können den RSS-Feed abonnieren oder Ihre E-Mail-Abonnements verwalten. Überwachen Sie Upgrades der von Ihnen verwendeten Amazon RDS-Datenbank, neue Funktionen, veröffentlichte Instances und neue Produkte wie Amazon Aurora Serverless. Überwachen Sie Branchenblogs, Produkte und Anbieter, die für Ihre Komponenten wichtig sind.
4. Dokumentieren Sie Ihren Prozess zur Evakuierung von Aktualisierungen und neuen Services. Geben Sie Ihren Kategorieeigentümern ausreichend Zeit und Raum zum Forschen, Testen, Experimentieren und zur Validierung von Aktualisierungen und neuen Services. Nutzen Sie die dokumentierten geschäftlichen Anforderungen und KPIs, um zu ermitteln, welche Aktualisierungen positive geschäftliche Auswirkungen haben werden.

Aufwand für den Implementierungsplan: Zur Einrichtung dieser bewährten Methode müssen Sie die derzeitigen Komponenten Ihres Workloads kennen sowie Kategorieeigentümer und Quellen für Serviceaktualisierungen identifizieren. Der Aufwand dafür ist anfangs gering, der Vorgang wird sich aber mit der Zeit deutlich weiterentwickeln.

Ressourcen

Zugehörige Dokumente:

- AWS-Blog
- Neuerungen bei AWS

Zugehörige Videos:

- YouTube-Kanal: AWS Events
- YouTube-Kanal: AWS Online Tech Talks
- YouTube-Kanal: Amazon Web Services

Zugehörige Beispiele:

- AWS Github
- AWS Skill Builder

PERF06-BP02 Definieren eines Prozesses zum Verbessern der Workload-Leistung

Definieren Sie einen Prozess, mit dem sich neu verfügbare Services, Designmuster, Ressourcentypen und Konfigurationen bewerten lassen. Führen Sie beispielsweise vorhandene Leistungstests für neue Instance-Angebote durch, um zu ermitteln, welche Verbesserungen sich für Ihre Workload ergeben.

Für Ihren Workload gibt es einige wesentliche Einschränkungen. Dokumentieren Sie diese, damit Sie besser einschätzen können, durch welche Art von Innovation die Leistung Ihres Workloads gesteigert werden könnte. Ziehen Sie diese Informationen heran, wenn Sie von neuen verfügbaren Services oder Technologien erfahren, um Möglichkeiten zur Beseitigung von Einschränkungen oder Engpässen zu identifizieren.

Gängige Antimuster:

- Sie gehen davon aus, dass Ihre aktuelle Architektur unverändert bleibt und im Laufe der Zeit nicht aktualisiert wird.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie begründen.

Vorteile der Einführung dieser bewährten Methode: Durch einen definierten Prozess zum Ändern der Architektur erhalten Sie die Möglichkeit, die gesammelten Daten langfristig in die Gestaltung Ihrer Workload einfließen zu lassen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Identifizieren wesentlicher Leistungseinschränkungen für Ihre Workload: Dokumentieren Sie die Leistungseinschränkungen Ihrer Workload, damit Sie besser einschätzen können, durch welche Art von Innovation die Leistung Ihrer Workload gegebenenfalls gesteigert werden kann.

Ressourcen

Ähnliche Dokumente:

- AWS-Blog
- Neuerungen bei AWS

Ähnliche Videos:

- AWS-Veranstaltungen: YouTube-Kanal
- AWS Online Tech Talks: YouTube-Kanal
- Amazon Web Services: YouTube-Kanal

Ähnliche Beispiele:

- AWS Github
- AWS Skill Builder

PERF06-BP03 Allmähliches Anpassen der Workload-Leistung

Nutzen Sie als Organisation die aus dem Evaluierungsprozess gewonnenen Informationen, um aktiv die frühzeitige Einführung neuer Services oder Ressourcen zu fördern, sobald diese zur Verfügung gestellt werden.

Nutzen Sie die Erkenntnisse, die Sie beim Bewerten neuer Services oder Technologien gewinnen, um Veränderungen auf den Weg zu bringen. Zusammen mit Ihrem Unternehmen oder Ihres Workloads verändern sich auch die Leistungsanforderungen. Nutzen Sie die aus den Workload-Metriken generierten Daten, um diejenigen Bereiche zu identifizieren, die das größte Potenzial für Effizienz- oder Leistungssteigerungen bieten. Führen Sie proaktiv neue Services und Technologien ein, um der Nachfrage gerecht zu werden.

Gängige Antimuster:

- Sie gehen davon aus, dass Ihre aktuelle Architektur unverändert bleibt und im Laufe der Zeit nicht aktualisiert wird.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie begründen.
- Sie ändern die Architektur nur, weil alle anderen in der Branche sie verwenden.

Vorteile der Einführung dieser bewährten Methode: Um Ihre Workloadleistung und -kosten zu optimieren, müssen Sie alle verfügbaren Software und Services auswerten, um die geeigneten für Ihre Workload zu bestimmen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Workload allmählich weiterentwickeln: Nutzen Sie die Erkenntnisse, die Sie beim Evaluieren neuer Services oder Technologien gewinnen, um Veränderungen auf den Weg zu bringen. Zusammen mit Ihrem Unternehmen bzw. Ihrer Workload verändern sich auch die Leistungsanforderungen. Nutzen Sie die aus den Workload-Metriken generierten Daten, um diejenigen Bereiche zu identifizieren, die das größte Potenzial für Effizienz- oder Leistungssteigerungen bieten. Führen Sie proaktiv neue Services und Technologien ein, um der Nachfrage gerecht zu werden.

Ressourcen

Ähnliche Dokumente:

- AWS-Blog
- Neuerungen bei AWS

Ähnliche Videos:

- AWS-Veranstaltungen: YouTube-Kanal
- AWS Online Tech Talks: YouTube-Kanal
- Amazon Web Services: YouTube-Kanal

Ähnliche Beispiele:

- AWS Github
- AWS Skill Builder

Überwachung

Frage

- LEIST 7 Wie lassen sich Ressourcen überwachen, um sicherzustellen, dass sie funktionieren?

LEIST 7 Wie lassen sich Ressourcen überwachen, um sicherzustellen, dass sie funktionieren?

Die Systemleistung kann sich mit der Zeit verschlechtern. Überwachen Sie die Systemleistung, um eine Verschlechterung frühzeitig zu erkennen und ihr entgegenzuwirken, etwa indem Sie interne oder externe Faktoren wie das Betriebssystem oder die Anwendungslast korrigieren.

Bewährte Methoden

- PERF07-BP01 Erfassen von Leistungsmetriken
- PERF07-BP02 Analysieren Sie Metriken bei Eintreten von Ereignissen oder Vorfällen
- PERF07-BP03 Legen Sie wichtige Leistungskennzahlen (KPIs) zum Messen der Workload-Leistung fest
- PERF07-BP04 Generieren alarmbasierter Benachrichtigungen per Überwachungssystem
- PERF07-BP05 Regelmäßiges Überprüfen von Metriken
- PERF07-BP06 Proaktives Überwachen und Benachrichtigen

PERF07-BP01 Erfassen von Leistungsmetriken

Verwenden Sie einen Überwachungs- und Beobachtungs-Service, um leistungsbezogene Metriken aufzuzeichnen. Metriken umfassen beispielsweise Datenbanktransaktionen, langsame Abfragen, I/O-Latenz, den Durchsatz von HTTP-Anforderungen, Servicelatenz und andere wichtige Daten.

Identifizieren Sie die für Ihren Workload relevanten Leistungskennzahlen und erfassen Sie sie. Diese Daten sind von wesentlicher Bedeutung, um festzustellen, welche Komponenten sich auf die Gesamtleistung und Effizienz Ihrer Workload auswirken.

Ermitteln Sie anhand des Kundenerlebnisses, auf welche Kennzahlen es ankommt. Identifizieren Sie für jede Kennzahl Ziel, Messverfahren und Priorität. Konfigurieren Sie darauf aufbauend Alarme und Benachrichtigungen, die eine proaktive Behandlung von Leistungsproblemen ermöglichen.

Gängige Antimuster:

- Sie überwachen nur Metriken auf Betriebssystemebene, um Einblicke in Ihre Workload zu erhalten.
- Sie legen Ihre Rechenbedürfnisse auf Workload-Anforderungen zu Spitzenzeiten aus.

Vorteile der Einführung dieser bewährten Methode: Um Leistung und Ressourcenauslastung zu optimieren, benötigen Sie einen Gesamtüberblick über Ihre wichtigsten Leistungsindikatoren. Sie

können Dashboards erstellen und Metrikberechnungen für Ihre Daten durchführen, um Einblicke in Betrieb und Nutzung zu erhalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Identifizieren Sie die für Ihre Workload relevanten Leistungsmetriken und erfassen Sie sie. Anhand dieser Daten können Sie feststellen, welche Komponenten sich auf die Gesamtleistung oder Effizienz Ihrer Workload auswirken.

Leistungsmetriken identifizieren: Ermitteln Sie anhand der Kundenerfahrungen die wichtigsten Metriken. Identifizieren Sie für jede Kennzahl Ziel, Messverfahren und Priorität. Nutzen Sie diese Datenpunkte, um Alarme und Benachrichtigungen zu konfigurieren, die eine proaktive Behandlung von Leistungsproblemen ermöglichen.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und On-Premises-Servern mit dem CloudWatch Agent
- Veröffentlichen von benutzerdefinierten Metriken
- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation
- Amazon CloudWatch RUM

Ähnliche Videos:

- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Verwaltung der Anwendungsleistung in AWS
- Erstellen eines Überwachungsplans

Ähnliche Beispiele:

- Level 100: Monitoring with CloudWatch Dashboards (Stufe 100: Überwachung mit Cloudwatch-Dashboards)

- Level 100: Monitoring Windows EC2 instance with CloudWatch Dashboards (Stufe 100: Überwachung einer Windows-EC2-Instance mit Cloudwatch-Dashboards)
- Level 100: Monitoring an Amazon Linux EC2 instance with CloudWatch Dashboards (Stufe 100: Überwachung einer Amazon-Linux-EC2-Instance mit Cloudwatch-Dashboards)

PERF07-BP02 Analysieren Sie Metriken bei Eintreten von Ereignissen oder Vorfällen

Ziehen Sie während eines Ereignisses oder Vorfalls oder als Reaktion darauf Überwachungs-Dashboards oder Berichte heran, um die Auswirkungen nachzuvollziehen und zu diagnostizieren. Diese Ansichten bieten Einblick in die Bereiche der Workload, die nicht die erwartete Leistung liefern.

Berücksichtigen Sie beim Beschreiben kritischer Benutzerszenarien für Ihre Architektur die Leistungsanforderungen. Geben Sie beispielsweise an, wie schnell die einzelnen kritischen Szenarien ausgeführt werden sollen. Implementieren Sie zusätzliche skriptbasierte Benutzerreisen in diese Szenarien, damit Sie genau wissen, wie sich die Leistung dieser Szenarien im Vergleich zu Ihren Anforderungen verhält.

Gängige Antimuster:

- Sie gehen davon aus, dass Leistungsereignisse einmalige Probleme sind und sich nur auf Anomalien beziehen.
- Vorhandene Leistungsmetriken werden nur ausgewertet, wenn Sie auf Leistungsereignisse reagieren.

Vorteile der Einführung dieser bewährten Methode: Um festzustellen, ob Ihre Workload auf erwartetem Niveau ausgeführt wird, müssen Sie auf Leistungsereignisse reagieren, indem Sie zusätzliche Metrikdaten für die Analyse erfassen. Diese Daten werden verwendet, um die Auswirkungen des Performance-Ereignisses zu verstehen und Änderungen zur Verbesserung der Workload-Leistung vorzuschlagen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Negativen Erlebnissen Priorität einräumen und kritische Benutzerszenarien beschreiben: Berücksichtigen Sie beim Beschreiben kritischer Benutzerszenarien für Ihre Architektur die Leistungsanforderungen. Geben Sie beispielsweise an, wie schnell die einzelnen kritischen Szenarien ausgeführt werden sollen. Implementieren Sie zusätzliche skriptbasierte Benutzerreisen

in diese kritischen Szenarien, damit Sie genau wissen, wie sich deren Leistung im Vergleich zu Ihren Anforderungen verhält.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Amazon CloudWatch Synthetics
- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation

Ähnliche Videos:

- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)
- Demo von Amazon CloudWatch Synthetics

Ähnliche Beispiele:

- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client

PERF07-BP03 Legen Sie wichtige Leistungskennzahlen (KPIs) zum Messen der Workload-Leistung fest

Identifizieren Sie die KPIs, die die Workload-Leistung quantitativ und qualitativ messen. Mithilfe von KPIs können Sie die Integrität einer Workload im Verhältnis zu einem Geschäftsziel messen. KPIs helfen dabei, Business- und Entwicklungsteams die Messung von Zielen und Strategien abzustimmen und wie diese gemeinsam zu Geschäftsergebnissen beitragen. KPIs sollten erneut aufgegriffen werden, wenn sich Geschäftsziele, Strategien oder Anforderungen von Endbenutzern ändern.

Beispielsweise könnte eine Website-Workload die Ladezeit der Seite als Indikator für die Gesamtleitstung heranziehen. Diese Metrik wäre einer von mehreren Datenpunkten, die ein Endbenutzererlebnis messen. Zusätzlich zum Ermitteln der Grenzwerte für Seitenladezeiten sollten

Sie das gewünschte Resultat dokumentieren bzw. das Geschäftsrisiko, wenn die Leistung nicht erreicht wird. Die lange Ladezeit einer Seite würde Ihre Endbenutzer direkt betreffen, die Bewertung ihres Benutzererlebnisses verringern und könnte zu einem Verlust von Kunden führen. Kombinieren Sie beim Definieren Ihrer KPI-Grenzwerte die Benchmarks der Branche und die Erwartungen Ihrer Endbenutzer. Beispielsweise, wenn die aktuelle Benchmark der Branche das Laden einer Webseite innerhalb von zwei Sekunden ist, Ihre Endbenutzer aber erwarten, dass eine Webseite innerhalb von einer Minute geladen wird, sollten Sie beim Einrichten des KPI beide Datenpunkte in Betracht ziehen. Ein weiteres Beispiel für eine KPI könnte der Fokus auf das Erfüllen von internen Leistungsanforderungen sein. Ein KPI-Grenzwert kann beim Erstellen von Vertriebsberichten innerhalb eines Tages, nachdem die Produktionsdaten erstellt wurden, eingerichtet werden. Diese Berichte beeinflussen möglicherweise direkt tägliche Entscheidungen und Geschäftsergebnisse.

Gewünschtes Ergebnis: Das Einführen von KPIs umfasst unterschiedliche Abteilungen und Stakeholder. Ihr Team muss Ihre Workload-KPIs mithilfe von detaillierten Echtzeitdaten und historischen Daten als Referenz evaluieren und Dashboards erstellen, die Metrikberechnungen für Ihre KPI-Daten durchführen, um Einblicke in Betrieb und Auslastung zu erhalten. KPIs sollten dokumentiert werden, sodass die vereinbarten KPIs und Grenzwerte, die Geschäftsziele und -strategien unterstützen, erklärt werden und den Metriken zugeordnet sind, die überwacht werden. Die KPIs identifizieren Leistungsanforderungen, werden absichtlich überprüft und häufig mit allen Teams geteilt und besprochen. Risiken und Kompromisse werden klar erkannt und es ist ersichtlich, wie das Geschäft beeinträchtigt wird, wenn KPI-Grenzwerte nicht erreicht werden.

Gängige Antimuster:

- Sie überwachen nur Metriken auf Systemebene, um Erkenntnisse über Ihre Workload zu gewinnen, und verstehen den geschäftlichen Einfluss dieser Metriken nicht.
- Sie gehen davon aus, dass Ihre KPIs bereits als standardmäßige Metrikdaten veröffentlicht und geteilt werden.
- Sie definieren KPIs, teilen Sie aber nicht mit allen Teams.
- Sie definieren keinen quantitativen, messbaren KPI.
- Sie richten KPIs nicht an Geschäftszielen oder -strategien aus.

Vorteile der Einführung dieser bewährten Methode: Das Identifizieren von bestimmten Metriken, die die Workload-Integrität darstellen, helfen Teams dabei, sich an ihren Prioritäten auszurichten und Geschäftsergebnisse erfolgreich zu definieren. Das Teilen dieser Metriken mit allen Abteilungen bietet Sichtbarkeit und die Ausrichtung an Grenzwerten, Erwartungen und Geschäftsauswirkungen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Alle Abteilungen und Geschäftsteams, die von der Integrität der Workload betroffen sind, sollten an der Definition der KPIs mitwirken. Eine einzelne Person sollte für die Zusammenarbeit, Zeitpläne, Dokumentation und Informationen in Bezug auf die KPIs eines Unternehmens zuständig sein. Dieser einzelne Eigentümer teilt häufig die Geschäftsziele und -strategien mit und weist Business-Stakeholdern Aufgaben zu, um KPIs in deren jeweiligen Abteilungen zu erstellen. Sobald KPIs definiert wurden, hilft das dem Betriebsteam oft beim Festlegen der Metriken, die in den Erfolg von unterschiedlichen KPIs einfließen und ihn unterstützen. KPIs sind nur dann wirksam, wenn sich alle Teammitglieder, die eine Workload unterstützten, der KPIs bewusst sind.

Implementierungsschritte

1. Identifizieren und dokumentieren Sie Business-Stakeholder.
2. Identifizieren Sie Unternehmensziele und -strategien.
3. Überprüfen Sie in der Branche gängige KPIs, die zu den Zielen und Strategien Ihres Unternehmens passen.
4. Überprüfen Sie die Erwartungen von Endbenutzern an Ihre Workload.
5. Definieren und dokumentieren Sie KPIs, die Ihre Unternehmensziele und -strategien unterstützen.
6. Identifizieren und dokumentieren Sie Kompromissstrategien zum Erreichen der KPIs.
7. Identifizieren und dokumentieren Sie Metriken, die in die KPIs einfließen.
8. Identifizieren und dokumentieren Sie KPI-Schwellenwerte für Schweregrad oder Alarmebene.
9. Identifizieren und dokumentieren Sie das Risiko und die Auswirkungen, wenn die KPIs nicht erreicht werden.
10. Identifizieren Sie die Überprüfungshäufigkeit pro KPI.
11. Kommunizieren Sie die KPI-Dokumentation allen Teams, die die Workload unterstützen.

Grad des Aufwands für den Implementierungsplan: Das Definieren und Kommunizieren von KPIs stellt einen niedrigen Arbeitsaufwand dar. Dies erfolgt üblicherweise innerhalb von einigen Wochen durch Treffen mit Stakeholdern und dem Überprüfen von Zielen, Strategien und Workload-Metriken.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation
- Verwendung von Amazon CloudWatch-Dashboards
- Amazon QuickSight-KPIs

Ähnliche Videos:

- AWS re:Invent 2019: Erweitern Sie den Umfang auf Ihre ersten 10 Millionen Benutzer (ARC211)
- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Erstellen eines Überwachungsplans

Ähnliche Beispiele:

- Erstellen eines Dashboards mit Amazon QuickSight

PERF07-BP04 Generieren alarmbasierter Benachrichtigungen per Überwachungssystem

Verwenden Sie basierend auf den von Ihnen definierten leistungsbezogenen wichtigen Kennzahlen (KPIs) ein Überwachungssystem, bei dem Alarme automatisch generiert werden, wenn sich die Messwerte außerhalb der erwarteten Grenzen bewegen.

Mit Amazon CloudWatch lassen sich Kennzahlen aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können auch benutzerdefinierte Kennzahlen erfassen und in Oberflächen-, Geschäfts- oder abgeleiteten Kennzahlen veröffentlichen. Legen Sie mit CloudWatch oder einem Überwachungsservice eines Drittanbieters Alarme fest, die bei Überschreitung bestimmter Schwellenwerte ausgelöst werden – mit einem solchen Alarm wird darauf hingewiesen, dass sich eine Metrik außerhalb des erwarteten Bereichs befindet.

Gängige Antimuster:

- Sie verlassen sich darauf, dass die Mitarbeiter Metriken überwachen und reagieren, wenn ein Problem auftritt.
- Sie verlassen sich ausschließlich auf betriebsbereite Runbooks, wenn Serverless-Workflows ausgelöst werden könnten, um dieselbe Aufgabe zu erledigen.

Vorteile der Einführung dieser bewährten Methode: Sie können Alarme festlegen und Aktionen basierend auf vordefinierten Schwellenwerten oder Algorithmen für Machine Learning automatisieren, die anormales Verhalten in Ihren Metriken identifizieren. Dieselben Alarme können auch Serverless-Workflows auslösen, die Leistungsmerkmale Ihrer Workload ändern können (z. B. Erhöhung der Rechenkapazität, Änderung der Datenbankkonfiguration).

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Überwachen von Metriken: Mithilfe von Amazon CloudWatch lassen sich Kennzahlen aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können benutzerdefinierte Metriken erfassen und veröffentlichen, um geschäftliche oder abgeleitete Metriken zu ermitteln. Richten Sie mit CloudWatch oder Überwachungsservices von Drittanbietern Alarme ein, die auf das Überschreiten von Schwellenwerten hinweisen.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation
- Verwendung von Alarmen und Alarmaktionen in CloudWatch

Ähnliche Videos:

- AWS re:Invent 2019: Erweitern Sie den Umfang auf Ihre ersten 10 Millionen Benutzer (ARC211)
- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Erstellen eines Überwachungsplans
- Verwenden von AWS Lambda mit Amazon CloudWatch Events

Ähnliche Beispiele:

- Cloudwatch-Protokolle: Konfigurieren von Alarmen

PERF07-BP05 Regelmäßiges Überprüfen von Metriken

Überprüfen Sie als routinemäßige Wartungsmaßnahme oder als Reaktion auf Ereignisse oder Vorfälle, welche Kennzahlen erfasst werden. Ermitteln Sie anhand dieser Überprüfung, welche Metriken für die Behebung von Problemen wesentlich waren und welche zusätzlichen Kennzahlen hilfreich wären, um Probleme zu identifizieren, zu beheben oder zu verhindern.

Bewerten Sie beim Reagieren auf Vorfälle oder Ereignisse diejenigen Kennzahlen, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Kennzahlen förderlich sein könnten. Verbessern Sie auf diese Weise die Qualität der erfassten Metriken, damit Sie zukünftige Probleme verhindern oder schneller beheben können.

Gängige Antimuster:

- Sie lassen zu, dass Metriken für einen längeren Zeitraum im Alarmstatus bleiben.
- Sie erstellen Alarme, die von einem Automatisierungssystem nicht umsetzbar sind.

Vorteile der Einführung dieser bewährten Methode: Überprüfen Sie kontinuierlich Metriken, die erfasst werden, um sicherzustellen, dass sie Probleme ordnungsgemäß identifizieren, beheben oder verhindern. Metriken können auch veralten, wenn sie für einen längeren Zeitraum im Alarmstatus bleiben.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Erfassung und Überwachung von Kennzahlen kontinuierlich verbessern: Bewerten Sie beim Reagieren auf Vorfälle oder Ereignisse diejenigen Kennzahlen, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Kennzahlen förderlich sein könnten. Verbessern Sie auf diese Weise die Qualität der erfassten Metriken, damit Sie zukünftige Probleme verhindern oder schneller beheben können.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und On-Premises-Servern mit dem CloudWatch Agent

- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation

Ähnliche Videos:

- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Verwaltung der Anwendungsleistung in AWS
- Erstellen eines Überwachungsplans

Ähnliche Beispiele:

- Erstellen eines Dashboards mit Amazon QuickSight
- Level 100: Monitoring with CloudWatch Dashboards (Stufe 100: Überwachung mit Cloudwatch-Dashboards)

PERF07-BP06 Proaktives Überwachen und Benachrichtigen

Verwenden Sie wichtige Leistungskennzahlen (KPIs) in Kombination mit Überwachungs- und Warnsystemen, um eine proaktive Behandlung leistungsbezogener Probleme zu ermöglichen. Verwenden Sie Alarme, um automatisierte Aktionen auszulösen und auf diese Weise Probleme nach Möglichkeit zu beheben. Leiten Sie den Alarm an die Personen weiter, die die richtigen Maßnahmen einleiten können, falls keine automatisierte Reaktion möglich ist. Beispielsweise können Sie ein System nutzen, das erwartete Werte wichtiger Leistungskennzahlen (KPIs) prognostiziert und bei Überschreiten bestimmter Schwellenwerte einen Alarm ausgibt. Denkbar ist auch ein Tool, das Bereitstellungen automatisch anhält oder zurücksetzt, wenn sich KPIs außerhalb der erwarteten Werte befinden.

Implementieren Sie Prozesse, die Ihnen Einblick in die Leistung gewähren, während Ihr Workload ausgeführt wird. Entwickeln Sie Dashboards für die Überwachung und legen Sie Leistungsnormen in Form von Grundwerten fest, um zu bestimmen, ob die Workload optimal funktioniert.

Gängige Antimuster:

- Sie geben dem Betriebspersonal nur die Möglichkeit, betriebliche Änderungen an der Workload vorzunehmen.
- Sie lassen alle Alarme ohne proaktive Behebung zum Betriebsteam filtern.

Vorteile der Einführung dieser bewährten Methode: Die proaktive Behebung von Alarmaktionen ermöglicht es dem Support-Personal, sich auf die Elemente zu konzentrieren, die nicht automatisch umsetzbar sind. Auf diese Weise wird sichergestellt, dass das Betriebspersonal nicht von allen Alarmen überfordert wird und sich stattdessen nur auf kritische Alarme konzentrieren kann.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Leistung im laufenden Betrieb überwachen: Implementieren Sie Prozesse, die Ihnen Einblick in die Leistung gewähren, während Ihr Workload ausgeführt wird. Erstellen Sie Überwachungs-Dashboards und legen Sie eine Basis für Leistungserwartungen fest.

Ressourcen

Ähnliche Dokumente:

- CloudWatch-Dokumentation
- Überwachung, Protokollierung und Leistung von APN-Partnern
- X-Ray-Dokumentation
- Verwendung von Alarmen und Alarmaktionen in CloudWatch

Ähnliche Videos:

- Ende des Chaos: Transparenz und Einblick in den Betrieb (MGT301-R1)
- Verwaltung der Anwendungsleistung in AWS
- Erstellen eines Überwachungsplans
- Verwenden von AWS Lambda mit Amazon CloudWatch Events

Ähnliche Beispiele:

- Cloudwatch-Protokolle: Konfigurieren von Alarmen

Kompromisse

Frage

- LEIST 8 Wie lässt sich Leistung durch Kompromisse verbessern?

LEIST 8 Wie lässt sich Leistung durch Kompromisse verbessern?

Durch die Festlegung von Kompromissen beim Gestalten von Lösungen lässt sich der optimale Ansatz einfacher bestimmen. Leistung lässt sich oft durch Zugeständnisse in anderen Bereichen verbessern, etwa bei Konsistenz, Beständigkeit, Zeit und Latenz.

Bewährte Methoden

- PERF08-BP01 Identifizieren von Bereichen mit kritischem Leistungsbedarf
- PERF08-BP02 Kennenlernen von Designmustern und Services
- PERF08-BP03 Identifizieren von Auswirkungen von Kompromissen auf Kunden und Effizienz
- PERF08-BP04 Messen der Auswirkung von Leistungsoptimierungen
- PERF08-BP05 Anwenden verschiedener Leistungsstrategien

PERF08-BP01 Identifizieren von Bereichen mit kritischem Leistungsbedarf

Ermitteln Sie die Bereiche, in denen sich durch Steigern der Workload-Leistung positive Auswirkungen auf die Effizienz oder den Kundenkomfort realisieren lassen. Beispiel: Eine Website mit zahlreichen Kundeninteraktionen kann von der Nutzung von Edge-Services profitieren, indem Inhalte näher bei den Kunden bereitgestellt werden.

Gewünschtes Ergebnis: Erhöhen Sie die Leistungseffizienz durch eingehendes Verständnis Ihrer Architektur, der Datenverkehrs- und der Datenzugriffmuster und identifizieren Sie Ihre Latenz- und Verarbeitungszeiten. Identifizieren Sie potenzielle Engpässe, die sich bei zunehmenden Workloads auf den Kundenkomfort auswirken könnten. Prüfen Sie im Rahmen der Identifizierung dieser Bereiche, welche Lösung Sie nutzen können, um diese Leistungsprobleme zu beseitigen.

Typische Anti-Muster:

- Sie gehen davon aus, dass Standard-Computing-Metriken wie `CPUUtilization` oder Speicherdruck ausreichen, um Leistungsprobleme zu identifizieren.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware Ihrer Wahl aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.

Vorteile der Nutzung dieser bewährten Methode: Das eingehende Verständnis kritischer Bereiche hilft Workload-Eigentümern dabei, KPIs zu überwachen und Verbesserungen mit größeren Auswirkungen zu priorisieren.

Risikostufe, wenn diese bewährte Methode nicht genutzt wird: Hoch

Implementierungsleitfaden

Richten Sie durchgehende Nachverfolgung ein, um Datenverkehrsmuster, Latenz und kritische Leistungsbereiche zu identifizieren. Überwachen Sie Ihre Datenzugriffsmuster auf langsame Abfragen oder schlecht fragmentierte und partitionierte Daten. Identifizieren Sie problematische Workload-Bereiche mithilfe von Lasttests oder -überwachung.

Implementierungsschritte

1. Richten Sie durchgehende Überwachung ein, um alle Workload-Komponenten und -Metriken zu erfassen.
   - Verwenden Sie Amazon CloudWatch Real-User Monitoring (RUM) zum Erfassen von Metriken zur Anwendungsleistung aus realen clientseitigen und Frontend-Sitzungen.
   - Richten Sie AWS X-Ray ein, um den Datenverkehr durch die Anwendungsebenen zu verfolgen und die Latenz zwischen Komponenten und Abhängigkeiten zu identifizieren. Verwenden Sie die X-Ray-Servicemaps, um Beziehungen und Latenz zwischen Workload-Komponenten zu erkennen.
   - Verwenden Sie Amazon Relational Database Service Performance Insights zum Anzeigen von Metriken zur Datenbankleistung und zum Identifizieren von Möglichkeiten zur Leistungsverbesserung.
   - Verwenden Sie Amazon RDS Enhanced Monitoring zum Anzeigen von Datenbank-BS-Leistungsmetriken.
   - Erfassen Sie CloudWatch-Metriken für die einzelnen Workload-Komponenten und Services und stellen Sie fest, welche Metriken Auswirkungen auf die Leistungseffizienz haben.
   - Richten Sie Amazon DevOps Guru für zusätzliche Einblicke in die Leistung und Empfehlungen ein.
2. Führen Sie Tests durch, um Metriken zu generieren sowie Datenverkehrsmuster, Engpässe und kritische Leistungsbereiche zu identifizieren.
   - Richten Sie CloudWatch Synthetic Canaries ein, um browserbasierte Benutzeraktivitäten programmgesteuert mit `cron` -Aufträgen oder Ratenausdrücken zu identifizieren und im Zeitverlauf konsistente Metriken zu erhalten.
   - Verwenden Sie die Lösung AWS Distributed Load Testing , um Spitzendatenverkehr zu generieren oder Workloads mit der erwarteten Wachstumsrate zu testen.

3. Evaluieren Sie die Metriken und die Telemetriedaten, um Ihre kritischen Leistungsbereiche zu identifizieren. Prüfen Sie diese Bereiche zusammen mit Ihrem Team und besprechen Sie Überwachung und Lösung zur Vermeidung von Engpässen.
4. Experimentieren Sie mit Leistungsverbesserungen und messen Sie diese Änderungen anhand von Daten.
   - Verwenden Sie CloudWatch Evidently zum Testen von neuen Verbesserungen und den Auswirkungen auf die Leistung des Workloads.

Aufwand für den Implementierungsplan: Um diese bewährte Methode zu nutzen, müssen Sie Ihre durchgehenden Metriken prüfen und die derzeitige Leistung Ihres Workloads kennen. Dies bedeutet mittleren Aufwand zur Einrichtung durchgehender Überwachung und zur Identifizierung Ihrer kritischen Leistungsbereiche.

Ressourcen

Zugehörige Dokumente:

- Amazon Builders' Library
- X-Ray-Dokumentation
- Amazon CloudWatch RUM
- Amazon DevOps Guru
- CloudWatch RUM und X-Ray

Zugehörige Videos:

- Introducing The Amazon Builders' Library (DOP328) (Einführung in die Amazon Builders' Library (DOP328))
- Demo von Amazon CloudWatch Synthetics

Zugehörige Beispiele:

- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client
- X-Ray SDK for Node.js
- X-Ray SDK for Python

- X-Ray SDK for Java
- X-Ray SDK for .Net
- X-Ray SDK for Ruby
- X-Ray Daemon
- Verteilte Lasttests auf AWS

PERF08-BP02 Kennenlernen von Designmustern und Services

Holen Sie Informationen zu den verschiedenen Designmustern und Services ein, die zu Leistungsoptimierungen beitragen, und machen Sie sich mit ihnen vertraut. Ermitteln Sie im Rahmen Ihrer Analyse, welche Kompromisse in Frage kommen, um eine höhere Leistung zu erzielen. Durch die Verwendung eines Cache-Service beispielsweise kann die Last von Datenbanksystemen verringert werden. Das Caching kann jedoch zu einer letztendlichen Datenkonsistenz führen und erfordert einen technischen Aufwand, um bei der Implementierung die geschäftlichen Anforderungen und die Erwartungen der Kunden zu erfüllen.

Gewünschtes Ergebnis: Die Prüfung von Designmustern wird Sie dazu bringen, ein Architekturdesign zu auswählen, das das leistungsfähigste System unterstützt. Machen Sie sich mit den Konfigurationsoptionen für die Leistung vertraut und finden Sie heraus, wie sich diese auf den Workload auswirken. Wie gut das Optimieren der Workload-Leistung gelingt, ist davon abhängig, wie gut Sie die Interaktion dieser Optionen mit Ihrer Architektur nachvollziehen können und davon, wie sich diese Optionen auf die gemessene und die von den Endbenutzern wahrgenommene Leistung auswirken.

Typische Anti-Muster:

- Sie gehen davon aus, dass alle herkömmlichen IT-Workload-Leistungsstrategien am besten für Cloud-Workloads geeignet sind.
- Sie erstellen und verwalten Caching-Lösungen, anstatt verwaltete Services zu verwenden.
- Sie verwenden dasselbe Designmuster für alle Ihre Workloads, ohne zu beurteilen, welches Muster die Workload-Leistung verbessern würde.

Vorteile der Nutzung dieser bewährten Methode: Durch die Auswahl des richtigen Designmusters und der richtigen Services für Ihre Workload können Sie die Leistung optimieren und so die operative Exzellenz verbessern und die Zuverlässigkeit erhöhen. Das richtige Designmuster wird Ihren

aktuellen Workload-Eigenschaften gerecht und erleichtert die Skalierung für zukünftiges Wachstum oder künftige Änderungen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Machen Sie sich mit den Konfigurationsoptionen für die Leistung vertraut und finden Sie heraus, wie sich diese auf die Workload auswirken. Der Erfolg beim Optimieren der Workload-Leistung ist davon abhängig, wie gut Sie die Interaktion dieser Optionen mit Ihrer Architektur nachvollziehen können und wie sich diese Optionen sowohl auf die gemessene als auch die von den Benutzern wahrgenommene Leistung auswirken.

Implementierungsschritte:

1. Evaluieren und prüfen Sie die Designmuster, die Ihre Workload-Leistung verbessern würden.
    a. Die Amazon Builders' Library enthält eine ausführliche Beschreibung dazu, wie Technologie von Amazon entwickelt und betrieben wird. Die dort enthaltenen Artikel werden von erfahrenen Technikern bei Amazon geschrieben und behandeln Themen in den Bereichen Architektur, Softwarebereitstellung und Betrieb.
    b. Die AWS-Lösungsbibliothek ist eine Sammlung von einsatzbereiten Lösungen, die Services, Code und Konfigurationen vereinen. Diese Lösungen wurden von AWS und AWS-Partnern auf der Grundlage von gängigen Anwendungsfällen und Designmustern erstellt, die nach Branche oder Workload-Typ gruppiert sind. Sie können beispielsweise eine Lösung für verteilte Lasttests für Ihre Workload einrichten.
    c. Im AWS-Architekturzentrum finden Sie Referenzarchitekturdiagramme, die nach Designmuster, Inhaltstyp und Technologie gruppiert sind.
    d. AWS Samples ist ein GitHub-Repository mit vielen praktischen Beispielen, anhand deren Sie gängige Architekturmuster, Lösungen und Services erkunden können. Das Repository wird häufig mit den neuesten Services und Beispielen aktualisiert.
2. Verbessern Sie Ihren Workload, um die ausgewählten Designmuster zu modellieren, und verwenden Sie Services und die Servicekonfigurationsoptionen, um Ihre Workload-Leistung zu verbessern.
    a. Schulen Sie Ihr internes Team mit den Ressourcen in AWS Skills Guild.
    b. Verwenden Sie das AWS Partner Network, um schnell Fachwissen zu bieten und Ihr Verbesserungspotenzial zu skalieren.

Aufwand für den Implementierungsplan: Um diese bewährte Methode einzuführen, müssen Sie sich über die Designmuster und Services im Klaren sein, die zur Verbesserung Ihrer Workload-Leistung beitragen könnten. Nach der Bewertung der Designmuster erfordert die Implementierung der Designmuster einen hohen Aufwand.

Ressourcen

Zugehörige Dokumente:

- Im AWS-Architekturzentrum
- AWS Partner Network
- Die AWS-Lösungsbibliothek
- AWS Knowledge Center
- In der Amazon Builders' Library
- Lastabwurf zur Vermeidung einer Überlastung
- Caching-Herausforderungen und -Strategien

Zugehörige Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- This is My Architecture:

Zugehörige Beispiele:

- AWS Samples
- AWS-SDK-Beispiele

PERF08-BP03 Identifizieren von Auswirkungen von Kompromissen auf Kunden und Effizienz

Ermitteln Sie beim Evaluieren von leistungsbezogenen Verbesserungen, welche gewählten Optionen sich auf Ihre Kunden und die Effizienz der Workloads auswirken. Wenn sich die Systemleistung beispielsweise bei Verwendung eines Schlüssel-Wert-Datenspeichers erhöht, sollten Sie unbedingt ermitteln, welche Auswirkungen sich bei einem dauerhaften Einsatz für die Kunden ergeben würden.

Identifizieren Sie anhand von Kennzahlen und Überwachung Bereiche Ihres Systems, die eine schlechte Leistung aufweisen. Stellen Sie fest, welche Verbesserungen möglich und welche Kompromisse damit verbunden sind und wie sich diese auf das System und das Benutzererlebnis

auswirken. So lässt sich beispielsweise durch Caching von Daten die Leistung deutlich steigern. Es ist aber eine eindeutige Strategie erforderlich, mit der festgelegt wird, wie und wann Cache-Daten aktualisiert oder ungültig werden, um unerwünschtes Systemverhalten zu verhindern.

Gängige Antimuster:

- Sie gehen davon aus, dass alle Leistungsgewinne implementiert werden sollten, auch wenn es Kompromisse für die Implementierung gibt, z. B. Eventual Consistency.
- Änderungen an Workloads werden nur dann ausgewertet, wenn ein Leistungsproblem einen kritischen Punkt erreicht hat.

Vorteile der Einführung dieser bewährten Methode: Wenn Sie potenzielle leistungsbezogene Verbesserungen bewerten, müssen Sie entscheiden, ob die Kompromisse für die Änderungen mit den Workload-Anforderungen übereinstimmen. In einigen Fällen müssen Sie möglicherweise zusätzliche Kontrollen implementieren, um Kompromisse zu kompensieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Ermitteln von Kompromissen: Identifizieren Sie anhand von Metriken und Überwachung die Bereiche Ihres Systems, die eine schlechte Leistung aufweisen. Bestimmen Sie, wie Verbesserungen vorgenommen werden können und wie Kompromisse sich auf das System und die Benutzererfahrung auswirken. So lässt sich beispielsweise durch Caching von Daten die Leistung deutlich steigern. Es ist aber eine eindeutige Strategie erforderlich, mit der festgelegt wird, wie und wann Cache-Daten aktualisiert oder ungültig werden, um unerwünschtes Systemverhalten zu verhindern.

Ressourcen

Ähnliche Dokumente:

- In der Amazon Builders' Library
- Amazon QuickSight-KPIs
- Amazon CloudWatch RUM
- X-Ray-Dokumentation

Ähnliche Videos:

- Einführung in die Amazon Builders' Library (DOP328)

- Erstellen eines Überwachungsplans
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)
- Demo von Amazon CloudWatch Synthetics

Ähnliche Beispiele:

- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client

PERF08-BP04 Messen der Auswirkung von Leistungsoptimierungen

Werten Sie die erfassten Metriken und Daten aus, wenn Änderungen zur Verbesserung der Leistung vorgenommen werden. Nutzen Sie diese Informationen, um die Auswirkungen zu ermitteln, die sich aufgrund der Leistungsverbesserung für die Workload, die zugehörigen Komponenten und Ihre Kunden ergeben haben. Anhand dieser Messungen lassen sich die dank des Kompromisses möglichen Verbesserungen einfacher nachvollziehen und Sie können feststellen, ob der Kompromiss eventuell zu unerwünschten Nebenwirkungen geführt hat.

In einem architektonisch guten System kommt meist eine Kombination verschiedener Leistungsstrategien zur Anwendung. Bestimmen Sie, welche Strategie die größte positive Wirkung auf einen bestimmten kritischen Punkt oder Engpass hat. Durch Sharding von Daten auf mehrere relationale Datenbanksysteme lässt sich der Gesamtdurchsatz verbessern, während Transaktionen weiterhin unterstützt werden. In den einzelnen Shards trägt Caching zur Lastreduzierung bei.

Gängige Antimuster:

- Sie stellen Technologien, die als verwaltete Services verfügbar sind, manuell bereit und verwalten sie.
- Sie konzentrieren sich auf nur eine Komponente, z. B. das Netzwerk, wenn mehrere Komponenten verwendet werden könnten, um die Leistung der Workload zu erhöhen.
- Sie verlassen sich auf Kundenfeedback und Kundenwahrnehmung als einzige Benchmark.

Vorteile der Einführung dieser bewährten Methode: Für die Implementierung von Leistungsstrategien müssen Sie mehrere Services und Funktionen auswählen, die es Ihnen ermöglichen, Ihre Workload-Anforderungen an die Leistung zu erfüllen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

In einem gut geplanten System kommt meist eine Kombination verschiedener Leistungsstrategien zur Anwendung. Ermitteln Sie, welche Strategie die größte positive Wirkung auf einen bestimmten kritischen Punkt oder Engpass hat. Durch Sharding von Daten auf mehrere relationale Datenbanksysteme lässt sich der Gesamtdurchsatz verbessern, während Transaktionen weiterhin unterstützt werden. In den einzelnen Shards trägt Caching zur Lastreduzierung bei.

Ressourcen

Ähnliche Dokumente:

- In der Amazon Builders' Library
- Amazon CloudWatch RUM
- Amazon CloudWatch Synthetics
- Verteilte Lasttests auf AWS

Ähnliche Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)
- Demo von Amazon CloudWatch Synthetics

Ähnliche Beispiele:

- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client
- Verteilte Lasttests auf AWS

PERF08-BP05 Anwenden verschiedener Leistungsstrategien

Wenden Sie nach Möglichkeit mehrere Strategien zur Leistungsoptimierung an. Verwenden Sie beispielsweise Strategien wie Daten-Caching, um exzessive Netzwerk- oder Datenbankaufrufe zu verhindern, und Lesereplikate für Datenbankmodule, um eine höhere Leserate zu erzielen.

Setzen Sie möglichst Sharding und Datenkomprimierung ein, um das Datenvolumen zu reduzieren, und nutzen Sie die Pufferung und das Streaming der verfügbaren Ergebnisse, um Blockaden zu vermeiden.

Wenn Sie Änderungen an Ihrer Workload vornehmen, sollten Sie Metriken erfassen und bewerten, um die Auswirkungen dieser Änderungen eindeutig zu bestimmen. Messen Sie die Auswirkungen auf System und Endbenutzer, um nachzuvollziehen, wie sich Ihre Kompromisse auf die Workload niederschlagen. Stellen Sie anhand eines systematischen Ansatzes fest (z. B. Lasttests), ob sich die Leistung durch den Kompromiss tatsächlich verbessert.

Gängige Antimuster:

- Sie gehen davon aus, dass die Workload-Leistung ausreichend ist, wenn sich Kunden nicht beschweren.
- Sie erfassen nur Daten zur Leistung, nachdem Sie leistungsbezogene Änderungen vorgenommen haben.

Vorteile der Einführung dieser bewährten Methode: Um Leistung und Ressourcenauslastung zu optimieren, benötigen Sie einen Gesamtüberblick über den Betrieb, detaillierte Echtzeitdaten und Referenzdaten aus der Vergangenheit. Sie können Dashboards erstellen und Metrikberechnungen für Ihre Daten durchführen, um Einblicke in Betrieb und Nutzung für Ihre Workloads zu erhalten, während sich diese im Laufe der Zeit ändern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Weiterentwicklung der Architektur mit datengestütztem Ansatz: Wenn Sie Änderungen an Ihrer Workload vornehmen, sollten Sie Metriken erfassen und bewerten, um die Auswirkungen dieser Änderungen eindeutig zu bestimmen. Messen Sie die Auswirkungen auf System und Endbenutzer, um nachzuvollziehen, wie sich Ihre Kompromisse auf die Workload auswirken. Stellen Sie anhand eines systematischen Ansatzes fest (z. B. Lasttests), ob sich die Leistung durch den Kompromiss tatsächlich verbessert.

Ressourcen

Ähnliche Dokumente:

- In der Amazon Builders' Library

- Bewährte Methoden für die Implementierung von Amazon ElastiCache
- AWS-Datenbank-Caching
- Amazon CloudWatch RUM
- Verteilte Lasttests auf AWS

Ähnliche Videos:

- Einführung in die Amazon Builders' Library (DOP328)
- Speziell entwickelte AWS-Datenbanken (DAT209-L)
- Optimize applications through Amazon CloudWatch RUM (Optimieren von Anwendungen mithilfe von CW RUM)

Ähnliche Beispiele:

- Messen der Seitenladezeit mit Amazon CloudWatch Synthetics
- Amazon CloudWatch RUM Web Client
- Verteilte Lasttests auf AWS

Kostenoptimierung

Themen

- Praxis für Cloud-Finanzmanagement
- Ausgabenerkennung und Nutzungsbewusstsein
- Kostengünstige Ressourcen
- Verwaltung von Nachfrage und Bereitstellung von Ressourcen
- Optimierung im Laufe der Zeit

Praxis für Cloud-Finanzmanagement

Frage

- KOSTEN 1 Wie implementieren Sie das Cloud Financial Management?

KOSTEN 1 Wie implementieren Sie das Cloud Financial Management?

Die Implementierung von Cloud Financial Management (CFM) ermöglicht es Unternehmen, geschäftlichen Nutzen und finanziellen Erfolg zu erzielen, wenn sie ihre Kosten und Nutzung optimieren und auf AWS skalieren.

Bewährte Methoden

- COST01-BP01 Implementieren einer Kostenoptimierungsfunktion
- COST01-BP02 Einrichten einer Partnerschaft zwischen Finanzen und Technologie
- COST01-BP03 Erstellen von Cloud-Budgets und -Prognosen
- COST01-BP04 Implementieren von Kostenbewusstsein in Ihre Organisationsprozesse
- COST01-BP05 Berichte und Benachrichtigungen zur Kostenoptimierung
- COST01-BP06 Proaktive Überwachung der Kosten
- COST01-BP07 Verfolgen neuer Serviceversionen

COST01-BP01 Implementieren einer Kostenoptimierungsfunktion

Richten Sie ein Team (Cloud Business Office oder Cloud Center of Excellence) ein, das für die Entwicklung und Wahrung eines Kostenbewusstseins in Ihrer gesamten Organisation verantwortlich ist. Das Team benötigt Mitarbeiter aus den Bereichen Finanzen, Technologie und Business in der gesamten Organisation.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Richten Sie ein Cloud Business Office (CBO)- oder Cloud Center of Excellence (CCOE)-Team ein, das für die Entwicklung und Wahrung einer Kultur des Kostenbewusstseins im Bereich Cloud-Computing verantwortlich ist. Dabei kann es sich um eine bestehende Person, ein Team innerhalb Ihres Unternehmens oder ein neues Team aus wichtigen Beteiligten in den Bereichen Finanzwesen und Technologie aus dem gesamten Unternehmen handeln.

Die Funktion (Einzelperson oder Team) priorisiert und verbraucht den erforderlichen Prozentsatz ihrer Zeit für Kostenmanagement- und Kostenoptimierungsaktivitäten. Bei kleinen Unternehmen kann die Funktion einen geringeren Prozentsatz der Zeit im Vergleich zu einer Vollzeitfunktion für ein größeres Unternehmen aufwenden.

Die Funktion erfordert einen multidisziplinären Ansatz, der Kompetenzen in den Bereichen Projektmanagement, Datenwissenschaft, Finanzanalyse und Software- oder Infrastrukturentwicklung erfordert. Die Funktion kann die Effizienz von Workloads durch Kostenoptimierungen auf drei unterschiedlichen Besitzebenen verbessern:

- Zentralisiert: Mit designierten Teams, beispielsweise in den Bereichen Finanzen, Kostenoptimierung, CBO oder CCOE, können Kunden Governance-Mechanismen entwerfen und implementieren sowie unternehmensweit Best Practices fördern.
- Dezentralisiert: Es wird Einfluss auf Technologieteams ausgeübt, um Optimierungen umzusetzen.
- Hybrid: Zentralisierte und dezentralisierte Teams arbeiten gemeinsam an der Umsetzung von Kostenoptimierungen.

Die Funktion kann anhand ihrer Fähigkeit zur Ausführung und Bereitstellung im Hinblick auf Kostenoptimierungsziele gemessen werden (z. B. Workload-Effizienzmetriken).

Sie müssen sicherstellen, dass Führungskräfte diese Funktion unterstützen, damit sie Änderungen einführen kann. Dies ist ein entscheidender Erfolgsfaktor. Der Förderer/Sponsor gilt als Befürworter für eine kosteneffiziente Cloud-Nutzung und bietet Eskalationsunterstützung für die Funktion, um sicherzustellen, dass die Aktivitäten zur Kostenoptimierung mit der vom Unternehmen definierten Priorität behandelt werden. Andernfalls werden Anleitungen nicht beachtet und Möglichkeiten für Kosteneinsparungen werden nicht priorisiert. Sponsor und Funktion stellen gemeinsam sicher, dass Ihre Organisation die Cloud effizient nutzt und weiterhin einen geschäftlichen Mehrwert erzielt.

Wenn Sie einen Business, Enterprise-On-Ramp oder Enterprise Support-Plan erworben haben und Hilfe bei der Einrichtung dieses Teams oder dieser Funktion benötigen, wenden Sie sich bitte über Ihr Account-Team an die Experten unseres Cloud Finance Management (CFM)-Teams.

Implementierungsschritte

- Definieren wichtiger Mitglieder: Sie müssen sicherstellen, dass alle relevanten Teile Ihres Unternehmens beitragen und einen Anteil an der Kostenverwaltung haben. Häufig handelt es sich hierbei um Teams mit Verantwortung für Finanzen, Anwendungen oder Produkte, das Management und technische Teams (DevOps). Einige Teams setzen ihre ganze Arbeitszeit hierfür ein (Finanzen, Technik), andere Teams werden wie erforderlich eingebunden. Die mit CFM befassten Personen oder Teams benötigen im Allgemeinen Kompetenzen in den folgenden Bereichen:
  - Softwareentwicklung – um Skripts und Automatisierungen entwickeln zu können.

  - Infrastrukturentwicklung – um Skripts, Automatisierungen, Services oder Ressourcen bereitstellen zu können.
  - Operatives Wissen – CFM stellt durch Messung, Überwachung, Änderung, Planung und Skalierung eine effiziente Nutzung der Cloud sicher.
- Definieren von Zielen und Metriken: Die Funktion muss der Organisation auf verschiedene Weise Mehrwert bieten. Diese Ziele werden definiert und mit der Entwicklung der Organisation kontinuierlich weiterentwickelt. Häufige Aktivitäten sind: Erstellen und Ausführen von Trainingsprogrammen zur Kostenoptimierung in der gesamten Organisation, Entwickeln von organisationsweiter Standards wie Überwachung und Berichterstellung zur Kostenoptimierung und zum Festlegen von Workload-Zielen für die Optimierung. Außerdem muss diese Funktion der Organisation regelmäßig über Möglichkeiten zur Kostenoptimierung Bericht erstatten.

  Sie können wertbasierte Leistungsindikatoren (Key Performance Indicators, KPIs) definieren. KPIs können kosten- oder wertbasiert sein. Wenn Sie KPIs definieren, können Sie die erwarteten Kosten in Bezug auf Effizienz und erwartete geschäftliche Ergebnisse berechnen. Wertbasierte KPIs verbinden Kosten- und Nutzungsmetriken mit Geschäftswertfaktoren und helfen, Änderungen bei AWS-Ausgaben zu begründen. Der erste Schritt bei der Formulierung wertbasierter KPIs besteht in der organisationsweiten Zusammenarbeit, um einen Standardsatz von KPIs auszuwählen und zu vereinbaren.
- Festlegen einer regulären Kadenz: Die Gruppe (Teams aus den Bereichen Finanzen, Technologie und Geschäft) sollte sich regelmäßig treffen, um Ziele und Metriken zu überprüfen. Dazu gehört in der Regel die Überprüfung des Status der Organisation, der aktuell ausgeführten Programme und der gesamten Finanz- und Optimierungsmetriken. Anschließend werden detaillierte Berichte zu wichtigen Workloads erstellt.

  Bei diesen regelmäßigen Besprechungen können Sie die Workload-Effizienz (Kosten) und die geschäftlichen Ergebnisse überprüfen. Eine Kostensteigerung von 20 % für einen Workload könnte beispielsweise mit einer erhöhten Nutzung durch Kunden zusammenhängen. In einem solchen Fall kann die Kostensteigerung von 20 % als Investition betrachtet werden. Diese regelmäßigen Besprechungen können Teams helfen, wertbasierte KPIs zu identifizieren, die für die gesamte Organisation sinnvoll sind.

Ressourcen

Zugehörige Dokumente:

- AWS CCOE-Blog

- Einrichtung von Cloud Business Office
- CCOE – Cloud Center of Excellence

Zugehörige Videos:

- Vanguard CCOE, eine Erfolgsgeschichte

Zugehörige Beispiele:

- Nutzung eines Cloud-Kompetenzzentrums (Center of Excellence, CCOE) zur Transformation des gesamten Unternehmens
- Einrichtung eines CCOE zur Transformation des gesamten Unternehmens
- 7 Fehler, die Sie bei der Einrichtung eines CCOE vermeiden sollten

COST01-BP02 Einrichten einer Partnerschaft zwischen Finanzen und Technologie

Beziehen Sie Finanz- und Technologieteams in Kosten- und Nutzungsgespräche in allen Phasen Ihrer Cloud-Reise mit ein. Teams treffen sich regelmäßig, um Themen wie Unternehmensziele, aktuellen Kosten- und Nutzungsstatus sowie Finanz- und Buchhaltungsmethoden zu besprechen.

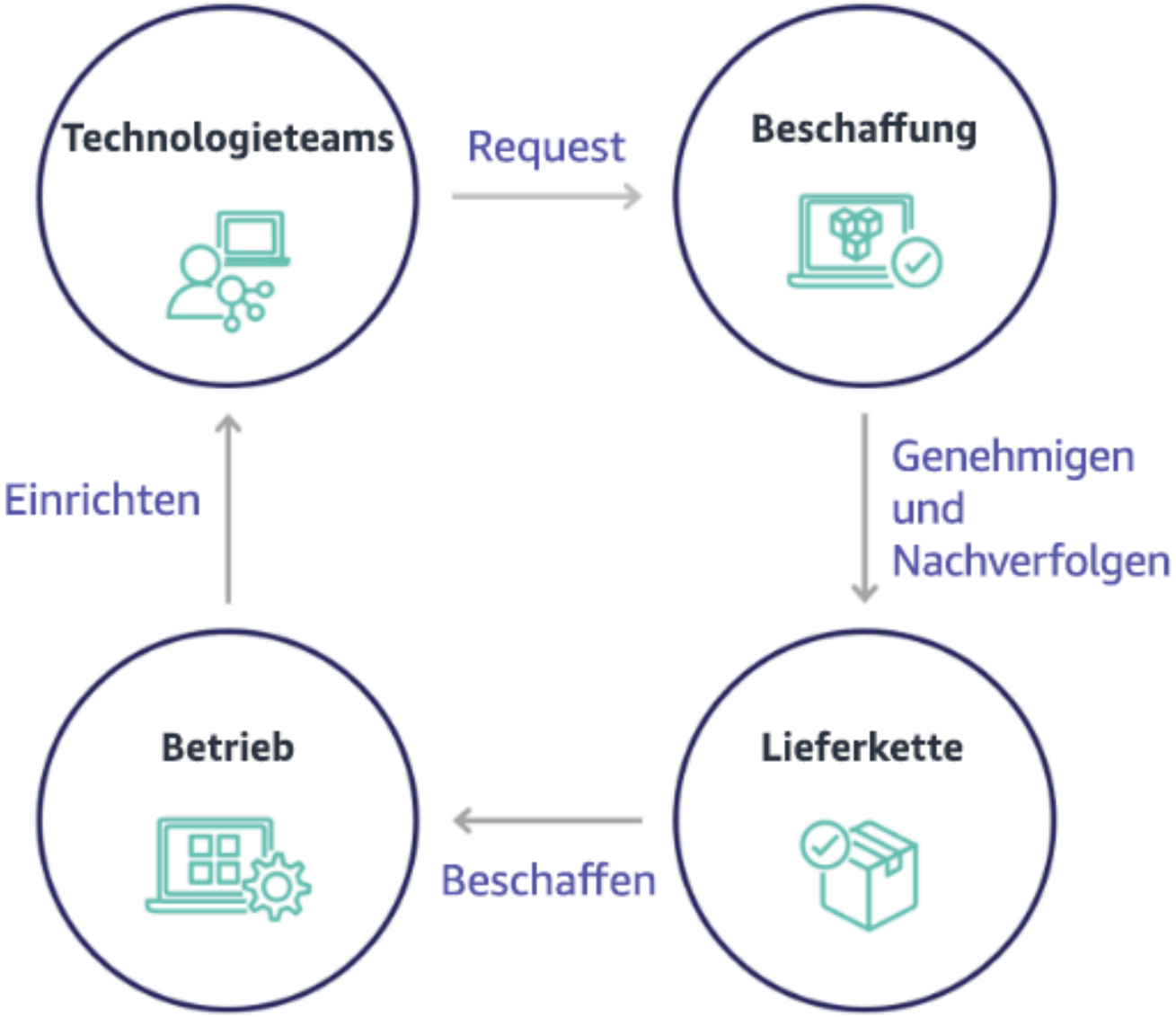
Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Technologieteams können in der Cloud dank verkürzter Genehmigungs-, Beschaffungs- und Infrastrukturbereitstellungszyklen schneller Innovationen vorantreiben. Dies kann eine Anpassung für Finanzunternehmen sein, die zuvor an die Ausführung zeitaufwändiger und ressourcenintensiver Prozesse zur Beschaffung und Bereitstellung von Kapital in Rechenzentrums- und lokalen Umgebungen und die Kostenzuordnung nur nach Projektgenehmigung gewöhnt waren.

Was die Finanz- und Beschaffungsabteilungen betrifft, wurden die Prozesse in den Bereichen Budgetierung, Kapitalbedarf, Genehmigung, Beschaffung und Installation der physischen Infrastruktur über Jahrzehnte hinweg weiterentwickelt und standardisiert.

- In der Regel fordern die Entwicklungs- oder IT-Teams die Geldmittel an.
- Die Finanzteams genehmigen und beschaffen die Geldmittel.
- Die operativen Teams stellen die Infrastruktur zusammen, sodass sie direkt eingesetzt werden kann.



Mit der Einführung der Cloud werden Beschaffung und Nutzung der Infrastruktur nicht mehr als Kette von Abhängigkeiten betrachtet. Im Cloud-Modell entwickeln Technologie- und Produktteams ihre Produkte nicht nur, sondern führen sie auch selbst aus und sind für sie verantwortlich. Dabei führen sie die meisten Aktivitäten aus, die bisher als Domäne der Finanz- und operativen Teams betrachtet wurden, einschließlich Beschaffung und Bereitstellung.

Zur Bereitstellung von Cloud-Ressourcen werden lediglich ein Benutzerkonto und der richtige Satz von Berechtigungen benötigt. Dies reduziert auch die Risiken in den Bereichen IT und Finanzen, da die Teams stets nur einige Klicks oder API-Aufrufe von der Einstellung nicht genutzter oder nicht notwendiger Cloud-Ressourcen entfernt sind. Technologieteams können so auch schneller Innovationen einführen und erhalten die nötige Agilität und Flexibilität, um Experimente zu starten

und zu beenden. Auch wenn sich die variable Natur der Cloud-Nutzung auf die Planbarkeit der Budgetierung und die Genauigkeit von Prognosen auswirken kann, bietet sie Organisationen jedoch auch die Möglichkeit, sowohl die Kosten für Überbereitstellungen als auch die Opportunitätskosten für konservative Unterbereitstellungen zu reduzieren.

Bauen Sie eine Partnerschaft zwischen wichtigen Beteiligten aus dem Finanzwesen und der Technologie auf, um ein gemeinsames Verständnis der organisatorischen Ziele zu schaffen und Mechanismen zu entwickeln, um im variablen Ausgabenmodell von Cloud Computing einen finanziellen Erfolg zu erzielen. Relevante Teams innerhalb Ihres Unternehmens müssen an Kosten- und Nutzungsdiskussionen in allen Phasen Ihrer Cloud-Reise beteiligt sein, einschließlich:

- Verantwortliche im Finanzbereich: CFOs, Finanzkontrolleure, Finanzplaner, Geschäftsanalysten, Beschaffung und Kreditorenbuchhaltung müssen das Cloud-Modell des Verbrauchs, Kaufoptionen und den monatlichen Rechnungsprozess verstehen. Die Teams in den Bereichen Finanzen und Technologie müssen zusammenarbeiten, um die IT-Wertschöpfung zu entwickeln und

darzustellen, damit die geschäftlichen Teams die Verbindung zwischen Technologieausgaben und Geschäftsergebnissen verstehen können. Auf diese Weise werden Technologieaufwendungen nicht als Kosten angesehen, sondern als Investitionen. Aufgrund der grundlegenden Unterschiede zwischen der Cloud (z. B. Änderungsrate der Nutzung, Pay-as-you-go-Preisgestaltung, gestaffelte Preise, Preismodelle und detaillierte Abrechnungs- und Nutzungsinformationen) im Vergleich zum Betrieb vor Ort ist es für die Finanzorganisation von entscheidender Bedeutung, dass sie versteht, wie sich die Nutzung der Cloud auf geschäftliche Aspekte wie Beschaffungsprozesse, Anreizverfolgung, Kostenzuordnung und Finanzberichte auswirken kann.

- Verantwortliche im Technologiebereich: Technologieverantwortliche (einschließlich Produkt- und Anwendungsbesitzer) müssen die finanziellen Anforderungen (z. B. Budgeteinschränkungen) sowie die geschäftlichen Anforderungen (z. B. Service Level Agreements) kennen. Damit kann das System implementiert werden, um die gewünschten Ziele des Unternehmens zu erreichen.

Die Partnerschaft zwischen Finanzen und Technologie bietet folgende Vorteile:

- Finanz- und Technologieteams haben nahezu in Echtzeit Einblicke in Kosten und Nutzung.
- Finanz- und Technologieteams legen ein standardmäßiges Betriebsverfahren für die Bewältigung von Ausgabeunterschieden in der Cloud fest.
- Stakeholder im Bereich Finanzen handeln als strategische Berater bei der Nutzung von Kapital für den Kauf rabattierter Programme (z. B. Reserved Instances oder AWS Savings Plans) und der Nutzung der Cloud zur Förderung des Wachstums der Organisation.
- Vorhandene Kreditorenbuchhaltungs- und Beschaffungsprozesse werden mit der Cloud verwendet.
- Die Finanz- und Technologieteams prognostizieren gemeinsam die Kosten und die Nutzung von AWS in der Zukunft, um die Budgets der Organisation entsprechend auszurichten und zu entwickeln.
- Bessere unternehmensübergreifende Kommunikation durch eine gemeinsame Sprache und ein gemeinsames Verständnis von Finanzkonzepten.

Weitere Beteiligte innerhalb Ihres Unternehmens, die an Kosten- und Nutzungsdiskussionen beteiligt sein sollten, sind:

- Besitzer von Geschäftseinheiten: Besitzer von Geschäftseinheiten müssen sich mit dem Cloud-Geschäftsmodell vertraut machen, sodass sie den Geschäftseinheiten und dem gesamten Unternehmen die Richtung weisen können. Dieses Cloud-Wissen ist wichtig, wenn es erforderlich

ist, das Wachstum und die Systemnutzung zu prognostizieren oder verschiedene Kaufoptionen zu bewerten, z. B. Reserved Instances oder Savings Plans.

- Entwicklungsteam: Eine Partnerschaft zwischen Finanz- und Technologieteams hat kritische Bedeutung für die Entwicklung einer kostenbewussten Kultur, die Entwickler motiviert, im Bereich Cloud Financial Management (CFM) aktiv zu werden. Ein häufiges Problem von CFM- und Finanzteams besteht darin, Entwicklern ein Verständnis des Geschäfts in der Cloud zu vermitteln und sie zur Umsetzung von Best Practices und empfohlenen Aktionen zu motivieren.
- Dritte: Wenn Ihr Unternehmen mit Dritten arbeitet (z. B. Berater oder Tools), dann stellen Sie sicher, dass diese an Ihren finanziellen Zielen ausgerichtet sind und sowohl die Ausrichtung durch ihre Engagement-Modelle als auch einen ROI (Return on Investment) nachweisen können. In der Regel beteiligen sich Dritte an der Berichterstellung und Analyse der von ihnen verwalteten Systeme, und sie stellen Kostenanalysen für die von ihnen konzipierten Workloads bereit.

Eine erfolgreiche CFM-Implementierung erfordert die Zusammenarbeit von Teams in den Bereichen Finanzen, Technologie und Geschäft sowie eine veränderte Kommunikation und Evaluierung in Bezug auf die Cloud-Ausgaben der Organisation. Beziehen Sie die Entwicklungsteams in alle Phasen der Diskussion über Kosten- und Nutzung ein und motivieren Sie sie zur Befolgung von Best Practices und zur Umsetzung vereinbarter Aktionen.

Implementierungsschritte

- Definieren wichtiger Mitglieder: Stellen Sie sicher, dass sich alle relevanten Mitglieder Ihrer Finanz- und Technologieteams aktiv an der Partnerschaft beteiligen. Relevante Mitglieder im Bereich Finanzen sind Personen, die mit Cloud-Ausgaben interagieren. Dies sind in der Regel CFOs, Finanzcontroller, Finanzplaner, Geschäftsanalysten und Mitarbeiter in Beschaffung und Einkauf. Technologiemitarbeiter sind in der Regel Produkt- und Anwendungsbesitzer, technische Manager und Vertreter aller Teams, die in der Cloud aktiv sind. Weitere Mitglieder können Geschäftsbereiche mit Einfluss auf die Nutzung von Produkten sein, zum Beispiel das Marketing, und Dritte wie Berater, die Sie bei der Ausrichtung an Ihren Zielen und Mechanismen und bei Berichten unterstützen.
- Definieren von Diskussionsthemen: Definieren Sie die Themen, die in den Teams häufig auftreten, oder ein gemeinsames Verständnis erfordern. Verfolgen Sie die Kosten ab dem Zeitpunkt, an dem sie generiert werden, bis zur Bezahlung der Rechnung. Beachten Sie alle beteiligten Mitglieder und organisatorischen Prozesse, die angewendet werden müssen. Informieren Sie sich über jeden einzelnen Schritt oder Prozess, den sie durchlaufen, sowie die zugehörigen Informationen, wie

z. B. verfügbare Preismodelle, gestaffelte Preise, Rabattmodelle, Budgetplanung und finanzielle Anforderungen.

- Festlegen einer regulären Kadenz: Richten Sie eine regelmäßige Kommunikationskadenz ein, um Finanz- und Technologieteams aneinander auszurichten und eine Partnerschaft zu unterstützen. Die Gruppe muss regelmäßig im Hinblick auf ihre Ziele und Metriken zusammenkommen. Dazu gehört in der Regel die Überprüfung des Status der Organisation, der aktuell ausgeführten Programme und der gesamten Finanz- und Optimierungsmetriken. Anschließend werden detaillierte Berichte zu wichtigen Workloads erstellt.

Ressourcen

Zugehörige Dokumente:

- AWS News-Blog

COST01-BP03 Erstellen von Cloud-Budgets und -Prognosen

Passen Sie vorhandene Budgetierungs- und Prognoseprozesse so an, dass sie mit der stark variablen Natur der Cloud-Kosten und -Nutzung kompatibel sind. Prozesse müssen dynamisch sein und Algorithmen anwenden, die auf Trends oder Geschäftsfaktoren oder einer Kombination von diesen basieren.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Kunden nutzen die Cloud für Effizienz, Geschwindigkeit und Agilität, wodurch sich Kosten und Nutzung in hohem Maße ändern. Die Kosten können durch eine höhere Workload-Effizienz oder durch die Bereitstellung neuer Workloads und Funktionen gesenkt werden. Es ist möglich, dass Kostensteigerungen auftreten, wenn die Workload-Effizienz steigt oder neue Workloads und Features bereitgestellt werden. Oder Workloads werden so skaliert, dass sie mehr Ihrer Kunden bedienen können, was die Cloud-Nutzung und -Kosten erhöht. Ressourcen sind heute einfacher verfügbar als je zuvor. Die Elastizität der Cloud bedeutet auch Elastizität bei Kosten und Prognosen. Bestehende organisatorische Budgetierungsprozesse müssen geändert werden, um diese Variabilität zu berücksichtigen.

Dynamisieren Sie vorhandene Budgetierungs- und Prognoseprozesse. Hierzu können Sie einen trendbasierten Algorithmus (mit historischen Kosten als Eingabe) oder einen Algorithmus verwenden,

der auf Geschäftsfaktoren basiert (z. B. auf der Einführung neuer Produkte oder auf einer regionalen Expansion). Sie können auch einen Algorithmus verwenden, der auf einer Kombination aus beidem basiert.

Mit AWS Budgets können Sie angepasste, detaillierte Budgets einrichten, indem Sie Zeitraum, Rekurrenz oder Betrag (fest oder variabel) angeben und Filter wie Service, AWS-Region und Tags hinzufügen. Um bei vorhandenen Budgets auf dem Laufenden zu bleiben, können Sie AWS Budgets-Berichte einrichten und planen, die Ihnen und Ihren Stakeholdern regelmäßig per E-Mail gesendet werden. Sie können auch reaktive AWS Budgets-Warnungen basierend auf den tatsächlichen Kosten einrichten oder mit Alarmen zu prognostizierten Kosten Maßnahmen zur Vermeidung möglicher Kostenüberschreitungen ermöglichen. Sie werden benachrichtigt, wenn Kosten oder Nutzung den budgetierten Betrag überschreiten oder in der Zukunft möglicherweise überschreiten werden.

Mit AWS erhalten Sie die nötige Flexibilität für die Entwicklung dynamischer Prognose- und Budgetierungsprozesse, damit Sie stets wissen, ob Ihre Kosten die Budgetlimits einhalten oder überschreiten.

Mit AWS Cost Explorer können Sie Kosten für einen definierten zukünftigen Zeitraum prognostizieren, basierend auf Ihren bisherigen Ausgaben. Die Prognose-Engine von AWS Cost Explorer segmentiert Ihre historischen Daten basierend auf Gebührentypen (z. B. Reserved Instances) und verwendet eine Kombination aus Machine Learning und regelbasierten Modellen, um die Ausgaben für alle Gebührentypen individuell zu prognostizieren. Verwendung Sie AWS Cost Explorer für tägliche (bis zu drei Monate) oder monatliche (bis zu 12 Monate) Cloud-Kosten-Prognosen, basierend auf Machine-Learning-Algorithmen, die auf Ihre historischen Kosten (trendbasiert) angewendet werden.

Sobald Sie mit Cost Explorer Ihre trendbasierte Prognose erstellt haben, verwenden Sie AWS Pricing Calculator , um Ihre AWS-Anwendungsfall- und künftigen Kosten auf der Grundlage der erwarteten Nutzung (Datenverkehr, Anfragen pro Sekunde, erforderliche Amazon Elastic Compute Cloud (Amazon EC2)-Instance usw.) zu schätzen. Sie können damit auch Ihre Ausgaben planen, Möglichkeiten für Kosteneinsparungen finden und informierte Entscheidungen bei der Verwendung von AWS treffen.

Verwendung Sie AWS Cost Anomaly Detection zur Vermeidung oder Verringerung von Kostenüberraschungen und für eine bessere Kontrolle, ohne dadurch Innovationen zu verlangsamen. AWS Cost Anomaly Detection nutzt modernste Machine-Learning-Technologien, um anomale Ausgaben und deren Ursachen zu identifizieren, damit Sie schnell handeln können. Mit drei einfachen Schrittenkönnen Sie Ihre eigene kontextsensitive Überwachung einrichten und benachrichtigt werden, wenn anomale Ausgaben erkannt werden. Lassen Sie die damit befassten Personen Dinge

erstellen und lassen Sie AWS Cost Anomaly Detection Ihre Ausgaben überwachen und das Risiko überraschend hoher Rechnungen senken.

Wie im Unterabschnitt Partnerschaft zwischen Finanzen und Technologie als Säule für eine gut gestaltete Kostenoptimierung bereits erwähnt, ist es wichtig, eine Partnerschaft mit regelmäßigen Konsultationen zwischen IT, Finanzabteilung und anderen Beteiligten zu schaffen, um sicherzustellen, dass alle in konsistenter Weise die gleichen Hilfsmittel oder Prozesse anwenden. Wenn Budgets geändert werden müssen, können häufigere Besprechungen dabei helfen, schneller darauf zu reagieren.

Implementierungsschritte

- Aktualisieren vorhandener Budget- und Prognoseprozesse: Implementieren Sie trendbasierte oder von geschäftlichen Faktoren unterstützte Algorithmen oder eine Kombination aus beidem in Ihre Budgetierungs- und Prognoseprozesse.
- Konfigurieren von Warnungen und Benachrichtigungen: Verwenden Sie AWS Budgets-Warnungen und Cost Anomaly Detection.
- Regelmäßige Prüfungen zusammen mit zentralen Beteiligten: Dazu gehören etwa Beteiligte in den Bereichen IT, Finanzen, Plattform und anderen, die an der geschäftlichen Ausrichtung und bestehenden Praktiken ausgerichtet werden müssen.

Ressourcen

Zugehörige Dokumente:

- AWS Cost Explorer
- AWS Budgets
- AWS Pricing Calculator
- AWS Cost Anomaly Detection
- AWS License Manager

Zugehörige Beispiele:

- Start: Nutzungsbasierte Prognosen, jetzt verfügbar in AWS Cost Explorer
- AWS Well-Architected Labs: Steuerung der Kosten und Nutzung

COST01-BP04 Implementieren von Kostenbewusstsein in Ihre Organisationsprozesse

Implementieren Sie Kostenbewusstsein und sorgen Sie für Transparenz und Verantwortlichkeit bei neuen oder bestehenden Prozessen, die sich auf die Nutzung auswirken, und greifen Sie auf vorhandene Prozesse zur Steigerung des Kostenbewusstseins zurück. Implementieren Sie Kostenbewusstsein in die Mitarbeiterschulung.

Risikostufe bei fehlender Befolgung dieser Best Practice: Hoch

Implementierungsleitfaden

Das Kostenbewusstsein muss in neuen und vorhandenen Organisationsprozessen implementiert werden. Dies ist eine der absoluten Grundlagen für weitere bewährte Methoden. Es wird empfohlen, vorhandene Prozesse nach Möglichkeit wiederzuverwenden und zu ändern. Dadurch werden die Auswirkungen auf Agilität und Geschwindigkeit minimiert. Informieren Sie die Technologieteams und die Entscheidungsträger in den Geschäfts- und Finanzteams über die Cloud-Kosten, um das Kostenbewusstsein zu verbessern, und richten Sie KPIs zur Effizienz für Beteiligte aus dem Finanz- und Geschäftsbereich ein. Die folgenden Empfehlungen helfen Ihnen bei der Implementierung der Kostenerkennung in Ihrem Workload:

- Stellen Sie sicher, dass das Änderungsmanagement eine Kostenmessung umfasst, um die finanziellen Auswirkungen Ihrer Änderungen zu quantifizieren. Auf diese Weise können Sie kostenbezogene Probleme proaktiv lösen und Kosteneinsparungen hervorheben.
- Stellen Sie sicher, dass die Kostenoptimierung eine zentrale Komponente Ihrer Betriebsfunktionen ist. Sie können beispielsweise vorhandene Vorfallmanagementprozesse nutzen, um die Ursache für Kosten- und Nutzungsanomalien (Kostenüberschreitungen) zu ermitteln und zu identifizieren.
- Beschleunigen Sie die Kosteneinsparungen und die Wertschöpfung des Unternehmens durch Automatisierung oder Tools. Wenn Sie über die Kosten der Implementierung nachdenken, sollten Sie das Gespräch so gestalten, dass es eine ROI-Komponente enthält, um die Investition von Zeit oder Geld zu rechtfertigen.
- Weisen Sie Cloud-Kosten zu, indem Sie Showbacks oder Chargebacks für Cloud-Aufwendungen implementieren, einschließlich Aufwendungen für verpflichtungsbasierte Kaufoptionen, gemeinsam genutzte Services und Markt-Einkäufe, um die Cloudnutzung in möglichst kostenbewusster Weise zu gestalten.
- Erweitern Sie vorhandene Schulungs- und Entwicklungsprogramme, um Schulungen zum Kostenbewusstsein in Ihrem gesamten Unternehmen einzubeziehen. Es wird empfohlen, dass dies fortlaufende Schulungen und Zertifizierungen umfasst. Dadurch entsteht ein Unternehmen, das Kosten und Nutzung selbst verwalten kann.

- Nutzen Sie kostenlose, native AWS-Tools, wie etwa AWS Cost Anomaly Detection, AWS Budgetsund AWS Budgets-Berichte.

Wenn Unternehmen Cloud Financial Management (CFM)-Praktiken in konsistenter Weise einsetzen, werden die entsprechenden Verhaltensweisen bald echte Bestandteile der Arbeitsweise und der Entscheidungsfindung. Das führt zu einer kostenbewussteren Kultur, in der Entwickler eine neue, in der Cloud entwickelte Anwendung bauen und Finanzmanager den ROI dieser neuen Cloud-Investitionen analysieren.

Implementierungsschritte

- Bestimmen relevanter organisatorischer Prozesse: Jede Organisationseinheit überprüft ihre Prozesse und identifiziert Prozesse, die sich auf Kosten und Nutzung auswirken. Alle Prozesse, die zur Erstellung oder Beendigung einer Ressource führen, müssen zur Überprüfung einbezogen werden. Suchen Sie auch nach Prozessen, die das Kostenbewusstsein in Ihrem Unternehmen unterstützen können, wie z. B. Vorfallmanagement und Schulungen.
- Schaffen einer sich selbst erhaltenden Kostenbewusstseinskultur: Sorgen Sie dafür, dass alle relevanten Beteiligten die Ursachen für Veränderungen und die damit verbundenen Kosten gut verstehen. So kann Ihr Unternehmen eine sich selbst erhaltende, kostenbewusste Innovationskultur entwickeln.
- Aktualisieren von Prozessen mit Kostenbewusstsein: Jeder Prozess wird so geändert, dass er kostenbewusst wird. Der Prozess erfordert möglicherweise zusätzliche Vorabprüfungen, z. B. die Bewertung der Auswirkungen von Kosten oder nachträgliche Prüfungen, die bestätigen, dass die erwarteten Kosten- und Nutzungsänderungen stattgefunden haben. Unterstützungsprozesse wie Schulungs- und Vorfallmanagement können auf Kosten- und Nutzungselemente erweitert werden.

Wenden Sie sich für Unterstützung über Ihr Account-Team an CFM-Sachverständige oder erkunden Sie die nachfolgend aufgeführten Ressourcen und Dokumente.

Ressourcen

Zugehörige Dokumente:

- AWS Cloud Financial Management

Zugehörige Beispiele:

- Strategie für effizientes Cloud-Kostenmanagement
- Blog-Serie zum Thema Kostenkontrolle Nr. 3: Umgang mit Kostenschocks
- AWS Cost Management für Anfänger

COST01-BP05 Berichte und Benachrichtigungen zur Kostenoptimierung

Konfigurieren Sie AWS Budgets und AWS Cost Anomaly Detection, um Benachrichtigungen über Kosten und Nutzung im Vergleich zu den Zielen zu ermöglichen. Analysieren Sie bei regelmäßig abgehaltenen Besprechungen die Kosteneffizienz Ihres Workloads und fördern Sie das Kostenbewusstsein im Unternehmen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Niedrig

Implementierungsleitfaden

Sie müssen regelmäßig Kosten- und Nutzungsoptimierungen in Ihrem Unternehmen melden. Sie können dedizierte Sitzungen zur Kostenoptimierung implementieren oder die Kostenoptimierung in Ihre regulären operativen Berichtszyklen für Ihre Workloads einschließen. Nutzen Sie Services und Tools, um Möglichkeiten für Kosteneinsparungen zu identifizieren und zu implementieren. AWS Cost Explorer stellt Dashboards und Berichte bereit. Sie können Ihren Kosten- und Nutzungsfortschritt anhand konfigurierter Budgets mit AWS Budgets-Berichten nachverfolgen.

Mit AWS Budgets können Sie angepasste Budgets einrichten, um Kosten und Nutzung nachzuverfolgen und schnell auf Warnungen zu reagieren, die Sie per E-Mail oder in Form von Amazon Simple Notification Service (Amazon SNS)-Benachrichtigungen erhalten, wenn Sie Ihren Schwellenwert überschreiten. Sie können den bevorzugten Budgetzeitraum auf täglich, monatlich, vierteljährlich oder jährlich festlegen und spezifische Budgetlimits einrichten, um zu sehen, wie sich die tatsächlichen oder prognostizierten Kosten in Bezug auf Ihren Budgetschwellenwert entwickeln. Sie können auch eine automatische Ausführung von Warnungen und Aktionen oder einen Genehmigungsprozess für den Fall einrichten, dass ein Budgetziel überschritten wird.

Darüber hinaus können Sie mit Benachrichtigungen zu Kosten und Nutzung schnell auf unerwartete Änderungen bei Kosten und Nutzung reagieren. AWS Cost Anomaly Detection ermöglicht Ihnen die Reduzierung von Überraschungen bei den Kosten und die Verbesserung der Kontrolle, ohne die Innovationsfähigkeit zu beeinträchtigen. AWS Cost Anomaly Detection identifiziert anomale Ausgaben und ihre Ursachen, was Ihnen hilft, das Risiko für Überraschungen bei Abrechnungen zu reduzieren. In drei einfachen Schritten können Sie Ihre eigene kontextorientierte Überwachung einrichten und Benachrichtigungen erhalten, wenn anomale Ausgaben entdeckt werden.

Sie können Amazon QuickSight mit AWS Cost and Usage Report (CUR)-Daten verwenden, um hoch angepasste Berichte mit detaillierteren Daten zu erstellen. Amazon QuickSight ermöglicht Ihnen die Planung von Berichten und den Erhalt regelmäßiger E-Mails mit Berichten zu historischen Kosten und zur Nutzung oder zu Möglichkeiten für Kosteneinsparungen.

Mit AWS Trusted Advisorerhalten Sie Anleitungen, mit denen Sie überprüfen können, ob bereitgestellte Ressourcen Best Practices für AWS zur Kostenoptimierung befolgen.

Sie können regelmäßige Berichte erstellen, die Savings Plans, Reserved Instances und Amazon Elastic Compute Cloud (Amazon EC2)-Empfehlungen aus AWS Cost Explorer für Anpassungen enthalten, um die Kosten für Steady-State-Workloads sowie nicht genutzte und nicht vollständig genutzte Ressourcen zu reduzieren. Identifizieren Sie unnötige Cloud-Ausgaben, die mit bereitgestellten Ressourcen verbunden sind, und gewinnen Sie diese zurück. Unnötige Cloud-Ausgaben entstehen, wenn Ressourcen mit der falschen Größe erstellt werden oder wenn andere als die erwarteten Nutzungsmuster beobachtet werden. Befolgen Sie die Best Practices für AWS, um unnötige Ausgaben zu reduzieren, Ihre Cloud-Kosten zu optimieren und zu sparen.

Generieren Sie regelmäßig Berichte zu besseren Kaufoptionen für Ihre Ressourcen, um die Kosten pro Einheit für Ihre Workloads zu senken. Kaufoptionen wie Savings Plans, Reserved Instances oder Amazon EC2 Spot Instances bieten die umfassendsten Kosteneinsparungen für fehlertolerante Workloads. Stakeholder (Geschäftsbereichsleiter, Finanz- und Technologieteams) können sich an den Diskussionen zu den damit verbundenen Verpflichtungen beteiligen.

Teilen Sie die Berichte, die Einsparmöglichkeiten beschreiben, oder Ankündigungen neuer Versionen, um die Gesamtbetriebskosten (TCO) der Cloud zu reduzieren. Führen Sie neue Services, Regionen, Funktionen, Lösungen oder neue Möglichkeiten für weitere Kostenreduzierungen ein.

Implementierungsschritte

- Konfigurieren Sie AWS Budgets: Konfigurieren Sie AWS Budgets für alle Konten Ihres Workloads. Legen Sie ein Budget für die Gesamtkontoausgaben und ein Budget für den Workload mithilfe von Tags fest.
  - Well-Architected Labs: Kosten und Steuerung der Nutzung
- Bericht zur Kostenoptimierung: Richten Sie einen regelmäßigen Zyklus ein, um die Effizienz des Workloads zu besprechen und zu analysieren. Melden Sie anhand der eingerichteten Metriken die erreichten Metriken und die Kosten für deren Erreichung. Identifizieren und beheben Sie negative Trends und suchen Sie nach positiven Trends, die Sie in der gesamten Organisation unterstützen können. Bei der Berichterstellung sollten Vertreter der Anwendungsteams und Besitzer, Finanz- und Geschäftsleitung einbezogen werden.

- Well-Architected Labs: Visualisierung

Ressourcen

Zugehörige Dokumente:

- AWS Cost Explorer
- AWS Trusted Advisor
- AWS Budgets
- AWS Budgets – Bewährte Methoden
- Amazon CloudWatch
- AWS CloudTrail
- Amazon S3-Analysen
- AWS Cost and Usage Report

Zugehörige Beispiele:

- Well-Architected Labs: Kosten und Steuerung der Nutzung
- Well-Architected Labs: Visualisierung
- Zentrale Methoden für die Optimierung Ihrer AWS-Cloud-Kosten

COST01-BP06 Proaktive Überwachung der Kosten

Implementieren Sie Tools und Dashboards, um die Kosten proaktiv für den Workload zu überwachen. Überprüfen Sie regelmäßig die Kosten mithilfe konfigurierter oder vorab erstellter Tools. Untersuchen Sie Kosten und Kategorien nicht erst, wenn Sie Benachrichtigungen erhalten. Die proaktive Überwachung und Analyse der Kosten hilft Ihnen, positive Trends zu identifizieren und diese in der gesamten Organisation zu unterstützen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Niedrig

Implementierungsleitfaden

Es wird empfohlen, die Kosten und die Nutzung innerhalb Ihres Unternehmens proaktiv zu überwachen, nicht nur, wenn Ausnahmen oder Anomalien vorliegen. Hoch sichtbare Dashboards in Ihrem Büro oder Ihrer Arbeitsumgebung stellen sicher, dass relevante Mitarbeiter Zugriff auf benötigte

Informationen haben, und signalisieren den Fokus des Unternehmens auf Kostenoptimierungen. Mit gut sichtbaren Dashboards können Sie den Erfolg aktiv unterstützen und und positive Ergebnisse in der gesamten Organisation implementieren.

Entwickeln Sie eine tägliche oder häufig ausgeführte Routine für die Verwendung von AWS Cost Explorer oder anderen Dashboards wie Amazon QuickSight, um die Kosten darzustellen und proaktiv zu analysieren. Analysieren Sie mithilfe von Gruppierung und Filterung Kosten und Nutzung von AWS-Services auf der Ebene von AWS-Konten, Workloads oder spezifischen AWS-Services und überprüfen Sie, ob es sich um erwartete oder unerwartete Ergebnisse handelt. Nutzen Sie die Granularität und die Tags auf Stunden- und Ressourcenbasis, um für die wichtigsten Ressourcen wiederkehrende Kosten herauszufiltern und zu identifizieren. Sie können auch über das Cost Intelligence Dashboardeigene Berichte erstellen. Dabei handelt es sich um eine Amazon QuickSight-Lösung, die von AWS Solution Architects entwickelt wurde. Sie ermöglicht Ihnen den Vergleich Ihrer Budgets mit den tatsächlichen Kosten und der tatsächlichen Nutzung.

Implementierungsschritte

- Bericht zur Kostenoptimierung: Richten Sie einen regelmäßigen Zyklus ein, um die Effizienz des Workloads zu besprechen und zu analysieren. Melden Sie anhand der eingerichteten Metriken die erreichten Metriken und die Kosten für deren Erreichung. Identifizieren und beheben Sie negative Trends und suchen Sie nach positiven Trends, um diese in der gesamten Organisation zu unterstützen. Bei der Berichterstellung sollten Vertreter der Anwendungsteams und Besitzer, Finanz- und Geschäftsleitung einbezogen werden.
- Erstellen und aktivieren Sie tägliche, detaillierte AWS Budgets für Kosten und Nutzung, um rechtzeitig Maßnahmen gegen potenzielle Kostenüberschreitungen ergreifen zu können. Mit AWS Budgets können Sie Warnungen konfigurieren, um stets zu wissen, ob ein Budgettyp außerhalb der vorab konfigurierten Schwellenwerte liegt. Die beste Art, AWS Budgets zu nutzen, besteht in der Einrichtung der erwarteten Kosten und der erwarteten Nutzung als Grenzwerte. So können alle Budgetüberschreitungen identifiziert werden.
- Erstellen Sie AWS Cost Anomaly Detection zur Kostenüberwachung: AWS Cost Anomaly Detection verwendet eine erweiterte Machine-Learning-Technologie, um anomale Ausgaben und ihre Ursachen schnell zu identifizieren, damit Sie schnell Maßnahmen ergreifen können. Sie können auf diese Weise Tools für die Überwachung der Kosten von Ausgabensegmenten konfigurieren, die Sie überwachen möchten (z. B. einzelne AWS-Services, Mitgliederkonten, Kostenzuweisungs-Tags und Kostenkategorien). Sie können auch festlegen, wann, wo und wie Sie Warnungen erhalten. Jedem Überwachungstool können Sie mehrere Warnungsabonnements für Geschäftsbereichsleiter und Technologieteams anfügen, einschließlich

Name, Kostenschwellenwert und Häufigkeit (einzelne Warnungen, tägliche Zusammenfassung, wöchentliche Zusammenfassung) für die einzelnen Abonnements.

- Verwenden Sie AWS Cost Explorer oder integrieren Sie Ihre AWS Cost and Usage Report (CUR)-Daten in Amazon QuickSight-Dashboards, um die Kosten Ihrer Organisation zu visualisieren: AWS Cost Explorer besitzt eine benutzerfreundliche Oberfläche, in der Sie AWS-Kosten und -Nutzung über die Zeit visualisieren, verstehen und verwalten können. Das Cost Intelligence Dashboard ist ein anpassbares und zugängliches Dashboard, mit dem Sie die Grundlagen für Ihr eigenes Tool für Kostenmanagement und Optimierung legen können.

Ressourcen

Zugehörige Dokumente:

- AWS Budgets
- AWS Cost Explorer
- Tägliche Kosten und Nutzungsbudgets
- AWS Cost Anomaly Detection

Zugehörige Beispiele:

- Well-Architected Labs: Visualisierung
- Well-Architected Labs: Erweiterte Visualisierung
- Well-Architected Labs: Cloud Intelligence Dashboards
- Well-Architected Labs: Kostenvisualisierung
- AWS Cost Anomaly Detection-Warnung mit Slack

COST01-BP07 Verfolgen neuer Serviceversionen

Konsultieren Sie regelmäßig Experten oder AWS-Partner, um zu prüfen, welche Services und Funktionen kostengünstiger sind. Lesen Sie AWS-Blogs und sonstige Informationsquellen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Niedrig

Implementierungsleitfaden

AWS fügt ständig neue Funktionen hinzu, so dass Ihnen die neuesten Technologien zur Verfügung stehen, damit Sie experimentieren und Innovationen schneller einführen können. Sie können

möglicherweise neue AWS-Services und -Funktionen implementieren, um die Kosteneffizienz Ihres Workloads zu erhöhen. Lesen Sie regelmäßig AWS Cost Management, den AWS News-Blog, den AWS Cost Management-Blogund Neuerungen bei AWS, um Informationen zur Veröffentlichung neuer Services und Funktionen zu erhalten. Die Posts in „Neuerungen“ bieten eine kurze Übersicht über alle Ankündigungen für AWS-Services, -Funktionen und -Regionserweiterungen bei Veröffentlichung.

Implementierungsschritte

- Abonnieren Sie Blogs: Rufen Sie die Seiten für AWS-Blogs auf und abonnieren Sie den Blog „Neuerungen“ und andere relevante Blogs. Sie können sich auf der Seite für die Kommunikationseinstellungen mit Ihrer E-Mail-Adresse registrieren.
- Abonnieren Sie AWS-Nachrichten: Lesen Sie regelmäßig den AWS News-Blog und Neuerungen bei AWS, um Informationen zur Veröffentlichung neuer Services und Funktionen zu erhalten. Abonnieren Sie den RSS-Feed oder registrieren Sie sich über Ihre E-Mail-Adresse, um Ankündigungen und Veröffentlichungen zu folgen.
- Verfolgen Sie AWSPreisreduzierungen: Wir geben die wirtschaftliche Effizienz, die wir aufgrund unserer Skalierbarkeit erzielen, mit regelmäßigen Preissenkungen für alle unsere Services als AWS-Standardverfahren an unsere Kunden weiter. Seit April 2022 hat AWS die Preise seit der Einführung im Jahr 2006 115 Mal gesenkt. Wenn geschäftliche Entscheidungen aufgrund von Preisbedenken ausstehen, können Sie die Preise nach der Reduzierung und der Integration neuer Services erneut prüfen. Informationen zu früheren Preissenkungen, einschließlich Preissenkungen für Amazon Elastic Compute Cloud (Amazon EC2)-Instances, finden Sie in der Kategorie „Preissenkungen“ im AWS News-Blog.
- AWS-Veranstaltungen und -Treffen: Nehmen Sie am lokalen AWS-Summit und weiteren lokalen Treffen mit anderen Organisationen aus Ihrer Region teil. Wenn eine persönliche Teilnahme nicht möglich ist, können Sie in virtuellen Veranstaltungen mehr von AWS-Experten und über die Business Cases anderer Kunden erfahren.
- Treffen Sie sich mit Ihrem Account-Team: Planen Sie regelmäßige Treffen mit Ihrem Account-Team, um über Branchentrends und AWS-Services zu sprechen. Sprechen Sie mit Ihrem Account Manager, Solutions Architekt und Support-Team.

Ressourcen

Zugehörige Dokumente:

- AWS Cost Management
- Neuerungen bei AWS,

- AWS News-Blog

Zugehörige Beispiele:

- Amazon EC2 – 15 Years of Optimizing and Saving Your IT Costs
- AWS News-Blog – Preisreduzierung

Ausgabenerkennung und Nutzungsbewusstsein

Fragen

- KOSTEN 2 Wie können Sie die Nutzung steuern?
- KOSTEN 3 Wie können Sie die Nutzung und Kosten überwachen?
- KOSTEN 4 Wie können Sie Ressourcen außer Betrieb nehmen?

KOSTEN 2 Wie können Sie die Nutzung steuern?

Definieren Sie Richtlinien und Verfahren, um sicherzustellen, dass sich die Kosten auf dem Weg zur Erreichung Ihrer Ziele in einem angemessenen Rahmen bewegen. Durch den Einsatz eines Kontrollsystems können Sie Innovationen vorantreiben, ohne das Budget zu überschreiten.

Bewährte Methoden

- COST02-BP01 Entwickeln von Richtlinien auf Basis Ihrer Organisationsanforderungen
- COST02-BP02 Implementieren von Zielen und Ergebnissen
- COST02-BP03 Implementieren einer Kontenstruktur
- COST02-BP04 Implementieren von Gruppen und Rollen
- COST02-BP05 Implementieren von Kostenkontrollen
- COST02-BP06 Verfolgen des Projektlebenszyklus

COST02-BP01 Entwickeln von Richtlinien auf Basis Ihrer Organisationsanforderungen

Entwickeln Sie Richtlinien, die definieren, wie Ressourcen von Ihrer Organisation verwaltet werden. Die Richtlinien sollten sich auch mit den Kostenaspekten der Ressourcen und Workloads befassen, einschließlich Erstellung, Änderung und Außerbetriebnahme während der gesamten Lebensdauer der Ressourcen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Die Kenntnis der Kostentreiber in Ihrem Unternehmen ist für die effektive Verwaltung Ihrer Ausgaben und Nutzung und die Identifizierung von Kostenreduzierungsmöglichkeiten von entscheidender Bedeutung. Unternehmen betreiben in der Regel mehrere Workloads, die von mehreren Teams ausgeführt werden. Diese Teams können sich in verschiedenen Organisationseinheiten befinden, die jeweils über eigene Einnahmequellen verfügen. Die Möglichkeit, die Ressourcenkosten den Workloads, der jeweiligen Organisation oder den Produkteigentümern zuzuordnen, fördert ein effizientes Nutzungsverhalten und hilft, Abfall zu reduzieren. Mit einer präzisen Kosten- und Nutzungszuordnung können Sie die Rentabilität von Organisationseinheiten und Produkten nachvollziehen und fundiertere Entscheidungen dazu treffen, welchen Bereichen in Ihrem Unternehmen Ressourcen zugeordnet werden. Das Bewusstsein der Nutzung auf allen Unternehmensebenen ist entscheidend für Veränderungen, da eine Änderung der Nutzung zu Kostenänderungen führt. Überlegen Sie sich, beim Ermitteln von Nutzungsmustern und Ausgaben einen mehrschichtigen Ansatz zu nutzen.

Der erste Schritt bei der Implementierung von Governance besteht darin, Richtlinien für die Cloud-Nutzung anhand der Anforderungen Ihres Unternehmens zu entwickeln. Diese Richtlinien definieren, wie Ihr Unternehmen die Cloud verwendet und wie Ressourcen verwaltet werden. Richtlinien sollten alle Aspekte von Ressourcen und Workloads abdecken, die sich auf Kosten oder Nutzung beziehen, einschließlich Erstellung, Änderung und Außerbetriebnahme über die Lebensdauer der Ressource.

Richtlinien sollten einfach sein, damit sie leicht verständlich sind und im gesamten Unternehmen effektiv implementiert werden können. Beginnen Sie mit umfangreichen allgemeinen Richtlinien, z. B. in welcher geografischen Region die Nutzung zulässig ist oder zu welchen Tageszeiten Ressourcen ausgeführt werden sollen. Verfeinern Sie schrittweise die Richtlinien für die verschiedenen Organisationseinheiten und Workloads. Häufige Richtlinien legen fest, welche Services und Funktionen verwendet werden können (z. B. Speicher mit niedrigerer Leistung in Test- oder Entwicklungsumgebungen) und welche Ressourcentypen von verschiedenen Gruppen verwendet werden können (z. B. ist die größte Ressourcen in einem Entwicklungskonto mittelgroß).

Implementierungsschritte

- Treffen mit Teammitgliedern: Um Richtlinien zu entwickeln, rufen Sie alle Teammitglieder aus Ihrer Organisation auf, ihre Anforderungen anzugeben und sie entsprechend zu dokumentieren. Führen Sie einen iterativen Ansatz aus, indem Sie bei jedem Schritt umfassend beginnen und kontinuierlich auf die kleinsten Einheiten verfeinern. Zu den Teammitgliedern gehören Personen

mit direktem Interesse am Workload, z. B. Organisationseinheiten oder Anwendungsbesitzer sowie unterstützende Gruppen wie Sicherheits- und Finanzteams.

- Festlegen von Speicherorten für Ihren Workload: Definieren Sie, wo Ihr Workload ausgeführt wird, einschließlich des Landes und der Region innerhalb des Landes. Diese Informationen werden für die Zuweisung zu AWS-Regionen und Availability Zones verwendet.
- Definieren und Gruppieren von Services und Ressourcen: Definieren Sie die Services, die für die Workloads erforderlich sind. Geben Sie für jeden Service die Typen, den Umfang und die Anzahl der erforderlichen Ressourcen an. Definieren Sie Gruppen für die Ressourcen nach Funktion, z. B. Anwendungsserver oder Datenbankspeicher. Ressourcen können mehreren Gruppen angehören.
- Definieren und Gruppieren der Benutzer nach Funktion: Definieren Sie die Benutzer, die mit dem Workload interagieren, und konzentrieren Sie sich darauf, was sie tun und wie sie den Workload verwenden, nicht auf die Benutzer oder ihre Position in der Organisation. Fassen Sie ähnliche Benutzer oder Funktionen in einer Gruppe zusammen. Sie können die von AWS verwalteten Richtlinien als Leitfaden verwenden.
- Definieren der Aktionen: Definieren Sie mithilfe der zuvor identifizierten Standorte, Ressourcen und Benutzer die Aktionen, die von jedem benötigt werden, um die Workload-Ergebnisse über die Lebensdauer (Entwicklung, Betrieb und Außerbetriebnahme) zu erzielen. Identifizieren Sie die Aktionen an jedem Standort basierend auf den Gruppen, nicht auf den einzelnen Elementen in den Gruppen. Beginnen Sie umfassend mit Lese- oder Schreibvorgängen und verfeinern Sie dann auf bestimmte Aktionen für jeden Service.
- Definieren des Überprüfungszeitraums: Workloads und Organisationsanforderungen können sich im Laufe der Zeit ändern. Definieren Sie den Zeitplan für die Überprüfung des Workloads, um sicherzustellen, dass er mit den Prioritäten der Organisation übereinstimmt.
- Dokumentieren der Richtlinien: Stellen Sie sicher, dass auf die definierten Richtlinien zugegriffen werden kann, wie von Ihrer Organisation gefordert. Diese Richtlinien werden verwendet, um den Zugriff auf Ihre Umgebungen zu implementieren, zu verwalten und zu prüfen.

Ressourcen

Zugehörige Dokumente:

- Von AWS verwaltete Richtlinien für Auftragsfunktionen
- AWS-Fakturierungsstrategie mit mehreren Konten
- Aktionen, Ressourcen und Bedingungsschlüssel für AWS-Services
- Cloud-Produkte

- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien
- Globale Infrastruktur-Regionen und -AZs

COST02-BP02 Implementieren von Zielen und Ergebnissen

Implementieren Sie sowohl Kosten- als auch Nutzungsziele für Ihren Workload. Ziele geben Ihrem Unternehmen Informationen zu Kosten und Nutzung und Ergebnisse liefern einen messbaren Wert für Ihre Workloads.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Entwickeln Sie Kosten- und Nutzungsziele sowie Vorgaben für Ihr Unternehmen. Ziele bieten Ihrem Unternehmen richtungsweisende Anleitungen hinsichtlich der erwarteten Ergebnisse. Vorgaben bieten spezifische messbare Ergebnisse, die erreicht werden müssen. Ein Beispiel für ein Ziel ist: Die Plattformnutzung sollte deutlich steigen, mit nur einer geringfügigen (nicht-linearen) Kostensteigerung. Ein Beispiel für eine Vorgabe ist eine 20 % höhere Plattformnutzung mit weniger als 5 % höheren Kosten. Ein weiteres häufiges Ziel ist, dass Workloads alle sechs Monate effizienter sein müssen. Die zugehörige Vorgabe wäre, dass die Kosten pro Ausgabe der Workload alle 6 Monate um 5 % gesenkt werden müssen.

Ein häufiges Ziel für Cloud-Workloads besteht darin, die Workload-Effizienz zu steigern, d. h. die Kosten pro Geschäftsergebnis des Workloads im Laufe der Zeit zu senken. Es wird empfohlen, dieses Ziel für alle Workloads zu implementieren und eine Vorgabe festzulegen, z. B. eine Effizienzsteigerung von 5 % alle 6–12 Monate. Dies kann in der Cloud durch die Entwicklung von Fähigkeiten bei der Kostenoptimierung und durch die Veröffentlichung neuer Services und Service-Funktionen erreicht werden.

Implementierungsschritte

- Definieren der erwarteten Nutzungsgrade: Konzentrieren Sie sich zunächst auf die Nutzungsebenen. Interagieren Sie mit den Anwendungsbesitzern, Marketing und größeren Geschäftsteams, um zu verstehen, wie die erwartete Nutzung für den Workload aussehen wird. Wie ändert sich die Kundennachfrage im Laufe der Zeit und gibt es Änderungen aufgrund saisonaler Erhöhungen oder Marketingkampagnen?
- Definieren von Ressourcen und Kosten für Workloads: Mit den definierten Nutzungsstufen quantifizieren Sie die Änderungen der Workload-Ressourcen, die erforderlich sind, um diese Nutzungsebenen zu erfüllen. Möglicherweise müssen Sie den Umfang oder die Anzahl der

Ressourcen für eine Workload-Komponente und die Datenübertragung erhöhen oder Workload-Komponenten in einen anderen Service auf einer bestimmten Ebene ändern. Geben Sie an, welche Kosten an jedem dieser wichtigen Punkte entstehen und welche Änderungen sich bei den Kosten ergeben, wenn sich die Nutzung ändert.

- Definieren von Geschäftszielen: Nehmen Sie die Ergebnisse zu den erwarteten Änderungen bei Nutzung und Kosten, kombinieren Sie sie mit den erwarteten Änderungen bei der Technologie oder sonstigen Programmen, die Sie ausführen, und entwickeln Sie Ziele für den Workload. Ziele müssen die Nutzung, die Kosten und die Beziehung zwischen den beiden berücksichtigen. Überprüfen Sie, dass es organisatorische Programme gibt, z. B. Kompetenzaufbau wie Schulungen und Fortbildungen, wenn sich zwar die Kosten ändern, aber die Nutzung nicht.
- Definieren der Ergebnisse: Geben Sie für jedes der definierten Ziele ein messbares Ergebnis an. Wenn ein Ziel darin besteht, die Effizienz des Workloads zu erhöhen, quantifiziert das Ergebnis den Grad der Verbesserung. Dies erfolgt typischerweise in Form einer Relation des Business-Outputs für jeden ausgegebenen Dollar und dem Zeitpunkt der Umsetzung.

Ressourcen

Zugehörige Dokumente:

- Von AWS verwaltete Richtlinien für Auftragsfunktionen
- AWS-Fakturierungsstrategie mit mehreren Konten
- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien

COST02-BP03 Implementieren einer Kontenstruktur

Implementieren Sie eine Kontenstruktur, die für Ihre Organisation geeignet ist. Dadurch werden die Zuweisung und Verwaltung der Kosten in der gesamten Organisation erleichtert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

AWS verfügt über eine Kontostruktur mit einem übergeordneten Konto und vielen untergeordneten Konten, die auch als Kontostruktur mit Verwaltungskonto (übergeordnet, zuvor Zahlerkonto) und Kontomitgliedern (untergeordnet, zuvor verknüpftes Konto) bezeichnet wird. Eine bewährte Methode besteht darin, immer mindestens ein Verwaltungs- mit einem Mitgliedskonto zu haben, unabhängig von der Unternehmensgröße oder Nutzung. Alle Workload-Ressourcen sollten sich nur innerhalb von Mitgliedskonten befinden.

Es gibt keine einheitliche Antwort dazu, über wie viele AWS-Konten Sie verfügen sollten. Zunächst sollten Sie Ihre aktuellen und künftigen Betriebs- und Kostenmodelle bewerten, um sicherzustellen, dass die Struktur Ihrer AWS-Konten die Ziele Ihres Unternehmens repräsentiert. Einige Unternehmen erstellen aus geschäftlichen Gründen mehrere AWS-Konten, z. B.:

- Es ist eine administrative und/oder fiskale und fakturierungsbezogene Abgrenzung zwischen Organisationseinheiten oder Kostenstellen oder spezifischen Workloads erforderlich.
- AWS-Service-Limits wurden für bestimmte Workloads definiert.
- Es besteht eine Anforderung für Isolierung und Trennung zwischen Workloads und Ressourcen.

Innerhalb von AWS Organizations, erstellt die konsolidierte Fakturierung das Konstrukt zwischen einem oder mehreren Mitgliedskonten und dem Managementkonto. Mit Mitgliedskonten können Sie Ihre Kosten und Nutzung nach Gruppen isolieren und unterscheiden. In diesem Kontext hat es sich bewährt, separate Mitgliedskonten für jede Organisationseinheit (z. B. Finanzen, Marketing und Vertrieb) oder für jeden Umgebungslebenszyklus (z. B. Entwicklung, Tests und Produktion) oder für jeden einzelnen Workload (Workload a, b und c) zu erstellen und diese verknüpften Konten dann über die konsolidierte Fakturierung zu aggregieren.

Mit der konsolidierten Fakturierung können Sie die Zahlung für mehrere AWS-Konten unter einem einzelnen Managementkonto konsolidieren und dabei weiterhin die Sichtbarkeit der Aktivitäten jedes verknüpften Kontos bereitstellen. Da Kosten und Nutzung im Managementkonto aggregiert werden, können Sie sowohl Ihre Service-Volumenrabatte als auch die Nutzung Ihrer an feste Kapazität gebundene Rabatte (Savings Plans und Reserved Instances) maximieren, und so die höchsten Vergünstigungen erzielen.

AWS Control Tower kann schnell mehrere AWS-Konten einrichten und konfigurieren, um sicherzustellen, dass Governance den Anforderungen Ihres Unternehmens entspricht.

Implementierungsschritte

- Definieren von Trennungsanforderungen: Die Trennungsanforderungen sind eine Kombination aus mehreren Faktoren, darunter fallen Sicherheit, Zuverlässigkeit und finanzielle Konstrukte. Arbeiten Sie die einzelnen Faktoren in der richtigen Reihenfolge durch und geben Sie an, ob der Workload oder die Workload-Umgebung von anderen Workloads getrennt sein sollte. Die Sicherheitsmaßnahmen gewährleisten, dass Zugriffs- und Datenanforderungen erfüllt werden. Die Zuverlässigkeit stellt sicher, dass Limits verwaltet werden, sodass Umgebungen und Workloads keine Auswirkungen auf andere Elemente haben. Finanzkonstrukte stellen sicher, dass eine strikte finanzielle Trennung und Verantwortlichkeit besteht. Häufige Beispiele für die Trennung

sind Produktions- und Test-Workloads, die in separaten Konten ausgeführt werden, oder die Verwendung eines separaten Kontos, sodass die Rechnungs- und Fakturierungsdaten einer Drittanbieterorganisation bereitgestellt werden können.

- Definieren von Gruppierungsanforderungen: Die Anforderungen für die Gruppierung überschreiben die Trennungsanforderungen nicht, sondern dienen zur Unterstützung der Verwaltung. Gruppieren Sie ähnliche Umgebungen oder Workloads, die keine Trennung erfordern. Ein Beispiel hierfür ist die Gruppierung mehrerer Test- oder Entwicklungsumgebungen aus einem oder mehreren Workloads.
- Definieren der Kontenstruktur: Geben Sie mit diesen Trennungen und Gruppierungen ein Konto für jede Gruppe an und stellen Sie sicher, dass die Trennungsanforderungen erfüllt werden. Diese Konten sind Ihre Mitgliedskonten oder verknüpfte Konten. Indem Sie diese Mitgliedskonten unter einem einzigen Management-/Zahlungskonto gruppieren, kombinieren Sie die Nutzung, was höhere Volumenrabatte für alle Konten ermöglicht und eine einzige Rechnung für alle Konten bereitstellt. Es ist möglich, Fakturierungsdaten zu trennen und jedem Mitgliedskonto eine individuelle Ansicht ihrer Fakturierungsdaten bereitzustellen. Definieren Sie mehrere Management-/Zahlungskoten, wenn die Nutzungs- oder Fakturierungsdaten eines Mitgliedskontos für kein anderes Konto sichtbar sein dürfen oder wenn eine separate Rechnung von AWS erforderlich ist. In diesem Fall hat jedes Mitgliedskonto ein eigenes Management-/Zahlungskonto. Ressourcen sollten immer in Mitgliedskonten oder verknüpften Konten platziert werden. Die Management-/Zahlungskonten sollten nur für die Verwaltung verwendet werden.

Ressourcen

Zugehörige Dokumente:

- Von AWS verwaltete Richtlinien für Auftragsfunktionen
- AWS-Fakturierungsstrategie mit mehreren Konten
- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien
- AWS Control Tower
- AWS Organizations
- Konsolidierte Fakturierung

Zugehörige Beispiele:

- Teilen des CUR und Freigabe des Zugangs

COST02-BP04 Implementieren von Gruppen und Rollen

Implementieren Sie Gruppen und Rollen, die Ihren Richtlinien entsprechen, und steuern Sie, wer Instances und Ressourcen in jeder Gruppe erstellen, ändern oder außer Betrieb nehmen kann. Implementieren Sie beispielsweise Entwicklungs-, Test- und Produktionsgruppen. Dies gilt sowohl für AWS-Services als auch für Lösungen anderer Anbieter.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Nachdem Sie Richtlinien entwickelt haben, können Sie logische Gruppen und Rollen von Benutzern innerhalb Ihrer Organisation erstellen. Auf diese Weise können Sie Berechtigungen zuweisen und die Nutzung steuern. Beginnen Sie mit allgemeinen Personengruppen. Dies entspricht in der Regel den Organisationseinheiten und Jobrollen (z. B. Systemadministrator in der IT-Abteilung oder Financial Controller). Den Gruppen treten Personen bei, die ähnliche Aufgaben ausführen und ähnlichen Zugriff benötigen. Rollen definieren, was eine Gruppe tun muss. Beispielsweise benötigt ein Systemadministrator in der IT Zugriff, um alle Ressourcen zu erstellen, aber ein Analyseteammitglied muss nur Analyseressourcen erstellen.

Implementierungsschritte

- Implementieren von Gruppen: Implementieren Sie bei Bedarf die entsprechenden Gruppen mithilfe der Benutzergruppen, die in Ihren Organisationsrichtlinien definiert sind. Bewährte Methoden für Benutzer, Gruppen und Authentifizierung finden Sie in der Säule der Sicherheit.
- Implementieren von Rollen und Richtlinien: Erstellen Sie mithilfe der Aktionen, die in Ihren Organisationsrichtlinien definiert sind, die erforderlichen Rollen und Zugriffsrichtlinien. Bewährte Methoden für Rollen und Richtlinien finden Sie in der Säule der Sicherheit.

Ressourcen

Zugehörige Dokumente:

- Von AWS verwaltete Richtlinien für Auftragsfunktionen
- AWS-Fakturierungsstrategie mit mehreren Konten
- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien
- Well-Architected: Säule „Sicherheit“

Zugehörige Beispiele:

- Well-Architected Lab: grundlegende Identität und Zugriff

COST02-BP05 Implementieren von Kostenkontrollen

Implementieren Sie Kontrollmechanismen, die auf den Organisationsrichtlinien sowie auf definierten Gruppen und Rollen basieren. Damit wird sichergestellt, dass die Kosten den Rahmen der festgelegten Organisationsanforderungen nicht sprengen. Mithilfe von AWS Identity and Access Management-Richtlinien (IAM) können Sie beispielsweise den Zugriff auf Regionen oder Ressourcentypen steuern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Ein häufiger erster Schritt bei der Implementierung von Kostenkontrollen ist die Einrichtung von Benachrichtigungen, wenn Kosten- oder Nutzungsereignisse außerhalb der Richtlinien der Organisation auftreten. Auf diese Weise können Sie schnell agieren und überprüfen, ob Korrekturmaßnahmen erforderlich sind, ohne Workloads oder neue Aktivitäten einzuschränken oder negativ zu beeinflussen. Nachdem Sie die Workload- und Umgebungslimits kennen, können Sie Governance erzwingen. In AWS werden Benachrichtigungen mit AWS Budgets durchgeführt. So können Sie ein monatliches Budget für Ihre AWS-Kosten, die Nutzung und an feste Kapazität gebundene Rabatte (Savings Plans und Reserved Instances) definieren. Sie können Budgets auf aggregierter Kostenebene (z. B. alle Kosten) oder auf einer detaillierteren Ebene erstellen, in der Sie nur bestimmte Dimensionen wie verknüpfte Konten, Services, Tags oder Availability Zones einschließen.

In einem zweiten Schritt können Sie Governance-Richtlinien in AWS über AWS Identity and Access Management (IAM) und AWS Organizations Service-Kontrollrichtlinien (Service Control Policies, SCP)durchsetzen. Mit IAM können Sie den Zugriff auf AWS-Services und -Ressourcen sicher verwalten. Mit IAM können Sie steuern, wer AWS-Ressourcen erstellen und verwalten kann, welche Art von Ressourcen erstellt werden kann und wo sie erstellt werden können. Dadurch wird die Erstellung von Ressourcen minimiert, die nicht erforderlich sind. Verwenden Sie die zuvor erstellten Rollen und Gruppen und weisen Sie IAM-Richtlinien zu, um die korrekte Nutzung zu erzwingen. SCP bietet eine zentrale Kontrolle über die maximal verfügbaren Berechtigungen für alle Konten in Ihrem Unternehmen und stellt sicher, dass Ihre Konten die Vorgaben Ihrer Zugriffskontrollrichtlinien erfüllen. SCPs sind nur in einem Unternehmen verfügbar, für das alle Funktionen aktiviert sind, und Sie

können die SCPs so konfigurieren, dass sie Aktionen für Mitgliedskonten standardmäßig verweigern oder zulassen. Weitere Informationen zur Implementierung der Zugriffsverwaltung finden Sie im Well-Architected Whitepaper zur Säule "Sicherheit" .

Über die Verwaltung von Service Quotas können Sie außerdem Governance implementieren. Indem Sie sicherstellen, dass Service Quotas mit minimalem Overhead definiert und ordnungsgemäß verwaltet werden, können Sie die Ressourcenerstellung über die Geschäftsanforderungen hinaus minimieren. Dazu müssen Sie nachvollziehen, wie schnell sich Ihre Anforderungen ändern können. Sie müssen die derzeit ausgeführten Projekte kennen (in Bezug auf die Erstellung und die Deaktivierung von Ressourcen) und berücksichtigen, wie schnell Kontingentänderungen implementiert werden können. Service Quotas können bei Bedarf verwendet werden, um Ihre Kontingente zu erhöhen.

Implementierungsschritte

- Implementieren von Benachrichtigungen zu Ausgaben: Erstellen Sie mithilfe Ihrer definierten Organisationsrichtlinien AWS-Budgets, um Benachrichtigungen zu erhalten, wenn Ausgaben außerhalb Ihrer Richtlinien liegen. Konfigurieren Sie mehrere Kostenbudgets, eines für jedes Konto, das Sie über die allgemeinen Kontoausgaben informiert. Konfigurieren Sie dann zusätzliche Kostenbudgets innerhalb jedes Kontos für kleinere Einheiten innerhalb des Kontos. Diese Einheiten variieren je nach Kontenstruktur. Einige gängige Beispiele sind AWS-Regionen, Workloads (mithilfe von Tags) oder AWS-Services. Stellen Sie sicher, dass Sie eine E-Mail-Verteilerliste als Empfänger für Benachrichtigungen konfigurieren und nicht das E-Mail-Konto einer Person. Sie können ein tatsächliches Budget für den Fall konfigurieren, dass ein Betrag überschritten wird, oder ein prognostiziertes Budget zur Benachrichtigung über die prognostizierte Nutzung verwenden.
- Implementieren von Nutzungskontrollen: Implementieren Sie mithilfe Ihrer definierten Organisationsrichtlinien IAM-Richtlinien und -Rollen, um anzugeben, welche Aktionen Benutzer ausführen und welche sie nicht ausführen können. In einer AWS-Richtlinie können mehrere Organisationsrichtlinien enthalten sein. Gehen Sie auf die gleiche Art und Weise vor, wie Sie Richtlinien definiert haben. Beginnen Sie umfassend und wenden dann bei jedem Schritt detailliertere Kontrollen an. Service Limits sind auch eine effektive Kontrolle der Nutzung. Implementieren Sie die richtigen Service Limits für alle Ihre Konten.

Ressourcen

Zugehörige Dokumente:

- Von AWS verwaltete Richtlinien für Auftragsfunktionen

- AWS-Fakturierungsstrategie mit mehreren Konten
- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien

Zugehörige Beispiele:

- Well-Architected Labs: Steuerung der Kosten und Nutzung
- Well-Architected Labs: Steuerung der Kosten und Nutzung

COST02-BP06 Verfolgen des Projektlebenszyklus

Verfolgen, bewerten und überprüfen Sie den Lebenszyklus von Projekten, Teams und Umgebungen, damit Sie keine unnötigen Ressourcen nutzen, für die Sie zahlen müssen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Stellen Sie sicher, dass Sie den gesamten Lebenszyklus des Workloads überwachen. Auf diese Weise wird sichergestellt, dass Workloads oder Workload-Komponenten außer Betrieb genommen oder geändert werden können, wenn sie nicht mehr benötigt werden. Dies ist besonders nützlich, wenn Sie neue Services oder Funktionen veröffentlichen. Die vorhandenen Workloads und Komponenten werden zwar u. U. als in Gebrauch angezeigt, sollten aber außer Betrieb genommen werden, um Kunden auf den neuen Service umzuleiten. Beachten Sie frühere Phasen von Workloads – nachdem eine Workload in der Produktion ist, können vorherige Umgebungen außer Betrieb genommen oder stark reduziert werden, bis sie wieder benötigt werden.

AWS bietet eine Reihe von Verwaltungs- und Governance-Services, die Sie für die Entitätslebenszyklus-Verfolgung verwenden können. Sie können AWS Config oder AWS Systems Manager verwenden, um eine detaillierte Bestandsaufnahme Ihrer AWS-Ressourcen und -Konfiguration bereitzustellen. Es wird empfohlen, dass Sie diese mit Ihren vorhandenen Projekt- bzw. Asset-Verwaltungssystemen integrieren, um aktive Projekte und Produkte in Ihrem Unternehmen zu verfolgen. Durch die Kombination Ihres aktuellen Systems mit dem umfassenden Angebot an Ereignissen und Kennzahlen in AWS können Sie eine Ansicht mit signifikanten Lebenszyklus-Ereignissen aufbauen und Ressourcen proaktiv verwalten, um so unnötige Kosten zu reduzieren.

Siehe das Well-Architected Whitepaper zur Säule für die betriebliche Exzellenz .

Implementierungsschritte

- Durchführen von Workload-Überprüfungen: Überprüfen Sie, wie in Ihren Organisationsrichtlinien definiert, Ihre vorhandenen Projekte. Der Aufwand für die Prüfung sollte proportional zum ungefähren Risiko, dem Wert oder den Kosten für die Organisation sein. Wichtige Bereiche, die in die Prüfung aufgenommen werden sollen, sind das Risiko eines Vorfalls oder eines Ausfalls, der Wert oder Beitrag für die Organisation (gemessen am Umsatz oder Ruf der Marke), die Kosten des Workloads (gemessen als Gesamtkosten für Ressourcen und Betriebskosten) und die Nutzung des Workloads (gemessen an der Anzahl der Ergebnisse der Organisation pro Zeiteinheit). Wenn sich diese Bereiche im Laufe des Lebenszyklus ändern, sind Anpassungen des Workloads erforderlich, z. B. die vollständige oder teilweise Außerbetriebnahme.

Ressourcen

Zugehörige Dokumente:

- AWS Config
- AWS Systems Manager
- Von AWS verwaltete Richtlinien für Auftragsfunktionen
- AWS-Fakturierungsstrategie mit mehreren Konten
- Steuern Sie den Zugang zu AWS-Regionen mit IAM-Richtlinien

KOSTEN 3 Wie können Sie die Nutzung und Kosten überwachen?

Definieren Sie Richtlinien und Verfahren, um Ihre Kosten überwachen und richtig zuordnen zu können. Dadurch können Sie die Kosteneffizienz des Workloads bewerten und verbessern.

Bewährte Methoden

- COST03-BP01 Konfigurieren detaillierter Informationsquellen
- COST03-BP02 Identifizieren von Kostenzuordnungskategorien
- COST03-BP03 Definieren von Organisationsmetriken
- COST03-BP04 Konfigurieren von Tools für die Fakturierung und Kostenverwaltung
- COST03-BP05 Hinzufügen von Organisationsinformationen zu Kosten und Nutzung
- COST03-BP06 Zuweisen von Kosten basierend auf Workload-Metriken

COST03-BP01 Konfigurieren detaillierter Informationsquellen

Konfigurieren Sie den AWS-Kosten- und Nutzungsbericht und die stündliche Granularität des Cost Explorer, um detaillierte Kosten- und Nutzungsinformationen bereitzustellen. Konfigurieren Sie Ihren Workload so, dass Protokolleinträge für jedes bereitgestellte Geschäftsergebnis vorhanden sind.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Aktivieren Sie die stündliche Granularität im AWS Cost Explorer und erstellen Sie einen AWS Cost and Usage Report (CUR). Diese Datenquellen bieten die genaueste Ansicht der Kosten und Nutzung in Ihrem gesamten Unternehmen. Der CUR bietet tägliche oder stündliche Nutzungsaufschlüsselung, Tarife, Kosten und Nutzungsattribute für alle kostenpflichtigen AWS-Services. Alle möglichen Dimensionen befinden sich im CUR, einschließlich: Tagging, Speicherort, Ressourcenattribute und Konto-IDs.

Konfigurieren Sie Ihren CUR mit den folgenden Anpassungen:

- Ressourcen-IDs einschließen
- Automatische Aktualisierung des CUR
- Stündliche Granularität
- Versionsverwaltung: Vorhandenen Bericht überschreiben
- Datenintegration: Amazon Athena (Parquet-Format und -Komprimierung)

Verwendung Sie AWS Glue , um die Daten für die Analyse vorzubereiten, und verwenden Sie Amazon Athena für die Datenanalyse mit SQL als Abfragesprache für die Daten. Sie können auch Amazon QuickSight verwenden, um benutzerdefinierte und komplexe Visualisierungen zu erstellen und diese in Ihrem gesamten Unternehmen zu verteilen.

Implementierungsschritte

- Konfigurieren des Kosten- und -Nutzungsberichts: Konfigurieren Sie über die Fakturierungskonsole mindestens einen Kosten- und Nutzungsbericht. Konfigurieren Sie einen Bericht mit stündlicher Granularität, der alle IDs und Ressourcen-IDs enthält. Sie können auch andere Berichte mit unterschiedlichen Granularitäten erstellen, um zusammenfassende Informationen bereitzustellen.
- Konfigurieren der stündlichen Granularität im Cost Explorer: Aktivieren Sie Hourly and Resource Level Data über die Fakturierungskonsole.

> **Note**
>
> Mit der Aktivierung dieser Funktion fallen Kosten an. Weitere Informationen finden Sie in den Preisen.

- Konfigurieren der Anwendungsprotokollierung: Überprüfen Sie, dass Ihre Anwendung jedes Geschäftsergebnis protokolliert, das sie liefert, sodass es nachverfolgt und gemessen werden kann. Stellen Sie sicher, dass die Granularität dieser Daten mindestens stündlich ist, um mit den Kosten- und Nutzungsdaten übereinzustimmen. Siehe Well-Architected: Säule „operative Exzellenz“ für weitere Details für Protokollierung und Überwachung.

Ressourcen

Zugehörige Dokumente:

- AWS-Kontoeinrichtung
- AWS Cost and Usage Report (CUR)
- AWS Glue
- Amazon QuickSight
- Preisberechnung des AWS-Kostenmanagements
- Tagging von AWS-Ressourcen
- Analysieren Ihrer Kosten mit AWS Budgets
- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten
- Well-Architected: Säule „operative Exzellenz“

Zugehörige Beispiele:

- AWS-Kontoeinrichtung

COST03-BP02 Identifizieren von Kostenzuordnungskategorien

Ermitteln Sie Organisationskategorien, die für die Kostenzuordnung innerhalb Ihrer Organisation genutzt werden können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Arbeiten Sie mit Ihrem Finanzteam und anderen relevanten Beteiligten zusammen, um zu verstehen, wie die Kosten innerhalb Ihres Unternehmens zugeordnet werden müssen. Workload-Kosten müssen über den gesamten Lebenszyklus hinweg zugeordnet werden, einschließlich Entwicklung, Tests, Produktion und Außerbetriebnahme. Analysieren Sie, welche Kosten durch Schulungen, Personalentwicklung und Ideenentwicklung im Unternehmen entstehen. Dies kann hilfreich sein, um Konten, die zu diesem Zweck verwendet werden, korrekt den Schulungs- und Entwicklungsbudgets zuzuordnen, anstatt allgemeinen IT-Kostenbudgets.

Implementierungsschritte

- Definieren der Organisationskategorien: Treffen Sie Beteiligte, um Kategorien zu definieren, die die Struktur und Anforderungen Ihrer Organisation widerspiegeln. Diese werden direkt der Struktur vorhandener Finanzkategorien zugeordnet, z. B. Geschäftsbereich, Budget, Kostenstelle oder Abteilung. Sehen Sie sich die Ergebnisse an, die die Cloud für Ihr Unternehmen bietet, z. B. Schulungen oder Fortbildungen, da es sich auch um Organisationskategorien handelt. Einer Ressource können mehrere Kategorien zugewiesen werden. Eine Ressource kann sich in mehreren verschiedenen Kategorien befinden. Definieren Sie daher so viele Kategorien wie nötig.
- Definieren der funktionalen Kategorien: Treffen Sie Beteiligte, um Kategorien zu definieren, die die Funktionen widerspiegeln, die Sie in Ihrem Unternehmen haben. Dabei kann es sich um den Workload- oder Anwendungsnamen und die Art der Umgebung handeln, z. B. Produktion, Test oder Entwicklung. Einer Ressource können mehrere Kategorien zugewiesen werden. Eine Ressource kann sich in mehreren verschiedenen Kategorien befinden. Definieren Sie daher so viele Kategorien wie nötig.

Ressourcen

Zugehörige Dokumente:

- Tagging von AWS-Ressourcen
- Analysieren Ihrer Kosten mit AWS Budgets
- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten

COST03-BP03 Definieren von Organisationsmetriken

Definieren Sie die Organisationsmetriken, die für diesen Workload erforderlich sind. Beispiele für Metriken eines Workloads sind erstellte Kundenberichte oder Webseiten, die den Kunden angezeigt werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Entwickeln Sie ein Verständnis dafür, wie die Ausgabe Ihres Workloads im Vergleich zum Geschäftserfolg gemessen wird. Jeder Workload verfügt in der Regel über einen kleinen Satz von Hauptausgaben, die auf die Leistung hinweisen. Wenn Sie einen komplexen Workload mit vielen Komponenten haben, können Sie die Liste priorisieren oder Metriken für jede Komponente definieren und nachverfolgen. Arbeiten Sie mit Ihren Teams zusammen, um zu verstehen, welche Metriken verwendet werden sollen. Diese Einheit wird verwendet, um die Effizienz des Workloads oder die Kosten für die einzelnen Geschäftsausgaben zu verstehen.

Implementierungsschritte

- Definieren von Workload-Ergebnissen: Treffen Sie sich mit den Beteiligten im Unternehmen und definieren Sie die Ergebnisse für den Workload. Hierbei handelt es sich um eine primäre Maßnahme für die Kundennutzung. Es müssen Geschäftsmetriken und keine technischen Metriken gemessen werden. Es sollte eine kleine Anzahl von High-Level-Metriken (weniger als fünf) pro Workload geben. Wenn der Workload mehrere Ergebnisse für verschiedene Anwendungsfälle erzeugt, gruppieren Sie sie in einer einzigen Metrik.
- Definieren der Ergebnisse von Workload-Komponenten: Wenn Sie einen großen und komplexen Workload haben oder Ihren Workload problemlos in Komponenten (z. B. Microservices) mit gut definierten Ein- und Ausgaben aufteilen können, definieren Sie optional Metriken für jede Komponente. Der Aufwand sollte den Wert und die Kosten der Komponente widerspiegeln. Beginnen Sie mit den größten Komponenten und arbeiten Sie sich zu den kleineren Komponenten vor.

Ressourcen

Zugehörige Dokumente:

- Tagging von AWS-Ressourcen
- Analysieren Ihrer Kosten mit AWS Budgets

- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten

COST03-BP04 Konfigurieren von Tools für die Fakturierung und Kostenverwaltung

Konfigurieren Sie AWS Cost Explorer und AWS Budgets in Übereinstimmung mit Ihren Organisationsrichtlinien.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Um die Nutzung zu ändern und die Kosten anzupassen, muss jede Person in Ihrem Unternehmen Zugriff auf ihre Kosten- und Nutzungsinformationen haben. Es wird empfohlen, dass für alle Workloads und Teams die folgenden Tools konfiguriert sind, wenn sie die Cloud verwenden:

- Berichte: Fassen Sie alle Kosten- und Nutzungsinformationen zusammen.
- Benachrichtigungen: Stellen Sie Benachrichtigungen bereit, wenn Kosten oder Nutzung außerhalb der festgelegten Limits liegen.
- Aktueller Status: Konfigurieren Sie ein Dashboard mit aktuellen Kosten- und Nutzungsraten. Das Dashboard sollte an einem gut sichtbaren Ort innerhalb der Arbeitsumgebung verfügbar sein (ähnlich wie bei einem Betriebs-Dashboard).
- Trending: Stellen Sie die Möglichkeit bereit, die Variabilität von Kosten und Nutzung über den erforderlichen Zeitraum mit der erforderlichen Aufschlüsselung anzuzeigen.
- Prognosen: Stellen Sie die Möglichkeit bereit, geschätzte zukünftige Kosten anzuzeigen.
- Nachverfolgung: Stellen Sie die aktuellen Kosten und die aktuelle Nutzung in Bezug zu den konfigurierten Zielen oder Vorgaben dar.
- Analyse: Stellen Sie Teammitgliedern die Möglichkeit bereit, benutzerdefinierte und tiefgreifende Analysen bis hin zur Stundenaufschlüsselung mit allen möglichen Dimensionen durchzuführen.

Sie können AWS-native Tools wie AWS Cost Explorer, AWS Budgetsund Amazon Athena mit Amazon QuickSight verwenden, um diese Funktion bereitzustellen. Sie können auch Tools von Drittanbietern verwenden. Sie müssen jedoch sicherstellen, dass diese Tools Ihrem Unternehmen einen Mehrwert bieten und somit ihre Kosten rechtfertigen.

Implementierungsschritte

- Erstellen einer Kostenoptimierungsgruppe: Konfigurieren Sie Ihr Konto und erstellen Sie eine Gruppe, die Zugriff auf die erforderlichen Kosten- und Nutzungsberichte hat. Diese Gruppe muss Vertreter aller Teams umfassen, die eine Anwendung besitzen oder verwalten. Auf diese Weise wird sichergestellt, dass jedes Team Zugriff auf seine Kosten- und Nutzungsinformationen hat.
- Konfigurieren von AWS Budgets: Konfigurieren Sie AWS Budgets für alle Konten Ihres Workloads. Legen Sie ein Budget für die Gesamtkontoausgaben und ein Budget für den Workload mithilfe von Tags fest.
- Konfigurieren von AWS Cost Explorer: Konfigurieren Sie AWS Cost Explorer für Ihren Workload und Ihre Konten. Erstellen Sie ein Dashboard für den Workload, das die Gesamtausgaben und die wichtigsten Nutzungsmetriken für den Workload aufzeichnet.
- Konfigurieren fortgeschrittener Tools: Optional können Sie benutzerdefinierte Tools für Ihre Organisation erstellen, die zusätzliche Details und Granularität bieten. Sie können fortgeschrittene Analysefunktionen mithilfe von Amazon Athenaund Dashboards mit Amazon QuickSightimplementieren.

Ressourcen

Zugehörige Dokumente:

- Tagging von AWS-Ressourcen
- Analysieren Ihrer Kosten mit AWS Budgets
- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten

Zugehörige Beispiele:

- Well-Architected Labs – AWS-Kontoeinrichtung
- Well-Architected Labs: Fakturierungsvisualisierung
- Well-Architected Labs: Kosten und Steuerung der Nutzung
- Well-Architected Labs: Analyse der Kosten und Nutzung
- Well-Architected Labs: Visualisierung der Kosten und Nutzung

COST03-BP05 Hinzufügen von Organisationsinformationen zu Kosten und Nutzung

Definieren Sie ein Markierungsschema auf Basis der Organisation sowie Workload-Attribute und Kostenzuordnungskategorien. Implementieren von Markierungen für alle Ressourcen. Verwenden Sie Cost Categories, um Kosten und Nutzung nach Organisationsattributen zu gruppieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Implementieren Sie Tagging in AWS, um Unternehmensinformationen zu Ihren Ressourcen hinzuzufügen, die dann zu Ihren Kosten- und Nutzungsinformationen hinzugefügt werden. Ein Tag ist ein Schlüssel-Wert-Paar – der Schlüssel ist definiert und muss innerhalb Ihres Unternehmens eindeutig sein und der Wert ist für eine Gruppe von Ressourcen eindeutig. Ein Beispiel für ein Schlüssel-Wert-Paar ist der Schlüssel „Umgebung“ mit dem Wert „Produktion“. Alle Ressourcen in der Produktionsumgebung verfügen über dieses Schlüssel-Wert-Paar. Mit dem Markieren können Sie Ihre Kosten mit aussagekräftigen, relevanten Unternehmensinformationen kategorisieren und nachverfolgen. Sie können Tags anwenden, die Unternehmenskategorien (z. B. Kostenstellen, Anwendungsnamen, Projekte oder Besitzer) darstellen und Workloads und Merkmale von Workloads (z. B. Test oder Produktion) identifizieren, um Ihre Kosten und Nutzung in Ihrem gesamten Unternehmen zuzuordnen.

Wenn Sie Tags auf Ihre AWS-Ressourcen anwenden (z. B. Amazon Elastic Compute Cloud-Instances oder Amazon Simple Storage Service-Buckets) und die Tags aktivieren, fügt AWS diese Informationen zu Ihren Kosten- und Nutzungsberichten hinzu. Sie können Berichte ausführen und Analysen für getaggte und nicht getaggte Ressourcen durchführen, um eine größere Compliance mit internen Kostenverwaltungsrichtlinien zu ermöglichen und eine genaue Zuordnung zu gewährleisten.

Mit der Erstellung und Implementierung eines AWS-Tagging-Standard für alle Konten in Ihrem Unternehmen können Sie Ihre AWS-Umgebungen auf konsistente und einheitliche Weise verwalten und steuern. Verwendung Sie Tag-Richtlinien in AWS Organizations, um Regeln für die Verwendung von Tags für AWS-Ressourcen in Ihren Konten in AWS Organizations zu definieren. Mit Tag-Richtlinien können Sie problemlos einen standardisierten Ansatz für das Taggen von AWS-Ressourcen anwenden.

Mit dem AWS Tag Editor können Sie Tags für mehrere Ressourcen hinzufügen, löschen und verwalten.

Mit AWS Cost Categories können Sie Ihren Kosten eine Unternehmensbedeutung zuweisen, ohne dass Tags für Ressourcen erforderlich sind. Sie können Ihre Kosten- und Nutzungsinformationen

eindeutigen internen Unternehmensstrukturen zuordnen. Sie definieren Kategorieregeln, um Kosten mithilfe von Fakturierungsdimensionen wie Konten und Tags zuzuordnen und zu kategorisieren. Dies bietet zusätzlich zum Tagging eine weitere Ebene der Verwaltungsfunktionen. Sie können auch bestimmte Konten und Tags mehreren Projekten zuordnen.

Implementierungsschritte

- Definieren eines Markierungsschemas: Sammeln Sie alle Beteiligten aus Ihrem gesamten Unternehmen, um ein Schema zu definieren. Dies umfasst in der Regel Techniker, Finanz- und Managementkräfte. Definieren Sie eine Liste der Tags, die alle Ressourcen haben müssen, sowie eine Liste der Tags, die Ressourcen haben sollten. Stellen Sie sicher, dass die Tag-Namen und -Werte in Ihrer Organisation konsistent sind.
- Markieren von Ressourcen: Platzieren Sie mithilfe Ihrer definierten Kostenzuordnungskategorien Tags für alle Ressourcen in Ihren Workloads entsprechend den Kategorien. Verwenden Sie Tools wie CLI, Tag Editor oder Systems Manager, um die Effizienz zu steigern.
- Implementieren von Cost Categories: Sie können Cost Categories erstellen, ohne Tags zu implementieren. Cost Categories verwenden die vorhandenen Kosten- und Nutzungsdimensionen. Erstellen Sie Kategorieregeln aus Ihrem Schema und implementieren Sie es in Cost Categories.
- Automatisiertes Markieren: Automatisieren Sie das Markieren, um sicherzustellen, dass Sie ein hohes Maß an Markierungen für alle Ressourcen aufrechterhalten, damit Ressourcen automatisch bei ihrer Erstellung markiert werden. Verwenden Sie die Funktionen innerhalb des Services oder Services wie AWS CloudFormation, um sicherzustellen, dass Ressourcen bei der Erstellung markiert werden. Sie können auch einen benutzerdefinierten Microservice erstellen, der den Workload regelmäßig scannt und alle Ressourcen entfernt, die nicht markiert sind. Dies ist ideal für Test- und Entwicklungsumgebungen.
- Überwachung von und Berichterstellung zu Tags: Um sicherzustellen, dass Sie in Ihrer Organisation ein hohes Maß an Markierungen aufrechterhalten, melden und überwachen Sie die Tags in Ihren Workloads. Sie können AWS Cost Explorer verwenden, um die Kosten für markierte und nicht markierte Ressourcen anzuzeigen. Alternativ können Sie auch Services wie Tag Editor verwenden. Überprüfen Sie regelmäßig die Anzahl der nicht markierten Ressourcen und ergreifen Sie Maßnahmen, um Tags hinzuzufügen, bis Sie die gewünschte Markierungsstufe erreichen.

Ressourcen

Zugehörige Dokumente:

- AWS CloudFormation-Ressource-Tag

- Mit AWS Cost Categories
- Tagging von AWS-Ressourcen
- Amazon EC2 und Amazon EBS fügen Support für das Tagging von Ressourcen bei der Erstellung hinzu.
- Analysieren Ihrer Kosten mit AWS Budgets
- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten

COST03-BP06 Zuweisen von Kosten basierend auf Workload-Metriken

Ordnen Sie die Kosten des betreffenden Workloads anhand von Metriken oder geschäftlichen Ergebnissen zu, um die Kosteneffizienz des Workloads zu bewerten. Implementieren Sie einen Prozess für die Analyse des AWS-Kosten- und Nutzungsberichts mit Amazon Athena,um von genaueren Einblicken und Rückbelastungsmöglichkeiten zu profitieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Die Kostenoptimierung liefert Geschäftsergebnisse zum niedrigsten Preis, was nur durch Zuweisung von Workload-Kosten nach Workload-Metriken (gemessen nach Workload-Effizienz) erreicht werden kann. Überwachen Sie die definierten Workload-Metriken durch Protokolldateien oder andere Anwendungsüberwachung. Kombinieren Sie diese Daten mit den Workload-Kosten, die Sie erhalten können, indem Sie Kosten mit einem bestimmten Tag-Wert oder einer Konto-ID betrachten. Es wird empfohlen, diese Analyse auf Stundenbasis durchzuführen. Ihre Effizienz ändert sich in der Regel, wenn Sie statische Kostenkomponenten haben (z. B. eine Backend-Datenbank, die rund um die Uhr ausgeführt wird) mit einer variierenden Anfragerate (z. B. Nutzungsspitzen um 9–17 Uhr, mit wenigen Anfragen in der Nacht). Wenn Sie die Beziehung zwischen den statischen und variablen Kosten verstehen, können Sie Ihre Optimierungsaktivitäten fokussieren.

Implementierungsschritte

- Zuweisen von Kosten zu Workload-Metriken: Erstellen Sie mit den definierten Metriken und konfigurierten Markierungen eine Metrik, die die Workload-Ausgabe und die Workload-Kosten kombiniert. Verwenden Sie Analyse-Services wie Amazon Athena und Amazon QuickSight, um ein Effizienz-Dashboard für den gesamten Workload und alle Komponenten zu erstellen.

Ressourcen

Zugehörige Dokumente:

- Tagging von AWS-Ressourcen
- Analysieren Ihrer Kosten mit AWS Budgets
- Analysieren Ihrer Kosten mit Cost Explorer
- Verwalten von AWS-Kosten- und -Nutzungsberichten

KOSTEN 4 Wie können Sie Ressourcen außer Betrieb nehmen?

Implementieren Sie vom Beginn bis zum Abschluss eines Projekts eine Änderungskontrolle und Ressourcenverwaltung. Auf diese Weise können Sie ungenutzte Ressourcen herunterfahren oder beenden, um Verschwendungen zu minimieren.

Bewährte Methoden

- COST04-BP01 Nachverfolgen von Ressourcen über ihre Lebensdauer
- COST04-BP02 Implementieren von Außerbetriebnahmeprozess
- COST04-BP03 Außerbetriebnahme von Ressourcen
- COST04-BP04 Automatische Stilllegung von Ressourcen

COST04-BP01 Nachverfolgen von Ressourcen über ihre Lebensdauer

Definieren und implementieren Sie eine Methode zur Verfolgung von Ressourcen und deren Verknüpfungen mit Systemen über ihre gesamte Lebensdauer hinweg. Mit einer entsprechenden Markierung können Sie den Workload oder die Funktion der Ressource identifizieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Nicht mehr benötigte Workload-Ressourcen werden außer Betrieb genommen. Ein gängiges Beispiel sind Ressourcen, die zum Testen verwendet werden. Nach Abschluss des Tests können die Ressourcen entfernt werden. Das Nachverfolgen von Ressourcen mit Tags (und Ausführen von Berichten zu diesen Tags) hilft Ihnen, Komponenten zu identifizieren, die außer Betrieb genommen werden sollen. Die Verwendung von Tags ist eine effektive Möglichkeit, Ressourcen zu verfolgen,

indem die Ressource mit ihrer Funktion oder einem bekannten Datum, an dem sie außer Betrieb genommen werden kann, gekennzeichnet wird. Berichte können dann zu diesen Tags ausgeführt werden. Beispielwerte für das Markieren von Funktionen sind `„feature-X-Testing“,` um den Zweck der Ressource in Bezug auf den Workload-Lebenszyklus zu identifizieren.

Implementierungsschritte

- Implementieren eines Markierungsschemas: Implementieren Sie ein Markierungsschema, das den Workload identifiziert, zu dem die Ressource gehört, und stellen Sie sicher, dass alle Ressourcen innerhalb des Workloads entsprechend markiert sind.
- Implementieren des Workload-Durchsatzes oder der Ausgabekontrolle: Implementieren Sie die Überwachung des Workload-Durchsatzes oder die Ausgabe von Alarmsignalen, die entweder bei der Eingabe oder Ausgabe ausgelöst werden. Konfigurieren Sie die Überwachung so, dass Benachrichtigungen erstellt werden, wenn Workload-Anforderungen oder -Ausgaben auf Null fallen. Dies bedeutet, dass die Workload-Ressourcen nicht mehr verwendet werden. Integrieren Sie einen Zeitfaktor, wenn der Workload unter normalen Bedingungen regelmäßig auf Null fällt.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Trusted Advisor
- Tagging von AWS-Ressourcen
- Veröffentlichen benutzerdefinierter Metriken

COST04-BP02 Implementieren von Außerbetriebnahmeprozess

Implementieren Sie einen Prozess für die Identifizierung und Außerbetriebnahme von verwaisten Ressourcen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Implementieren Sie einen standardisierten Prozess in Ihrem gesamten Unternehmen, um ungenutzte Ressourcen zu identifizieren und zu entfernen. Der Prozess sollte definieren, wie häufig

Suchvorgänge durchgeführt werden, und die Prozesse zum Entfernen der Ressource festlegen, um sicherzustellen, dass alle Unternehmensanforderungen erfüllt sind.

Implementierungsschritte

- Erstellen und Implementieren eines Prozesses für die Außerbetriebnahme: Erstellen Sie in Zusammenarbeit mit den Workload-Entwicklern und -Besitzern einen Prozess zur Außerbetriebnahme des Workloads und seiner Ressourcen. Der Prozess sollte die Methode abdecken, um zu überprüfen, ob der Workload verwendet wird, und auch, ob jede der Workload-Ressourcen verwendet wird. Der Prozess deckt auch die Schritte ab, die erforderlich sind, um die Ressource außer Betrieb zu nehmen und gleichzeitig die Einhaltung gesetzlicher Anforderungen sicherzustellen. Alle zugeordneten Ressourcen sind ebenfalls abgedeckt, z. B. Lizenzen oder dazugehöriger Speicher. Der Prozess soll die Besitzer des Workloads darüber informieren, dass die Außerbetriebnahme ausgeführt wurde.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Trusted Advisor

COST04-BP03 Außerbetriebnahme von Ressourcen

Außerbetriebnahme von Ressourcen, die durch Ereignisse wie regelmäßige Prüfungen oder Änderungen der Nutzung ausgelöst werden. Die Außerbetriebnahme erfolgt in der Regel regelmäßig und erfolgt manuell oder automatisiert.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Die Häufigkeit und der Aufwand für die Suche nach ungenutzten Ressourcen sollten die potenziellen Einsparungen widerspiegeln, sodass ein Konto mit geringen Kosten seltener analysiert werden sollte als ein Konto mit größeren Kosten. Suchanfragen und Außerbetriebnahmeereignisse können durch Statusänderungen im Workload ausgelöst werden, z. B. ein Produkt, das sich dem Ende seiner Lebensdauer nähert oder ersetzt wird. Suchen und Außerbetriebnahme können auch durch externe Ereignisse ausgelöst werden, wie z. B. Änderungen der Marktbedingungen oder Produktterminierung.

Implementierungsschritte

- Ressourcen außer Betrieb nehmen: Verwenden Sie den Außerbetriebnahme-Prozess, um jede der Ressourcen, die als verwaist identifiziert wurde, außer Betrieb zu nehmen.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Trusted Advisor

COST04-BP04 Automatische Stilllegung von Ressourcen

Gestalten Sie Ihren Workload so, dass er die Beendigung von Ressourcen reibungslos handhabt, wenn Sie unkritische Ressourcen, nicht benötigte Ressourcen oder Ressourcen mit geringer Auslastung identifizieren und außer Betrieb nehmen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Verwenden Sie die Automatisierung, um die damit verbundenen Kosten für die Außerbetriebnahme zu reduzieren oder zu entfernen. Wenn Sie Ihren Workload so konzipieren, dass er eine automatische Außerbetriebnahme durchführt, werden die gesamten Workload-Kosten während der Nutzungsdauer gesenkt. Sie können AWS Auto Scaling verwenden, um die Außerbetriebnahme durchzuführen. Sie können auch benutzerdefinierten Code mithilfe der API oder des SDK implementieren, um Workload-Ressourcen automatisch außer Betrieb zu nehmen.

Implementierungsschritte

- Implementieren von AWS Auto Scaling: Konfigurieren Sie unterstützte Ressourcen mit AWS Auto Scaling
- Konfigurieren von CloudWatch zum Beenden von Instances: Instances können so konfiguriert werden, dass sie mit CloudWatch-Alarmen beendet werden. Implementieren Sie mithilfe der Metriken aus dem Außerbetriebnahmeprozess einen Alarm mit einer Amazon Elastic Compute Cloud-Aktion (Amazon EC2). Stellen Sie sicher, dass Sie den Vorgang in einer Nicht-Produktionsumgebung überprüfen, bevor Sie den Vorgang ausführen.
- Implementieren von Code innerhalb des Workloads: Sie können das AWS SDK oder die AWS CLI verwenden, um Workload-Ressourcen außer Betrieb zu nehmen. Implementieren Sie Code

innerhalb der in AWS integrierten Anwendung, die nicht mehr verwendete Ressourcen beendet oder entfernt.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Trusted Advisor
- Create Alarms to Stop, Terminate, Reboot, or Recover an Instance (Alarme erstellen, um eine Instance zu stoppen, zu beenden, neu zu starten oder wiederherzustellen)
- Erste Schritte mit Amazon EC2 Auto Scaling

Kostengünstige Ressourcen

Fragen

- KOSTEN 5 Wie können Sie die Kosten bei der Auswahl von Services einschätzen?
- KOSTEN 6 Wie können Sie bei der Auswahl des Ressourcentyps, -umfangs und der Anzahl der Ressourcen Kostenziele erfüllen?
- KOSTEN 7 Wie können Sie Kosten mithilfe von Preismodellen senken?
- KOSTEN 8 Wie können Sie die Kosten für Datenübertragungen planen?

KOSTEN 5 Wie können Sie die Kosten bei der Auswahl von Services einschätzen?

Bei Amazon EC2, Amazon EBS und Amazon S3 handelt es sich um AWS-Services, die als einzelne Bausteine angeboten werden. Verwaltete Services, etwa Amazon RDS und Amazon DynamoDB, sind AWS-Services auf einer höheren Ebene oder Anwendungsebene. Wenn Sie sich für die richtigen Bausteine und verwalteten Services entscheiden, können Sie die Kosten dieses Workloads optimieren. Durch die Nutzung von verwalteten Services können Sie einen Großteil Ihres administrativen und betrieblichen Overheads reduzieren oder beseitigen und damit Kapazitäten für anwendungs- und geschäftsbezogene Aktivitäten gewinnen.

Bewährte Methoden

- COST05-BP01 Ermitteln der Organisationsanforderungen zur Kosteneinschätzung
- COST05-BP02 Analysieren sämtlicher Komponenten dieses Workloads

- COST05-BP03 Durchführen einer gründlichen Analyse der einzelnen Komponenten
- COST05-BP04 Auswahl von Software mit kostengünstiger Lizenzierung
- COST05-BP05 Auswahl von Komponenten dieses Workloads zur Optimierung der Kosten im Einklang mit den Prioritäten der Organisation
- COST05-BP06 Durchführen einer Kostenanalyse für unterschiedliche Nutzungen im Lauf der Zeit

COST05-BP01 Ermitteln der Organisationsanforderungen zur Kosteneinschätzung

Definieren Sie gemeinsam mit den Teammitgliedern für diesen Workload das Gleichgewicht zwischen Kostenoptimierung und anderen Säulen wie Leistung und Zuverlässigkeit.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Bei der Auswahl von Services für Ihren Workload ist es wichtig, dass Sie die Prioritäten Ihres Unternehmens verstehen. Stellen Sie ein Gleichgewicht zwischen Kosten und anderen Well-Architected-Säulen wie Leistung und Zuverlässigkeit sicher. Ein vollständig kostenoptimierter Workload ist die Lösung, die am meisten an den Anforderungen Ihres Unternehmens ausgerichtet ist, nicht notwendigerweise an den niedrigsten Kosten. Treffen Sie sich mit allen Teams innerhalb Ihres Unternehmens, um Informationen zu sammeln, z. B. mit den Produkt-, Geschäfts-, Technik- und Finanz-Teams.

Implementierungsschritte

- Ermitteln der Organisationsanforderungen zur Kosteneinschätzung: Treffen Sie sich mit Teammitgliedern aus Ihrem Unternehmen, darunter Produktmanagement, Anwendungsbesitzern, Entwicklungs- und Betriebsteams, Management und Finanzen. Priorisieren Sie die Well-Architected-Säulen für diesen Workload und seine Komponenten. Die Ausgabe erfolgt als Liste mit den Säulen in der entsprechenden Reihenfolge. Sie können auch jeweils eine Gewichtung hinzufügen. Diese kann angeben, wie viel zusätzlicher Fokus auf einer Säule liegt oder wie ähnlich der Fokus zwischen zwei Säulen ist.

Ressourcen

Zugehörige Dokumente:

- AWS-Gesamtbetriebskostenrechner (Total Cost of Ownership, TCO)

- Amazon S3-Speicherklassen
- Cloud-Produkte

COST05-BP02 Analysieren sämtlicher Komponenten dieses Workloads

Stellen Sie sicher, dass jede Workload-Komponente unabhängig von der derzeitigen Größe oder den aktuellen Kosten analysiert wird. Der Überprüfungsaufwand sollte in einem angemessenen Verhältnis zu dem potenziellen Nutzen stehen, z. B. bei einer Prüfung der derzeitigen und prognostizierten Kosten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Führen Sie eine gründliche Analyse aller Komponenten in Ihrem Workload durch. Stellen Sie ein Gleichgewicht zwischen den Analysekosten und den potenziellen Einsparungen im Workload über dessen Lebenszyklus hinweg sicher. Sie müssen die aktuellen und potenziellen zukünftigen Auswirkungen der Komponente ermitteln. Wenn zum Beispiel die Kosten der vorgeschlagenen Ressource 10 USD/Monat betragen und bei prognostizierter Belastung 15 USD/Monat nicht überschreiten würden, könnte ein Tag Aufwand, um die Kosten um 50 % zu reduzieren (5 USD pro Monat), den potenziellen Nutzen über die Lebensdauer des Systems übersteigen. Durch eine schnellere und effizientere datenbasierte Schätzung wird das beste Gesamtergebnis für diese Komponente erzielt.

Workloads können sich im Laufe der Zeit ändern. Die richtigen Services sind möglicherweise nicht optimal, wenn sich die Workload-Architektur oder -Nutzung ändert. Die Analyse für die Auswahl von Services muss aktuelle und zukünftige Workload-Zustände und Nutzungsebenen umfassen. Die Implementierung eines Service für den zukünftigen Workload-Status oder die Nutzung kann die Gesamtkosten senken, indem der Aufwand reduziert oder beseitigt wird, der für zukünftige Änderungen erforderlich ist.

AWS Cost Explorer und AWS Cost and Usage Report (CUR) können die Kosten eines Machbarkeitsnachweises (Proof of Concept, PoC) oder einer laufenden Umgebung analysieren. Sie können AWS Pricing Calculator zur Schätzung der Workload-Kosten nutzen.

Implementierungsschritte

- Auflisten der Workload-Komponenten: Erstellen Sie die Liste aller Workload-Komponenten. Diese wird als Verifizierung verwendet, um zu überprüfen, ob jede Komponente analysiert wurde. Der Aufwand sollte die Kritikalität für den Workload widerspiegeln, die durch die Prioritäten Ihrer

Organisation definiert wird. Die Gruppierung von Ressourcen verbessert die Effizienz, z. B. die Speicherung von Produktionsdatenbanken, wenn es mehrere Datenbanken gibt.

- Priorisieren der Komponentenliste: Priorisieren Sie die Komponentenliste nach Aufwand. In der Regel erfolgt die Priorisierung nach den Kosten der Komponente – von der teuersten zur günstigsten. Alternativ kann sie auch nach der von den Prioritäten Ihrer Organisation definierten Kritikalität erfolgen.
- Durchführen der Analyse: Überprüfen Sie für jede Komponente auf der Liste die verfügbaren Optionen und Services und wählen Sie die Option aus, die am besten mit Ihren Organisationsprioritäten übereinstimmt.

Ressourcen

Zugehörige Dokumente:

- AWS Pricing Calculator
- AWS Cost Explorer
- Amazon S3-Speicherklassen
- Cloud-Produkte

COST05-BP03 Durchführen einer gründlichen Analyse der einzelnen Komponenten

Nehmen Sie die Gesamtkosten, die der Organisation durch die einzelnen Komponenten entstehen, unter die Lupe. Betrachten Sie die Gesamtbetriebskosten unter Berücksichtigung der Betriebs- und Verwaltungskosten, insbesondere bei der Nutzung von verwalteten Services. Der Überprüfungsaufwand sollte in einem angemessenen Verhältnis zum potenziellen Nutzen stehen, z. B. muss die Zeit, die für die Analyse benötigt wird, den Komponentenkosten entsprechen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Bedenken Sie die Zeitersparnis, die es Ihrem Team ermöglicht, sich auf das Aufholen technischen Rückstands, Innovation und wertschöpfende Funktionen zu konzentrieren. So könnten Sie beispielsweise Ihre lokale Umgebung so schnell wie möglich in die Cloud verlagern und die Optimierung im Nachgang ausführen. Es lohnt sich, die Einsparungen zu untersuchen, die Sie durch den Einsatz von verwalteten Services erzielen könnten, die Lizenzkosten entfernen oder reduzieren. Verwaltete Services eliminieren den betrieblichen und administrativen Aufwand für die Wartung

eines Service, sodass Sie sich auf Innovationen konzentrieren können. Da verwaltete Services in der großen Cloud-Umgebung ausgeführt werden, profitieren Sie hier außerdem von geringeren Kosten pro Transaktion oder Service.

Verwaltete Services weisen in der Regel Attribute auf, die Sie festlegen können, um zu gewährleisten, dass ausreichend Kapazität bereitsteht. Sie müssen diese Attribute festlegen und überwachen, damit Ihre überschüssige Kapazität auf ein Minimum begrenzt und die Leistung maximiert werden. Sie können die Attribute der AWS Managed Services mithilfe der AWS Management Console oder AWS-APIs und SDKs ändern, um den Ressourcenbedarf an den sich ändernden Bedarf anzupassen. So können Sie beispielsweise die Anzahl der Knoten in einem Amazon EMR-Cluster (oder einem Amazon Redshift-Cluster) auf- oder abskalieren.

Außerdem können Sie mehrere Instances in eine AWS-Ressource legen, um eine Nutzung mit höherer Dichte zu aktivieren. Sie können beispielsweise mehrere kleine Datenbanken auf einer einzelnen Amazon Relational Database Service (Amazon RDS) Datenbank-Instance bereitstellen. Mit zunehmendem Wachstum können Sie eine der Datenbanken über einen Snapshot- und Wiederherstellungsprozess auf eine spezielle Amazon RDS-Datenbank-Instance migrieren.

Wenn Sie Workloads auf verwalteten Services bereitstellen, müssen Sie sich mit den Anforderungen für das Anpassen der Service-Kapazität vertraut machen. Diese Anforderungen sind in der Regel Zeit, Aufwand und die Auswirkungen auf den normalen Workload-Betrieb. Die bereitgestellte Ressource muss Zeit für Änderungen einräumen und den erforderlichen Overhead bereitstellen, damit dies möglich ist. Der laufende Aufwand für das Ändern von Services kann praktisch auf null reduziert werden, wenn Sie APIs und SDKs verwenden, die mit System- und Überwachungs-Tools wie Amazon CloudWatch integriert sind.

Amazon RDS, Amazon Redshiftund Amazon ElastiCache stellen einen verwalteten Datenbankservice bereit. Amazon Athena, Amazon EMRund Amazon OpenSearch Service bieten einen verwalteten Analyseservice.

AMS ist ein Service, der die AWS-Infrastruktur für Unternehmenskunden und -partner betreibt. Es bietet eine sichere und konforme Umgebung, in der Sie Ihre Workloads bereitstellen können. AMS verwendet Enterprise-Cloud-Betriebsmodelle mit Automatisierung, damit Sie Ihre Unternehmensanforderungen erfüllen, schneller in die Cloud wechseln und Ihre laufenden Verwaltungskosten senken können.

Implementierungsschritte

- Durchführen einer gründlichen Analyse: Arbeiten Sie anhand der Komponentenliste jede Komponente von der höchsten Priorität bis zur niedrigsten Priorität ab. Führen Sie für die

Komponenten mit höherer Priorität sowie für die teureren Komponenten zusätzliche Analysen durch und bewerten Sie alle verfügbaren Optionen und deren langfristige Auswirkungen. Bewerten Sie bei Komponenten mit niedrigerer Priorität, ob Änderungen in der Nutzung die Priorität der Komponente ändern. Führen Sie anschließend eine Analyse des angemessenen Aufwands durch.

Ressourcen

Zugehörige Dokumente:

- AWS-Gesamtbetriebskostenrechner (Total Cost of Ownership, TCO)
- Amazon S3-Speicherklassen
- Cloud-Produkte

COST05-BP04 Auswahl von Software mit kostengünstiger Lizenzierung

Open-Source-Software eliminiert Softwarelizenzkosten, die in Workloads erhebliche Kosten verursachen können. Wenn lizenzierte Software erforderlich ist, vermeiden Sie Lizenzen, die an beliebige Attribute wie CPUs gebunden sind, und suchen Sie nach Lizenzen, die an die Ausgabe oder Ergebnisse gebunden sind. Die Kosten dieser Lizenzen lassen sich besser auf die von ihnen bereitgestellten Vorteile skalieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Die Kosten für Softwarelizenzen können durch die Verwendung von Open-Source-Software eliminiert werden. Dies kann erhebliche Auswirkungen auf die Workload-Kosten haben, da die Größe des Workloads skaliert wird. Messen Sie die Vorteile von lizenzierter Software anhand der Gesamtkosten, um sicherzustellen, dass Sie den Workload optimiert haben. Modellieren Sie Änderungen bei der Lizenzierung und wie sich diese auf Ihre Workload-Kosten auswirken würden. Wenn ein Anbieter die Kosten Ihrer Datenbanklizenz ändert, untersuchen Sie, wie sich dies auf die Gesamteffizienz Ihres Workloads auswirkt. Berücksichtigen Sie historische Preisankündigungen von Ihren Anbietern für Trends bei Lizenzänderungen in ihren Produkten. Die Lizenzkosten können auch unabhängig vom Durchsatz oder der Nutzung skaliert werden, z. B. Lizenzen, die nach Hardware skaliert werden (CPU-gebundene Lizenzen). Diese Lizenzen sollten vermieden werden, da sich die Kosten ohne entsprechende Ergebnisse schnell erhöhen können.

Implementierungsschritte

- Analyse von Lizenzoptionen: Überprüfen Sie die Lizenzbedingungen der verfügbaren Software. Suchen Sie nach Open-Source-Versionen, die über die erforderliche Funktionalität verfügen, und stellen Sie fest, ob die Vorteile der lizenzierten Software die Kosten überwiegen. Bei günstigen Bedingungen stimmen die Kosten der Software mit ihren Vorteilen überein.
- Analysieren des Softwareanbieters: Überprüfen Sie alle historischen Preise oder Lizenzänderungen des Anbieters. Suchen Sie nach Änderungen, die nicht im Einklang mit den Ergebnissen stehen, wie z. B. Strafen für die Ausführung auf Hardware oder Plattformen bestimmter Anbieter. Achten Sie zudem darauf, wie mögliche Prüfungen und Strafen durchgeführt werden.

Ressourcen

Zugehörige Dokumente:

- AWS-Gesamtbetriebskostenrechner (Total Cost of Ownership, TCO)
- Amazon S3-Speicherklassen
- Cloud-Produkte

COST05-BP05 Auswahl von Komponenten dieses Workloads zur Optimierung der Kosten im Einklang mit den Prioritäten der Organisation

Berücksichtigen Sie bei der Auswahl sämtlicher Komponenten die Kosten. Dies beinhaltet auch die Verwendung von Services auf Anwendungsebene sowie verwalteter Services wie etwa Amazon Relational Database Service (Amazon RDS), Amazon DynamoDB, Amazon Simple Notification Service (Amazon SNS) und Amazon Simple Email Service (Amazon SESzur Reduzierung der Gesamtkosten der Organisation. Verwenden Sie Serverless-Lösungen und Container für die Datenverarbeitung, zum Beispiel AWS Lambda, Amazon Simple Storage Service (Amazon S3) für statische Websites und Amazon Elastic Container Service (Amazon ECS). Minimieren Sie Lizenzkosten, indem Sie Open-Source-Software oder Software ohne Lizenzgebühren verwenden, wie z. B. Amazon Linux für Datenverarbeitungs-Workloads. Alternativ können Sie Datenbanken auch zu Amazon Auroramigrieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Sie können serverlose Services oder Services auf Anwendungsebene wie AWS Lambda, Amazon Simple Queue Service (Amazon SQS), Amazon SNSund Amazon SES. Mit diesen Services

müssen Sie keine Ressourcen mehr verwalten und sie stellen die Funktion der Codeausführung, Warteschlangenservices und Nachrichtenzustellung bereit. Der andere Vorteil besteht darin, dass die Leistung und Kosten entsprechend der Nutzung skaliert werden, was eine effiziente Kostenzuordnung ermöglicht.

Weitere Informationen zu serverlosen Services finden Sie im Whitepaper „Well-Architected Serverless Application Lens“.

Implementierungsschritte

- Auswahl der einzelnen Services zur Kostenoptimierung: Wählen Sie unter Verwendung Ihrer Prioritätenliste und Analyse jede Option aus, die am besten mit Ihren Organisationsprioritäten übereinstimmt.

Ressourcen

Zugehörige Dokumente:

- AWS-Gesamtbetriebskostenrechner (Total Cost of Ownership, TCO)
- Amazon S3-Speicherklassen
- Cloud-Produkte

COST05-BP06 Durchführen einer Kostenanalyse für unterschiedliche Nutzungen im Lauf der Zeit

Workloads können sich im Laufe der Zeit ändern. Einige Services oder Funktionen sind auf unterschiedlichen Nutzungsebenen kostengünstiger. Wenn Sie jede Komponente im zeitlichen Verlauf und mit einer prognostizierten Nutzung analysieren, bleibt dieser Workload über seine gesamte Lebensdauer hinweg kostengünstig.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Wenn AWS neue Services und Funktionen veröffentlicht, können sich die optimalen Services für Ihren Workload ändern. Der erforderliche Aufwand sollte potenzielle Vorteile widerspiegeln. Die Häufigkeit der Workload-Überprüfung hängt von den Anforderungen Ihres Unternehmens ab. Wenn es sich um einen Workload mit erheblichen Kosten handelt, wird die Implementierung neuer Services früher die Kosteneinsparungen maximieren, sodass eine häufigere Überprüfung von Vorteil sein

kann. Ein weiterer Auslöser für die Überprüfung ist die Änderung der Nutzungsmuster. Signifikante Änderungen bei der Nutzung können darauf hinweisen, dass alternative Services optimaler wären. Bei höheren Datenübertragungsraten kann ein Direct Connect-Service beispielsweise günstiger als ein VPN sein und die erforderliche Konnektivität bereitstellen. Prognostizieren Sie die potenziellen Auswirkungen von Service-Änderungen, damit Sie diese Auslöser auf Nutzungsebene überwachen und die kosteneffektivsten Services früher implementieren können.

Implementierungsschritte

- Definieren vorhergesagter Nutzungsmuster: Dokumentieren Sie in Zusammenarbeit mit Unternehmensbereichen, wie z. B. Marketing- und Produktbesitzern, wie die erwarteten und vorausgesagten Nutzungsmuster für den Workload aussehen werden.
- Durchführen einer Kostenanalyse bei vorhergesagter Nutzung: Führen Sie mithilfe der definierten Nutzungsmuster die Analyse an jedem dieser Punkte durch. Der Analyseaufwand sollte das potenzielle Ergebnis widerspiegeln. Wenn beispielsweise die Änderung der Nutzung groß ist, sollte eine gründliche Analyse durchgeführt werden, um etwaige Kosten und Änderungen zu überprüfen.

Ressourcen

Zugehörige Dokumente:

- AWS-Gesamtbetriebskostenrechner (Total Cost of Ownership, TCO)
- Amazon S3-Speicherklassen
- Cloud-Produkte

KOSTEN 6 Wie können Sie bei der Auswahl des Ressourcentyps, -umfangs und der Anzahl der Ressourcen Kostenziele erfüllen?

Stellen Sie sicher, dass Sie den geeigneten Ressourcenumfang und die Anzahl der Ressourcen für die jeweilige Aufgabe auswählen. Durch die Auswahl des kostengünstigsten Typs, Umfangs und der kostengünstigsten Anzahl minimieren Sie die Verschwendung von Ressourcen.

Bewährte Methoden

- COST06-BP01 Durchführen einer Kostenmodellierung
- COST06-BP02 Auswahl von Ressourcentyp, -umfang und -anzahl basierend auf Daten
- COST06-BP03 Auswahl von Ressourcentyp, -umfang und -anzahl basierend auf Metriken

COST06-BP01 Durchführen einer Kostenmodellierung

Ermitteln Sie die Organisationsanforderungen und führen Sie eine Kostenmodellierung des Workloads und ihrer einzelnen Komponenten durch. Führen Sie Benchmark-Aktivitäten für den Workload unter verschiedenen prognostizierten Belastungen durch und vergleichen Sie die Kosten. Der Modellierungsaufwand sollte in einem angemessenen Verhältnis zu dem potenziellen Nutzen stehen, z. B. muss der Zeitaufwand den Komponentenkosten entsprechen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Führen Sie eine Kostenmodellierung für Ihre Workload und jede ihrer Komponenten durch, um das Gleichgewicht zwischen Ressourcen zu verstehen und die richtige Größe für jede Ressource im Workload zu finden, unter Berücksichtigung eines bestimmten Leistungsgrads. Führen Sie Benchmark-Aktivitäten für den Workload unter verschiedenen prognostizierten Belastungen durch und vergleichen Sie die Kosten. Der Modellierungsaufwand sollte in einem angemessenen Verhältnis zu dem potenziellen Nutzen stehen, z. B. muss der Zeitaufwand proportional zu den Komponentenkosten oder prognostizierten Einsparungen sein. Bewährte Methoden finden Sie im Abschnitt „Prüfen“ im Whitepaper zur Säule der Leistungseffizienz.

AWS Compute Optimizer unterstützt Sie bei der Kostenmodellierung für die Ausführung von Workloads. Es bietet Empfehlungen zur richtigen Dimensionierung für Datenverarbeitungsressourcen basierend auf der bisherigen Nutzung. Dies ist die ideale Datenquelle für Datenverarbeitungsressourcen, da es sich um einen kostenlosen Service handelt und er Machine Learning nutzt, um je nach Risikograd mehrere Empfehlungen zu geben. Sie können auch Amazon CloudWatch und Amazon CloudWatch Logs mit benutzerdefinierten Protokollen als Datenquellen für die richtige Dimensionierung anderer Services und Workload-Komponenten verwenden.

Im Folgenden finden Sie Empfehlungen für die Kostenmodellierung von Daten und Metriken:

- Die Überwachung muss die Endbenutzererfahrung genau widerspiegeln. Wählen Sie die richtige Detaillierung für die Dauer aus, und wählen Sie das Maximum oder den 99. Perzentil statt des Durchschnitts aus.
- Wählen Sie die richtige Aufschlüsselung für die Dauer der Analyse aus, die für die Deckung der Workload-Zyklen erforderlich ist. Bei einer zweiwöchigen Analyse könnten Sie beispielsweise einen monatlichen Zyklus mit hoher Nutzung übersehen, der zu einer Unterbereitstellung führen könnte.

Implementierungsschritte

- Durchführen einer Kostenmodellierung: Stellen Sie den Workload oder einen Machbarkeitsnachweis in einem separaten Konto mit den spezifischen zu testenden Ressourcentypen und -umfängen bereit. Führen Sie den Workload mit den Testdaten aus und zeichnen die Ergebnisse zusammen mit den Kostendaten zum Zeitpunkt der Testausführung auf. Stellen Sie anschließend den Workload erneut bereit oder ändern Sie die Ressourcentypen und -umfänge und führen Sie den Test noch einmal aus.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- Amazon CloudWatch-Funktionen
- Kostenoptimierung: Richtige Amazon EC2-Dimensionierung
- AWS Compute Optimizer

COST06-BP02 Auswahl von Ressourcentyp, -umfang und -anzahl basierend auf Daten

Wählen Sie den Ressourcenumfang oder -typ basierend auf Daten zum Workload und der Ressourcenmerkmale aus. Zu berücksichtigen sind hier beispielsweise Datenverarbeitung, Speicher, Durchsatz oder Schreibintensität. Diese Schätzung erfolgt in der Regel unter Verwendung einer früheren (On-Premises)-Version des Workloads, der Dokumentation oder anderer Informationsquellen über den Workload.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Wählen Sie den Ressourcenumfang oder -typ auf Basis des Workloads und der Ressourcenmerkmale aus; zu berücksichtigen sind hier beispielsweise Datenverarbeitung, Speicher, Durchsatz oder Schreibintensität. Diese Auswahl erfolgt in der Regel unter Verwendung der Kostenmodellierung, einer früheren Version des Workloads (z. B. eine On-Premises-Version), mithilfe der Dokumentation oder unter Verwendung anderer Informationsquellen über den Workload (Whitepaper, veröffentlichte Lösungen).

Implementierungsschritte

- Auswahl von Ressourcen basierend auf Daten: Wählen Sie anhand Ihrer Kostenmodelldaten den erwarteten Workload-Nutzungsgrad aus und dann den angegebenen Ressourcentyp und den -umfang.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- Amazon CloudWatch-Funktionen
- Kostenoptimierung: Richtige EC2-Dimensionierung

COST06-BP03 Auswahl von Ressourcentyp, -umfang und -anzahl basierend auf Metriken

Nutzen Sie Metriken aus dem derzeit aktiven Workload für die Auswahl des richtigen Umfangs und Typs, um Kosten zu optimieren. Sorgen Sie für die richtige Bereitstellung von Durchsatz, Umfang und Speicher für Services wie Amazon Elastic Compute Cloud (Amazon EC2), Amazon DynamoDB, Amazon Elastic Block Store (Amazon EBS) (PIOPS), Amazon Relational Database Service (Amazon RDS), Amazon EMR und Netzwerkbetrieb. Dies kann mit einer Feedback-Schleife wie Auto Scaling oder durch benutzerdefinierten Code im Workload erfolgen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Erstellen Sie eine Feedback-Schleife innerhalb des Workloads, die aktive Metriken aus dem laufenden Workload verwendet, um Änderungen an diesem Workload vorzunehmen. Sie können einen verwalteten Service wie AWS Auto Scalingverwenden, den Sie so konfigurieren, dass er die richtigen Dimensionierungsvorgänge für Sie durchführt. AWS bietet auch APIs, SDKsund Funktionen, mit denen Ressourcen mit minimalem Aufwand angepasst werden können. Sie können einen Workload so programmieren, dass eine Amazon Elastic Compute Cloud-Instance (Amazon EC2) angehalten und gestartet wird, um eine Änderung der Instance-Größe oder des Instance-Typs zuzulassen. Dies bietet die Vorteile der richtigen Dimensionierung und eliminiert nahezu alle Betriebskosten, die für die Änderung erforderlich sind.

Einige AWS-Services verfügen über eine automatische Auswahl von Typ oder Größe, z. B. Amazon Simple Storage Service(Amazon S3) Intelligent-Tiering. Amazon S3 Intelligent-Tiering verschiebt Ihre Daten automatisch zwischen zwei Zugriffsebenen: Häufiger Zugriff und seltener Zugriff, basierend auf Ihren Nutzungsmustern.

Implementierungsschritte

- Konfigurieren von Workload-Metriken: Stellen Sie sicher, dass Sie die Schlüsselmetriken für den Workload erfassen. Diese Metriken geben die Kundenerfahrung an, z. B. die Workload-Ausgabe. Sie passen sich außerdem an die Unterschiede zwischen Ressourcentypen und -umfängen, z. B. CPU- und Speichernutzung, an.
- Anzeige von Empfehlungen zur Umfangsanpassung: Verwenden Sie die Empfehlungen zur Umfangsanpassung in AWS Compute Optimizer, um Anpassungen an Ihrem Workload vorzunehmen.
- Automatische Auswahl des Ressourcentyps und -umfangs basierend auf Metriken: Mithilfe der Workload-Metriken können Sie Ihre Workload-Ressourcen manuell oder automatisch auswählen. Die Konfiguration von AWS Auto Scaling oder die Implementierung von Code in Ihrer Anwendung kann den Aufwand reduzieren, der bei häufigen Änderungen erforderlich ist, und möglicherweise Änderungen früher implementieren, als dies mit einem manuellen Prozess der Fall wäre.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Compute Optimizer
- Amazon CloudWatch-Funktionen
- Einrichten von CloudWatch
- CloudWatch – Veröffentlichen benutzerdefinierter Metriken
- Kostenoptimierung: Richtige Amazon EC2-Dimensionierung
- Erste Schritte mit Amazon EC2 Auto Scaling
- Amazon S3 Intelligent-Tiering
- Starten einer Amazon EC2-Instance mit SDK

KOSTEN 7 Wie können Sie Kosten mithilfe von Preismodellen senken?

Verwenden Sie das Preismodell, das sich für Ihre Ressourcen am besten eignet. So halten Sie die Ausgaben möglichst niedrig.

Bewährte Methoden

- COST07-BP01 Durchführen einer Preismodellanalyse
- COST07-BP02 Implementieren von Regionen auf Basis der Kosten
- COST07-BP03 Auswahl von Drittanbietervereinbarungen mit kosteneffizienten Bedingungen
- COST07-BP04 Implementieren von Preismodellen für alle Komponenten dieses Workloads
- COST07-BP05 Durchführen einer Preismodellanalyse auf Masterkontoebene

COST07-BP01 Durchführen einer Preismodellanalyse

Analysieren Sie die einzelnen Komponenten des Workloads. Stellen Sie fest, ob die Komponente und die Ressourcen über einen längeren Zeitraum (für Bindungsrabatte) oder dynamisch und kurz ausgeführt werden (für Spot- oder On-Demand-Instances). Analysieren Sie den Workload mithilfe der AWS Cost Explorer-Empfehlungsfunktion.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

AWS verfügt über mehrere Preismodelle, mit denen Sie für Ihre Ressourcen auf die kostengünstigste Art und Weise bezahlen können, die den Anforderungen Ihres Unternehmens entspricht.

Implementierungsschritte

- Durchführen einer Analyse des Bindungsrabatts: Sehen Sie sich unter Verwendung des Cost Explorer in Ihrem Konto die Empfehlungen für Savings Plans und Reserved Instances an. Um sicherzustellen, dass Sie die richtigen Empfehlungen mit den erforderlichen Rabatten und Risiken implementieren, befolgen Sie die Well-Architected Labs.
- Analysieren der Workload-Elastizität: Verwenden Sie die stündliche Granularität im Cost Cost Explorer oder ein benutzerdefiniertes Dashboard. Analysieren der Workload-Elastizität. Suchen Sie nach regelmäßigen Änderungen in der Anzahl der Instances, die ausgeführt werden. Instances mit kurzer Dauer sind Kandidaten für Spot Instances oder Spot-Flotte.
  - Well-Architected Lab: Cost Explorer
  - Well-Architected Lab: Kostenvisualisierung

Ressourcen

Zugehörige Dokumente:

- Zugreifen auf Empfehlungen für Reserved Instances
- Kaufoptionen für Instances

Relevante Videos:

- Einsparen von bis zu 90 % und Ausführen der Produktions-Workloads mit Spot

Zugehörige Beispiele:

- Well-Architected Lab: Cost Explorer
- Well-Architected Lab: Kostenvisualisierung
- Well-Architected Lab: Preismodelle

COST07-BP02 Implementieren von Regionen auf Basis der Kosten

Die Ressourcenpreise können je nach Region abweichen. Die Berücksichtigung der Regionskosten stellt sicher, dass Sie den niedrigsten Gesamtpreis für diesen Workload zahlen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Wenn Sie die Architektur Ihrer Lösungen aufbauen, hat es sich bewährt zu versuchen, Computing-Ressourcen zugunsten einer geringeren Latenz und einer stärkeren Datensouveränität näher an die Benutzer zu bringen. Für weltweite Zielgruppen sollten Sie mehrere Standorte verwenden, um diese Anforderungen zu erfüllen. Sie sollten den geografischen Standort auswählen, der Ihre Kosten minimiert.

Die AWS Cloud-Infrastruktur basiert auf Regionen und Availability Zones. Eine Region ist ein physischer Standort auf der Welt, an dem wir mehrere Availability Zones haben. Availability Zones bestehen aus mindestens einem eigenständigen Rechenzentrum mit einer redundanten Stromversorgung, einem Netzwerk sowie Konnektivität. Sie sind jeweils in getrennten Einrichtungen untergebracht.

Jede AWS-Region wird im Rahmen der jeweilig gültigen lokalen Marktbedingungen betrieben, und die Ressourcenpreise können von Region zu Region variieren. Wählen Sie eine spezifische Region aus, in der Sie eine Komponente oder Ihre gesamte Lösung ausführen möchten, sodass Sie

weltweit einen Betrieb zu den geringstmöglichen Kosten gewährleisten. Sie können den AWS Pricing Calculator verwenden, um die Kosten Ihres Workload in verschiedenen Regionen einzuschätzen.

Implementierungsschritte

- Überprüfung der Preise für die Region: Analysieren Sie die Workload-Kosten in der aktuellen Region. Berechnen Sie die Kosten in anderen verfügbaren Regionen, beginnend mit den höchsten Kosten nach Service und Verwendungstyp. Migrieren Sie in die neue Region, wenn die prognostizierte Einsparung die Kosten für das Verschieben der Komponente oder des Workloads überwiegt.

Ressourcen

Zugehörige Dokumente:

- Zugreifen auf Empfehlungen für Reserved Instances
- Amazon EC2-Preise
- Kaufoptionen für Instances
- Tabelle „Region“

Relevante Videos:

- Einsparen von bis zu 90 % und Ausführen der Produktions-Workloads mit Spot

COST07-BP03 Auswahl von Drittanbietervereinbarungen mit kosteneffizienten Bedingungen

Kosteneffiziente Vereinbarungen und Bedingungen stellen sicher, dass die Kosten dieser Services mit den von ihnen bereitgestellten Vorteilen skaliert werden. Wählen Sie Vereinbarungen und Preise aus, die skaliert werden, wenn sie Ihrem Unternehmen zusätzliche Vorteile bieten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Wenn Sie Drittanbieterlösungen oder -services in der Cloud nutzen, ist es wichtig, dass die Preisstrukturen an den Ergebnissen der Kostenoptimierung ausgerichtet sind. Die Preise sollten mit den Ergebnissen und dem Wert skaliert werden, den sie bieten. Ein Beispiel hierfür ist Software, die einen Prozentsatz der Einsparungen in Anspruch nimmt, je mehr Sie sparen (Ergebnis), desto mehr Gebühren fallen an. Vereinbarungen, die mit Ihrer Rechnung skaliert werden, sind in der Regel

nicht auf die Kostenoptimierung ausgerichtet, es sei denn, sie liefern Ergebnisse für jeden Teil Ihrer spezifischen Rechnung. Beispiel: Eine Lösung, die Empfehlungen für Amazon Elastic Compute Cloud(Amazon EC2) bereitstellt und einen Prozentsatz Ihrer gesamten Rechnung berechnet, wird teurer, wenn Sie andere Services nutzen, für die sie keinen Vorteil bietet. Ein weiteres Beispiel ist ein verwalteter Service, der zu einem Prozentsatz der Kosten für verwaltete Ressourcen in Rechnung gestellt wird. Eine höhere Instance-Größe erfordert möglicherweise nicht notwendigerweise mehr Verwaltungsaufwand, wird jedoch mehr in Rechnung gestellt. Stellen Sie sicher, dass diese Service-Preisvereinbarungen ein Kostenoptimierungsprogramm oder entsprechende Funktionen in ihrem Service enthalten, um die Effizienz zu steigern.

Implementierungsschritte

- Analyse von Vereinbarungen und Bedingungen Dritter: Überprüfen Sie die Preise in Drittanbietervereinbarungen. Führen Sie die Modellierung für verschiedene Nutzungsebenen durch und berücksichtigen Sie neue Kosten, wie z. B. die Nutzung neuer Services oder Erweiterungen der aktuellen Services aufgrund des Workload-Wachstums. Entscheiden Sie, ob die zusätzlichen Kosten Ihrem Unternehmen die erforderlichen Vorteile bieten.

Ressourcen

Zugehörige Dokumente:

- Zugreifen auf Empfehlungen für Reserved Instances
- Kaufoptionen für Instances

Relevante Videos:

- Einsparen von bis zu 90 % und Ausführen der Produktions-Workloads mit Spot

COST07-BP04 Implementieren von Preismodellen für alle Komponenten dieses Workloads

Dauerhaft ausgeführte Ressourcen sollten reservierte Kapazität wie Savings Plans oder Reserved Instances nutzen. Die kurzfristige Kapazität wird für die Verwendung von Spot Instances oder einer Spot-Flotte konfiguriert. On-Demand-Instances werden nur für kurzfristige Workloads verwendet, die nicht unterbrochen werden können und nicht lange genug für reservierte Kapazitäten ausgeführt werden – typischerweise 25 bis 75 % des Zeitraums, je nach Ressourcentyp.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Berücksichtigen Sie die Anforderungen der Workload-Komponenten und verstehen Sie die potenziellen Preismodelle. Definieren Sie die Verfügbarkeitsanforderung der Komponente. Stellen Sie fest, ob mehrere unabhängige Ressourcen vorhanden sind, die die Funktion im Workload ausführen, und welche Workload-Anforderungen im Laufe der Zeit gelten. Vergleichen Sie die Kosten der Ressourcen unter Verwendung des standardmäßigen On-Demand-Preismodells und anderer anwendbarer Modelle. Beziehen Sie potenzielle Änderungen in Ressourcen oder Workload-Komponenten in Ihre Überlegungen ein.

Implementierungsschritte

- Implementieren von Preismodellen: Nutzen Sie Ihre Analyseergebnisse, um Savings Plans (SPs) bzw. Reserved Instances (RIs) zu erwerben oder Spot Instances zu implementieren. Wenn es sich um Ihren ersten RI-Kauf handelt, wählen Sie die besten 5 oder 10 Empfehlungen in der Liste aus und überwachen und analysieren Sie dann die Ergebnisse in den nächsten ein oder zwei Monaten. Erwerben Sie in regelmäßigen Zyklen eine geringe Anzahl von Bindungsrabatten, z. B. alle zwei Wochen oder monatlich. Implementieren Sie Spot Instances für Workloads, die unterbrochen werden können oder zustandslos sind.
- Workload-Überprüfungszyklus: Implementieren Sie einen Überprüfungszyklus für den Workload, der speziell die Abdeckung des Preismodells analysiert. Erwerben Sie alle zwei bis vier Wochen weitere Bindungsrabatte sobald der Workload über die erforderliche Abdeckung verfügt oder wenn sich die Nutzung Ihrer Organisation ändert.

Ressourcen

Zugehörige Dokumente:

- Zugreifen auf Empfehlungen für Reserved Instances
- EC2-Flotte
- Erwerb von Reserved Instances
- Kaufoptionen für Instances
- Spot Instances

Relevante Videos:

- Einsparen von bis zu 90 % und Ausführen der Produktions-Workloads mit Spot

COST07-BP05 Durchführen einer Preismodellanalyse auf Masterkontoebene

Verwenden Sie Cost Explorer Savings Plans und Reserved Instance-Empfehlungen, um regelmäßige Analysen auf Managementkontoebene für Bindungsrabatte durchzuführen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Durch die regelmäßige Kostenmodellierung wird sichergestellt, dass Möglichkeiten zur Optimierung über mehrere Workloads hinweg implementiert werden können. Wenn beispielsweise mehrere Workloads On-Demand-Instances verwenden, ist das Änderungsrisiko insgesamt niedriger und die Nutzung eines auf fester Kapazität basierenden Rabatts führt zu niedrigeren Gesamtkosten. Es wird empfohlen, Analysen in regelmäßigen Zyklen von zwei Wochen bis zu einem Monat durchzuführen. Auf diese Weise können Sie kleine Anpassungskäufe tätigen, sodass sich die Abdeckung Ihrer Preismodelle mit Ihren sich ändernden Workloads und ihren Komponenten weiter entwickelt.

Verwenden Sie das AWS Cost Explorer -Empfehlungstool, um Möglichkeiten für an feste Kapazität gebundene Rabatte zu finden.

Um Möglichkeiten für Spot-Workloads zu finden, verwenden Sie eine stündliche Ansicht Ihrer Gesamtnutzung und suchen Sie nach regelmäßigen Zeiträumen mit sich ändernder Nutzung oder Elastizität.

Implementierungsschritte

- Durchführen einer Analyse des Bindungsrabatts: Sehen Sie sich unter Verwendung des Cost Explorer in Ihrem Konto die Empfehlungen für Savings Plans und Reserved Instances an. Um sicherzustellen, dass Sie die richtigen Empfehlungen mit den erforderlichen Rabatten und Risiken implementieren, befolgen Sie die Well-Architected Labs.

Ressourcen

Zugehörige Dokumente:

- Zugreifen auf Empfehlungen für Reserved Instances
- Kaufoptionen für Instances

Relevante Videos:

- Einsparen von bis zu 90 % und Ausführen der Produktions-Workloads mit Spot

Zugehörige Beispiele:

- Well-Architected Lab: Preismodelle

KOSTEN 8 Wie können Sie die Kosten für Datenübertragungen planen?

Damit Sie architekturbezogene Entscheidungen zur Kostenminimierung treffen können, müssen Sie unbedingt die Datenübertragungskosten einplanen und überwachen. Eine geringfügige, aber effektive Änderung an der Architektur kann Ihre Betriebskosten über einen längeren Zeitraum hinweg erheblich senken.

Bewährte Methoden

- COST08-BP01 Durchführen einer Datenübertragungsmodellierung
- COST08-BP02 Auswahl von Komponenten zur Optimierung der Datenübertragungskosten
- COST08-BP03 Implementieren von Services zur Senkung der Datenübertragungskosten

COST08-BP01 Durchführen einer Datenübertragungsmodellierung

Stellen Sie die Organisationsanforderungen zusammen und führen Sie eine Datenübertragungsmodellierung des Workloads und ihrer einzelnen Komponenten durch. Dadurch wird der niedrigste Kostenpunkt für die jeweiligen aktuellen Datenübertragungsanforderungen ermittelt.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Analysieren Sie, wo die Datenübertragung in Ihrem Workload stattfindet, welche Kosten für die Übertragung entstehen und welche Vorteile damit verbunden sind. Auf diese Weise können Sie eine fundierte Entscheidung treffen, die Architekturentscheidung zu ändern oder zu akzeptieren. Sie können beispielsweise über eine Multi-Availability Zone-Konfiguration verfügen, in der Sie Daten zwischen den Availability Zones replizieren. Sie modellieren die Kosten der Struktur und entscheiden, dass dies akzeptable Kosten sind (ähnlich wie bei der Zahlung für Datenverarbeitung und Speicher in beiden Availability Zones), um die erforderliche Zuverlässigkeit und Ausfallsicherheit zu erreichen.

Modellieren Sie die Kosten über verschiedene Nutzungsstufen. Die Workload-Nutzung kann sich im Laufe der Zeit ändern und verschiedene Services können auf verschiedenen Ebenen kostengünstiger sein.

Verwendung Sie AWS Cost Explorer oder dem AWS Cost and Usage Report (CUR), um Ihre Datenübertragungskosten zu verstehen und zu modellieren. Konfigurieren Sie einen Machbarkeitsnachweis (PoC) oder testen Sie Ihren Workload und führen Sie einen Test mit einer realistischen simulierten Last aus. Sie können Ihre Kosten bei verschiedenen Workload-Nachfragen modellieren.

Implementierungsschritte

- Berechnen der Datenübertragungskosten: Verwenden Sie die AWS-Preisseiten und berechnen Sie die Datenübertragungskosten für den Workload. Berechnen Sie die Datenübertragungskosten auf verschiedenen Nutzungsebenen für Erhöhungen und Verringerungen der Workload-Nutzung. Wenn es mehrere Optionen für die Workload-Architektur gibt, berechnen Sie zum Vergleich die Kosten für die einzelnen Optionen.
- Verbindung von Kosten mit Ergebnissen: Geben Sie für alle anfallenden Datenübertragungskosten das Ergebnis an, das für den Workload erzielt wird. Erfolgt der Transfer zwischen Komponenten, kann dies für die Entkopplung verwendet werden. Erfolgt der Transfer zwischen Availability Zones, kann dies zur Redundanz verwendet werden.

Ressourcen

Zugehörige Dokumente:

- AWS-Caching-Lösungen
- AWS-Preise
- Amazon EC2-Preise
- Amazon VPC-Preise
- Schnellere Bereitstellung von Inhalten mit Amazon CloudFront

COST08-BP02 Auswahl von Komponenten zur Optimierung der Datenübertragungskosten

Alle Komponenten sind ausgewählt und die Architektur ist so konzipiert, dass die Datenübertragungskosten gesenkt werden. Dies umfasst auch die Verwendung von Komponenten wie WAN-Optimierung und Multi-Availability Zone-Konfigurationen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Der Aufbau einer Architektur für die Datenübertragung gewährleistet, dass Sie die Kosten für die Datenübertragung minimieren. Dies kann auch die Nutzung von Inhaltsbereitstellungsnetzwerken bedeuten, um Daten näher an Nutzern zu platzieren, oder die Verwendung spezieller Netzwerk-Links von Ihrem Standort zu AWS. Sie können auch WAN-Optimierung und Anwendungsoptimierung verwenden, um die Datenmenge zu reduzieren, die zwischen Komponenten übertragen wird.

Implementierungsschritte

- Auswahl von Komponenten für die Datenübertragung: Konzentrieren Sie sich mithilfe des Datenübertragungsmodells darauf, wo die größten Datenübertragungskosten liegen oder wo sie sich befinden würden, wenn sich die Workload-Nutzung ändert. Suchen Sie nach alternativen Architekturen oder zusätzlichen Komponenten, die den Datenübertragungsbedarf beseitigen oder reduzieren oder die Kosten senken.

Ressourcen

Zugehörige Dokumente:

- AWS-Caching-Lösungen
- Schnellere Bereitstellung von Inhalten mit Amazon CloudFront

COST08-BP03 Implementieren von Services zur Senkung der Datenübertragungskosten

Implementieren Sie Services zur Verringerung der Datenübertragung. Sie können beispielsweise ein Content Delivery Network (CDN) wie Amazon CloudFront für die Übermittlung von Inhalten an Endbenutzer, Caching-Layer mit Amazon ElastiCache oder AWS Direct Connect anstelle von VPN für die Verbindung mit AWS verwenden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Amazon CloudFront ist ein weltweites Inhaltsbereitstellungsnetzwerk, das Daten bei niedriger Latenz und hohen Datenübertragungsgeschwindigkeiten bereitstellt. Es stellt Daten an Edge-Standorten rund um die Welt in den Cache und reduziert damit die Belastung Ihrer Ressourcen. Durch die

Verwendung von CloudFront können Sie den administrativen Aufwand für die Bereitstellung von Inhalten für eine große Anzahl an Benutzern weltweit bei minimaler Latenz reduzieren.

AWS Direct Connect können Sie eine dedizierte Netzwerkverbindung zu AWS aufbauen. Damit können Sie Netzwerkkosten reduzieren, die Bandbreite erhöhen und eine im Vergleich zu Internet-basierten Verbindungen gleichbleibendere Netzwerkerfahrung bieten.

Mit AWS VPN können Sie eine sichere und private Verbindung zwischen Ihrem privaten Netzwerk und dem globalen AWS-Netzwerk herstellen. Es ist ideal für kleine Niederlassungen oder Geschäftspartner, da es eine schnelle und einfache Konnektivität bietet und ein vollständig verwalteter und elastischer Service ist.

VPC-Endpunkte ermöglichen die Konnektivität zwischen AWS-Services über private Netzwerke und können verwendet werden, um Kosten für öffentliche Datenübertragungen und NAT-Gateways zu reduzieren. Für Gateway-VPC-Endpunkte fallen keine stündlichen Gebühren an und sie unterstützen Amazon Simple Storage Service(Amazon S3) und Amazon DynamoDB. Schnittstellen-VPC-Endpunkte werden von AWS PrivateLink bereitgestellt und für sie fällt eine Gebühr pro Stunde und Nutzungskosten pro GB an.

Implementierungsschritte

- Implementieren von Services: Sehen Sie sich mit der Datenübertragungsmodellierung an, wo sich die höchsten Kosten und Volumenströme befinden. Überprüfen Sie die AWS-Services und prüfen Sie, ob es einen Service gibt, der die Übertragung reduziert oder entfernt, insbesondere die Netzwerk- und Inhaltsbereitstellung. Suchen Sie auch nach Caching-Services, bei denen wiederholt auf Daten oder große Datenmengen zugegriffen wird.

Ressourcen

Zugehörige Dokumente:

- AWS Direct Connect
- Unsere AWS-Produkte entdecken
- AWS-Caching-Lösungen
- Amazon CloudFront
- Schnellere Bereitstellung von Inhalten mit Amazon CloudFront

Verwaltung von Nachfrage und Bereitstellung von Ressourcen

Frage

- KOSTEN 9 Wie verwalten Sie die Nachfrage und stellen Ressourcen bereit?

KOSTEN 9 Wie verwalten Sie die Nachfrage und stellen Ressourcen bereit?

Stellen Sie bei einem Workload mit ausgewogenen Ausgaben und Leistungen sicher, dass alles, wofür Sie bezahlen, genutzt wird, und vermeiden Sie eine erhebliche Unterauslastung der Instances. Eine verschobene Auslastungsmetrik in einer der Richtungen wirkt sich nachteilig auf Ihr Unternehmen aus, entweder im Hinblick auf die Betriebskosten (verschlechterte Leistung aufgrund von Überbelegung) oder auf die verschwendeten AWS-Ausgaben (aufgrund von Überversorgung).

Bewährte Methoden

- COST09-BP01 Analyse des Workload-Bedarfs
- COST09-BP02 Implementieren eines Puffers oder einer Drosselung zur Bedarfsverwaltung
- COST09-BP03 Dynamische Bereitstellung von Ressourcen

COST09-BP01 Analyse des Workload-Bedarfs

Analysieren Sie den Bedarf des Workloads im gesamten Zeitverlauf. Stellen Sie sicher, dass die Analyse saisonale Trends berücksichtigt und die Betriebsbedingungen über die gesamte Lebensdauer des Workloads genau wiedergibt. Der Analyseaufwand sollte in einem angemessenen Verhältnis zum potenziellen Nutzen stehen, z. B. muss der Zeitaufwand den Workload-Kosten entsprechen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Informieren Sie sich über die Anforderungen des Workloads. Die Anforderungen des Unternehmen sollten die Reaktionszeiten des Workloads für Anforderungen angeben. Die Reaktionszeit kann verwendet werden, um zu bestimmen, ob der Bedarf verwaltet wird oder ob sich das Angebot an Ressourcen ändert, um der Nachfrage gerecht zu werden.

Die Analyse sollte die Vorhersehbarkeit und Wiederholbarkeit der Nachfrage, die Änderungsrate der Nachfrage und die Menge der Nachfrageänderungen umfassen. Stellen Sie sicher, dass die

Analyse über einen ausreichend langen Zeitraum ausgeführt wird, um saisonale Abweichungen wie die Verarbeitung am Ende des Monats oder Feiertagsspitzen einzubeziehen.

Stellen Sie sicher, dass der Analyseaufwand die potenziellen Vorteile der Implementierung der Skalierung widerspiegelt. Sehen Sie sich die erwarteten Gesamtkosten der Komponente sowie etwaige Erhöhungen oder Verringerungen der Nutzung und der Kosten während der Lebensdauer des Workloads an.

Sie können AWS Cost Explorer oder Amazon QuickSight mit AWS Cost and Usage Report (CUR) oder Ihren Anwendungsprotokollen verwenden, um eine visuelle Analyse des Workload-Bedarfs durchzuführen.

Implementierungsschritte

- Analysieren vorhandener Workload-Daten: Analysieren Sie Daten aus dem vorhandenen Workload, früheren Versionen des Workloads oder vorhergesagten Nutzungsmustern. Verwenden Sie Protokolldateien und Überwachungsdaten, um Einblicke in die Nutzung des Workloads durch Kunden zu erhalten. Typische Metriken sind der tatsächliche Bedarf nach Anfragen pro Sekunde, die Zeiten, in denen sich die Bedarfsrate ändert, oder wenn sie sich auf verschiedenen Ebenen befindet, sowie die Rate der Bedarfsänderung. Stellen Sie sicher, dass Sie einen vollständigen Workload-Zyklus analysieren und dass Sie Daten für saisonale Änderungen erfassen, z. B. Ereignisse am Monatsende oder am Ende des Jahres. Der in der Analyse reflektierte Aufwand sollte die Workload-Merkmale widerspiegeln. Der größte Aufwand sollte für hochwertige Workloads mit den größten Nachfrageänderungen betrieben werden. Der geringste Aufwand sollte für Workloads mit geringfügigen Nachfrageänderungen betrieben werden. Häufige Metriken für den Wert sind Risiko, Markenbewusstsein, Umsatz oder Workload-Kosten.
- Vorhersage externer Einflüsse: Treffen Sie Teammitglieder aus der gesamten Organisation, die die Nachfrage im Workload beeinflussen oder ändern können. Häufig betroffene Teams sind Vertrieb, Marketing oder Geschäftsentwicklung. Arbeiten Sie mit ihnen zusammen, um die Zyklen kennenzulernen, mit denen sie arbeiten, und um zu erfahren, ob es Ereignisse gibt, die die Nachfrage des Workloads ändern könnten. Erstellen Sie eine Prognose des Workload-Bedarfs anhand dieser Daten.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling

- AWS Instance Scheduler
- Erste Schritte mit Amazon SQS
- AWS Cost Explorer
- Amazon QuickSight

COST09-BP02 Implementieren eines Puffers oder einer Drosselung zur Bedarfsverwaltung

Pufferung und Drosselung ändern den Bedarf Ihres Workloads und glätten alle Spitzen. Implementieren Sie die Drosselung, wenn Ihre Clients Wiederholungen durchführen. Implementieren Sie die Pufferung, um die Anforderung zu speichern und die Verarbeitung auf einen späteren Zeitpunkt zu verschieben. Stellen Sie sicher, dass Ihre Drosselungen und Puffer so konzipiert sind, dass Clients in der erforderlichen Zeit eine Antwort erhalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Drosselung: Wenn die Quelle der Nachfrage über eine Wiederholungsfunktion verfügt, können Sie die Drosselung implementieren. Die Drosselung teilt der Quelle mit, dass wenn sie die Anforderung zum aktuellen Zeitpunkt nicht bedienen kann, sie es später erneut versuchen sollte. Die Quelle wartet einen bestimmten Zeitraum und wiederholt die Anforderung. Die Implementierung der Drosselung hat den Vorteil, dass die maximale Menge an Ressourcen und Kosten des Workloads begrenzt wird. In AWS können Sie Amazon API Gateway verwenden, um die Drosselung zu implementieren. Weitere Informationen zur Implementierung der Drosselung finden Sie im Well-Architected Whitepaper zur Säule "Zuverlässigkeit" .

Pufferung: Ähnlich wie bei der Drosselung verschiebt ein Puffer die Anforderungsverarbeitung, sodass Anwendungen, die mit unterschiedlichen Raten ausgeführt werden, effektiv kommunizieren können. Bei der Pufferung werden Nachrichten (Arbeitseinheiten) von Produzenten in eine Warteschlange gestellt. Nachrichten können dadurch von Verbrauchern in der für ihre Geschäftsanforderungen passenden Geschwindigkeit gelesen und verarbeitet werden. Sie brauchen sich keine Gedanken darüber zu machen, wie Produzenten mit Drosselungsproblemen, z. B. der Datenbeständigkeit und dem Gegendruck, umgehen. Bei Gegendruck werden die Produzenten langsamer, damit die langsameren Verbraucher die Daten aufnehmen können.

In AWS können Sie zur Implementierung eines pufferbasierten Ansatzes aus mehreren Services wählen. Amazon Simple Queue Service(Amazon SQS) ist ein verwalteter Service, der Warteschlangen bietet, die es einem einzelnen Verbraucher ermöglichen, individuelle Nachrichten zu

lesen. Amazon Kinesis stellt einen Stream bereit, der es vielen Verbrauchern ermöglicht, dieselben Nachrichten zu lesen.

Stellen Sie bei der Architektur mit einem pufferbasierten Ansatz sicher, dass Sie Ihren Workload so gestalten, dass er die Anforderung in der erforderlichen Zeit erfüllt, und dass Sie doppelte Arbeitsanfragen verarbeiten können.

Implementierungsschritte

- Analysieren der Client-Anforderungen: Analysieren Sie die Client-Anforderungen, um zu bestimmen, ob sie Wiederholungen durchführen können. Für Clients, die keine Wiederholungen durchführen können, müssen Puffer implementiert werden. Analysieren Sie den Gesamtbedarf, die Änderungsrate und die erforderliche Reaktionszeit, um die Größe der erforderlichen Drosselung oder des Puffers zu bestimmen.
- Implementieren eines Puffers oder einer Drosselung: Implementieren Sie einen Puffer oder eine Drosselung im Workload. Eine Warteschlange wie Amazon Simple Queue Service (Amazon SQS) kann für Ihre Workload-Komponenten einen Puffer bereitstellen. Amazon API Gateway kann eine Drosselung für Ihre Workload-Komponenten bereitstellen.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Instance Scheduler
- Amazon API Gateway
- Amazon Simple Queue Service
- Erste Schritte mit Amazon SQS
- Amazon Kinesis

COST09-BP03 Dynamische Bereitstellung von Ressourcen

Ressourcen werden geplant bereitgestellt. Dies kann bedarfsbasiert sein, z. B. durch Auto Scaling, oder zeitbasiert, wobei der Bedarf vorhersehbar ist und Ressourcen basierend auf der Zeit bereitgestellt werden. Diese Methoden führen dazu, dass die Über- oder Unterversorgung am geringsten ist.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Sie können AWS Auto Scalingverwenden oder die Skalierung in Ihren Code mit der AWS API oder dem SDK einbinden. Dies reduziert Ihre Gesamtkosten für den Workload, da die Betriebskosten durch manuelle Änderungen an Ihrer Umgebung wegfallen, und kann viel schneller durchgeführt werden. Auf diese Weise wird sichergestellt, dass die Workload-Ressourcen jederzeit am besten mit der Nachfrage übereinstimmen.

Nachfragebasiertes Angebot: Nutzen Sie die Elastizität der Cloud, um Ressourcen bereitzustellen, die sich ändernde Anforderungen erfüllen. Nutzen Sie die Vorteile von APIs oder Service-Funktionen, um die Menge der Cloud-Ressourcen in Ihrer Architektur dynamisch zu variieren. Auf diese Weise können Sie Komponenten in Ihrer Architektur skalieren und die Anzahl der Ressourcen in Bedarfsspitzenzeiten zur Aufrechterhalten der Leistung automatisch erhöhen und die Kapazität zur Reduzierung der Kosten herabsetzen, wenn der Bedarf abklingt.

AWS Auto Scaling können Sie Ihre Kapazität anpassen, um eine stabile, vorhersehbare Leistung zu möglichst niedrigen Kosten aufrechtzuerhalten. Es handelt sich um einen vollständig verwalteten und kostenlosen Service, der sich in Amazon Elastic Compute Cloud-Instances (Amazon EC2) und Spot-Flotten, Amazon Elastic Container Service (Amazon ECS), Amazon DynamoDB und Amazon Aurora integrieren lässt.

Auto Scaling bietet eine automatische Ressourcenerkennung, um zu helfen, Ressourcen in Ihrem Workload zu finden, die konfiguriert werden können. Es verfügt über integrierte Skalierungsstrategien zur Optimierung der Leistung, der Kosten oder eines Gleichgewichts zwischen beiden Ressourcen und bietet eine prädiktive Skalierung, um regelmäßig auftretende Spitzen zu unterstützen.

Auto Scaling kann eine manuelle, geplante oder bedarfsbasierte Skalierung implementieren. Sie können auch Metriken und Alarme von Amazon CloudWatch verwenden, um Skalierungsereignisse für Ihren Workload auszulösen. Typische Metriken können Amazon EC2-Standardmetriken sein, z. B. CPU-Auslastung, Netzwerkdurchsatz und von Elastic Load Balancing(ELB) beobachtete Anforderungs-/Antwortlatenz. Wenn möglich, sollten Sie eine Metrik verwenden, die auf das Kundenerlebnis hinweist. In der Regel handelt es sich um eine benutzerdefinierte Metrik, die aus Anwendungscode innerhalb Ihres Workloads stammen kann.

Beim Aufbau der Architektur mit einem bedarfsbasierten Ansatz sollten Sie die folgenden beiden wichtigen Aspekte berücksichtigen: 1. Machen Sie sich damit vertraut, wie schnell Sie neue Ressourcen bereitstellen müssen. 2. Machen Sie sich damit vertraut, dass sich die Größe der Marge zwischen Angebot und Nachfrage ändern wird. Sie müssen darauf vorbereitet sein, das Intervall der Änderung in Bezug auf die Nachfrage zu verarbeiten, und auch Ressourcenfehler einkalkulieren.

ELB unterstützt Sie bei der Skalierung durch die Verteilung der Nachfrage auf mehrere Ressourcen. Wenn Sie weitere Ressourcen implementieren, fügen Sie sie dem Load Balancer hinzu, um die Nachfrage zu erfüllen. Elastic Load Balancing unterstützt Amazon EC2-Instances, Container, IP-Adressen und AWS Lambda-Funktionen.

Zeitbasiertes Angebot: Ein zeitbasierter Ansatz richtet die Ressourcenkapazität an Bedarfen aus, die prognostizierbar sind oder zeitlich gut definiert werden können. Dieser Ansatz ist in der Regel nicht abhängig vom Nutzungsgrad der Ressourcen. Mit einem zeitbasierten Ansatz können Sie sicherstellen, dass Ressourcen zu dem Zeitpunkt zur Verfügung stehen, zu dem sie benötigt werden, und ohne Verzögerung aufgrund von Startverfahren und System- oder Konsistenzprüfungen bereitgestellt werden können. Durch die Verwendung eines zeitbasierten Ansatzes können Sie zusätzliche Ressourcen hinzufügen oder die Kapazität in Spitzenzeiten erhöhen.

Sie können geplantes Auto Scaling verwenden, um einen zeitbasierten Ansatz zu implementieren. Workloads können zu bestimmten Zeiten auf Basis eines Zeitplans hoch- oder runterskaliert werden, z. B. zu Beginn der Geschäftszeiten, und damit sicherstellen, dass ausreichende Ressourcen verfügbar sind, wenn die Benutzer oder Nachfrage ankommen.

Sie können die AWS-APIs und SDKs und AWS CloudFormation nutzen, um vollständige Umgebungen bei Bedarf bereitzustellen oder zu deaktivieren. Dieser Ansatz eignet sich hervorragend für Entwicklungs- und Testumgebungen, die nur zu Geschäftszeiten oder in bestimmten Zeiträumen ausgeführt werden.

Mit APIs können Sie die Größe der Ressourcen innerhalb einer Umgebung skalieren (Stichwort: vertikales Skalieren). So könnten Sie beispielsweise einen Produktions-Workload hochskalieren, indem Sie die Instance-Größe oder -Klasse ändern. Stoppen und starten Sie dazu die Instance, und wählen Sie eine andere Instance-Größe oder -Klasse aus. Diese Technik kann auch auf andere Ressourcen angewendet werden, z. B. Amazon Elastic Block Store (Amazon EBS), bei denen Sie im laufenden Betrieb die Größe ändern, die Leistung anpassen (IOPS) oder den Volume-Typ ändern können.

Beim Aufbau der Architektur mit einem zeitbasierten Ansatz sollten Sie die beiden folgenden wichtigen Aspekte berücksichtigen: 1: Wie konsistent ist das Nutzungsmuster? 2. Wie wirken sich Musteränderungen aus? Sie können die Treffergenauigkeit für Prognosen durch die Überwachung Ihrer Workloads und die Verwendung von Business Intelligence erhöhen. Wenn Sie signifikante Änderungen im Nutzungsmuster erkennen, können Sie die Zeiten ändern, um eine Deckung zu gewährleisten.

Implementierungsschritte

- Konfigurieren der zeitbasierten Planung: Für vorhersehbare Änderungen des Bedarfs kann die zeitbasierte Skalierung die richtige Anzahl an Ressourcen in einem angemessenen Zeitraum bereitstellen. Es ist auch nützlich, wenn die Ressourcenerstellung und -konfiguration nicht schnell genug ist, um bei Bedarf auf Änderungen zu reagieren. Mithilfe der Workload-Analyse konfigurieren Sie die geplante Skalierung mithilfe von AWS Auto Scaling.
- Konfigurieren von Auto Scaling: Verwenden Sie Amazon Auto Scaling, um die Skalierung basierend auf aktiven Workload-Metriken zu konfigurieren. Verwenden Sie die Analyse und konfigurieren Sie Auto Scaling so, dass es auf den richtigen Ressourcenebenen ausgelöst wird. Stellen Sie sicher, dass der Workload in der erforderlichen Zeit skaliert wird.

Ressourcen

Zugehörige Dokumente:

- AWS Auto Scaling
- AWS Instance Scheduler
- Erste Schritte mit Amazon EC2 Auto Scaling
- Erste Schritte mit Amazon SQS
- Geplante Skalierung für Amazon EC2 Auto Scaling

Optimierung im Laufe der Zeit

Frage

- KOSTEN 10 Wie können Sie neue Services bewerten?

KOSTEN 10 Wie können Sie neue Services bewerten?

Im Zuge der Veröffentlichung neuer Services und Funktionen durch AWS empfiehlt es sich, dass Sie Ihre bestehenden Entscheidungen zur Architektur überdenken, um sicherzustellen, dass diese weiterhin so kostengünstig wie möglich sind.

Bewährte Methoden

- COST10-BP01 Entwickeln eines Prüfprozesses für Workloads
- COST10-BP02 Regelmäßige Prüfung und Analyse des betreffenden Workloads

COST10-BP01 Entwickeln eines Prüfprozesses für Workloads

Entwickeln Sie einen Prozess, der die Kriterien und den Prozess für die Workload-Prüfung definiert. Der Überprüfungsaufwand sollte in einem angemessenen Verhältnis zum potenziellen Nutzen stehen. Beispielsweise ist es sinnvoll, zentrale Workloads oder Workloads, deren Wert mehr als 10 % der Rechnung ausmacht, vierteljährlich zu prüfen, während Workloads mit einem Wert von weniger als 10 % der Rechnung jährlich überprüft werden sollten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Hoch

Implementierungsleitfaden

Um sicherzustellen, dass Sie immer den kosteneffizientesten Workload haben, müssen Sie den Workload regelmäßig überprüfen, um zu wissen, ob es Möglichkeiten gibt, neue Services, Funktionen und Komponenten zu implementieren. Um sicherzustellen, dass Sie insgesamt niedrigere Kosten erzielen, muss der Prozess proportional zu den potenziellen Einsparungen sein. Beispielsweise sollten Workloads, die 50 % Ihrer Gesamtausgaben ausmachen, regelmäßiger und gründlicher überprüft werden als Workloads, die 5 % Ihrer Gesamtausgaben ausmachen. Lassen Sie auch externe Faktoren oder Volatilität in Ihre Überlegungen einfließen. Wenn der Workload eine bestimmte Geografie oder ein bestimmtes Marktsegment abzielt und Änderungen in diesem Bereich vorhergesagt werden, können häufigere Überprüfungen zu Kosteneinsparungen führen. Ein weiterer Faktor bei der Überprüfung ist die Implementierung von Änderungen. Wenn es erhebliche Kosten für das Testen und Validieren von Änderungen gibt, sollten Überprüfungen seltener erfolgen.

Denken Sie an die langfristigen Kosten für die Wartung veralteter Komponenten und Ressourcen sowie die Unfähigkeit, neue Funktionen in diese zu implementieren. Die aktuellen Kosten für Tests und Validierung können den vorgeschlagenen Vorteil übersteigen. Doch mit der Zeit können sich die Kosten für die Änderung erheblich erhöhen, da die Lücke zwischen dem Workload und den aktuellen Technologien zunimmt, was zu noch größeren Kosten führt. Beispielsweise sind die Kosten für den Wechsel zu einer neuen Programmiersprache derzeit möglicherweise nicht günstig. In fünf Jahren können sich jedoch die Kosten für Personen, die in dieser Sprache qualifiziert sind, erhöhen. Aufgrund des Wachstums des Workloads würden Sie ein noch größeres System in die neue Sprache verlagern, was noch mehr Aufwand erfordert als zuvor.

Unterteilen Sie Ihren Workload in Komponenten, weisen Sie die Kosten der Komponente zu (eine Schätzung reicht aus) und listen Sie dann die Faktoren (z. B. Aufwand und externe Märkte) neben den einzelnen Komponenten auf. Verwenden Sie diese Indikatoren, um eine Überprüfungshäufigkeit für jeden Workload zu bestimmen. Zum Beispiel können bei Webservern hohe Kosten, geringer Änderungsaufwand und hohe externe Faktoren anfallen, was zu einer hohen Überprüfungshäufigkeit

führt. Bei einer zentralen Datenbank können mittlere Kosten, hoher Änderungsaufwand und niedrige externe Faktoren anfallen, was zu einer mittleren Überprüfungshäufigkeit führt.

Implementierungsschritte

- Definition der Überprüfungsfrequenz: Legen Sie fest, wie häufig der Workload und seine Komponenten überprüft werden sollen. Dies ist eine Kombination von Faktoren und kann sich von Workload zu Workload innerhalb Ihrer Organisation oder auch zwischen Komponenten im Workload unterscheiden. Häufige Faktoren sind u. a. die Bedeutung für die Organisation, gemessen in Bezug auf Umsatz oder Marke, die Gesamtkosten für die Ausführung des Workloads (einschließlich Betriebs- und Ressourcenkosten), die Komplexität des Workloads (wie einfach es ist, eine Änderung zu implementieren), Softwarelizenzvereinbarungen sowie Änderungen, die aufgrund mangelhafter Lizenzen erhebliche Erhöhungen der Lizenzkosten verursachen würden. Komponenten können funktional oder technisch definiert werden, z. B. Webserver und Datenbanken oder Rechen- und Speicherressourcen. Gleichen Sie die Faktoren entsprechend aus und entwickeln Sie einen Zeitraum für den Workload und seine Komponenten. Sie können sich entscheiden, den vollständigen Workload alle 18 Monate, die Webserver alle 6 Monate, die Datenbank alle 12 Monate, die Datenverarbeitungs- und Kurzzeitspeicherung alle 6 Monate und die Langzeitspeicherung alle 6 Monate zu überprüfen.
- Definition einer gründlichen Überprüfung: Legen Sie fest, wie viel Aufwand für die Prüfung des Workloads oder der Workload-Komponenten aufgewendet wird. Ähnlich wie bei der Überprüfungsfrequenz geht es hier um mehrere Faktoren, die ausgeglichen sein müssen. Sie können beispielsweise entscheiden, für die Analyse der Datenbankkomponente eine Woche und für die Analyse von Speicherprüfungen vier Stunden zu verwenden.

Ressourcen

Zugehörige Dokumente:

- AWS-Blog mit Neuigkeiten
- Arten von Cloud Computing
- Neuerungen bei AWS

COST10-BP02 Regelmäßige Prüfung und Analyse des betreffenden Workloads

Bestehende Workloads werden gemäß den einzelnen definierten Prozessen regelmäßig überprüft.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Um die Vorteile neuer AWS-Services und -Funktionen zu nutzen, müssen Sie den Überprüfungsprozess für Ihre Workloads ausführen und bei Bedarf neue Services und Funktionen implementieren. Sie können beispielsweise Ihre Workloads überprüfen und die Messaging-Komponente durch Amazon Simple Email Service (Amazon SES) ersetzen. Dadurch entfallen die Kosten für den Betrieb und die Wartung einer Flotte von Instances, während die gesamte Funktionalität zu geringeren Kosten bereitgestellt wird.

Implementierungsschritte

- Regelmäßige Überprüfung des Workloads: Führen Sie mit Ihrem definierten Prozess Überprüfungen mit der angegebenen Häufigkeit durch. Stellen Sie sicher, dass Sie den richtigen Aufwand für jede Komponente aufwenden. Dieser Prozess ähnelt dem anfänglichen Designprozess, bei dem Sie Services für die Kostenoptimierung ausgewählt haben. Analysieren Sie die Services und die Vorteile, die sie mit sich bringen würden, sowie den Zeitfaktor bei den Änderungskosten. Analysieren Sie nicht nur die langfristigen Vorteile.
- Implementieren neuer Services: Wenn es das Ziel der Analyse ist, Änderungen zu implementieren, führen Sie zunächst eine Analyse der Basisanforderungen des Workloads durch, um die aktuellen Kosten für jede Ausgabe festzustellen. Implementieren Sie die Änderungen und führen Sie dann eine Analyse durch, um die neuen Kosten für jede Ausgabe zu bestätigen.

Ressourcen

Zugehörige Dokumente:

- AWS-Blog mit Neuigkeiten
- Arten von Cloud Computing
- Neuerungen bei AWS

Nachhaltigkeit

Themen

- Auswahl von Regionen
- Verhaltensmuster von Benutzern
- Software- und Architekturmuster

- Datenmuster
- Hardwaremuster
- Entwicklungs- und Bereitstellungsprozess

Auswahl von Regionen

Frage

- SUS 1 Wie wählen Sie Regionen aus, um Ihre Nachhaltigkeitsziele zu unterstützen?

SUS 1 Wie wählen Sie Regionen aus, um Ihre Nachhaltigkeitsziele zu unterstützen?

Wählen Sie die Regionen, in denen Sie Ihre Workloads implementieren, anhand Ihrer geschäftlichen Anforderungen und Ihrer Nachhaltigkeitsziele aus.

Bewährte Methode:

SUS01-BP01 Auswählen von Regionen in der Nähe von Amazon-Projekten für erneuerbare Energien. Es sollte sich um Regionen handeln, in denen das Stromnetz nachweislich geringere Kohlendioxidemissionen generiert als andere Standorte (oder Regionen).

Wählen Sie Regionen in der Nähe von Amazon-Projekten für erneuerbare Energien aus. Es sollte sich um Regionen handeln, in denen das Stromnetz nachweislich geringere Kohlendioxidemissionen generiert als andere Standorte (oder Regionen).

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

Wählen Sie Regionen in der Nähe von Amazon-Projekten für erneuerbare Energien aus. Es sollte sich um Regionen handeln, in denen das Stromnetz nachweislich geringere Kohlendioxidemissionen generiert als andere Standorte (oder Regionen).

Ressourcen

Ähnliche Dokumente:

- Amazon auf der ganzen Welt
- Methodik für erneuerbare Energien

- What to Consider when Selecting a Region for your Workloads (Relevante Aspekte bei der Wahl einer Region für Ihre Workloads)

Verhaltensmuster von Benutzern

Frage

- SUS 2 Wie können Sie Verhaltensmuster von Benutzern zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

SUS 2 Wie können Sie Verhaltensmuster von Benutzern zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Die Art und Weise, wie Benutzer Ihre Workloads und andere Ressourcen nutzen, kann Sie bei der Identifizierung von Verbesserungen unterstützen, um Nachhaltigkeitsziele zu erreichen. Skalieren Sie Ihre Infrastruktur, um die Benutzerlast kontinuierlich anzupassen. Sorgen Sie dafür, dass zur Unterstützung der Benutzer stets nur die mindestens erforderlichen Ressourcen bereitgestellt werden. Richten Sie Service-Levels an den Kundenanforderungen aus. Positionieren Sie Ressourcen so, dass die für ihre Nutzung erforderlichen Netzwerkkapazitäten begrenzt werden. Entfernen Sie vorhandene, nicht verwendete Komponenten. Identifizieren Sie erstellte, aber nicht verwendete Komponenten und beenden Sie ihre Generierung. Stellen Sie Teammitgliedern Geräte zur Verfügung, die ihre Anforderungen bei geringstmöglichen Auswirkungen auf die Nachhaltigkeit erfüllen.

Bewährte Methoden:

SUS02-BP01 Skalieren der Infrastruktur anhand der Benutzerlast

Identifizieren Sie Zeiträume mit geringer oder gar keiner Nutzung und skalieren Sie Ressourcen herunter, um überschüssige Kapazitäten zu entfernen und die Effizienz zu verbessern.

Typische Anti-Muster:

- Sie skalieren Ihre Infrastruktur nicht mit der Benutzerlast.
- Sie skalieren Ihre Infrastruktur stets manuell.
- Sie belassen die erhöhte Kapazität nach dem Hochskalieren, anstatt wieder herunterzuskalieren.

Vorteile der Nutzung dieser bewährten Methode: Durch das Konfigurieren und Testen der Workload-Elastizität können Sie die Umweltauswirkungen von Workloads verringern, Geld sparen und Leistungs-Benchmarks einhalten. Sie können die Elastizität in der Cloud nutzen, um die Kapazität während und nach Lastspitzen automatisch zu skalieren, um sicherzustellen, dass Sie nur die genaue Anzahl von Ressourcen nutzen, die Sie benötigen, um die Anforderungen Ihrer Kunden zu erfüllen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Mittel

Implementierungsleitfaden

- Elastizität ermöglicht das Anpassen der verfügbaren Ressourcen an den Bedarf. Instances, Container und Funktionen bieten Mechanismen für Elastizität, sei es in Kombination mit automatischer Skalierung oder als Merkmal des Service. Nutzen Sie Elastizität in Ihrer Architektur, um sicherzustellen, dass ein Workload bei geringer Benutzerlast schnell und problemlos herunterskaliert werden kann.
  - Verwendung Sie Amazon EC2 Auto Scaling , um zu prüfen, ob Sie über die korrekte Anzahl von Amazon EC2-Instances verfügen, um die Nutzerlast für Ihre Anwendung zu bewältigen.
  - Verwendung Sie Application Auto Scaling zur automatischen Skalierung der Ressourcen für einzelne AWS-Services über Amazon EC2 hinaus, wie etwa Lambda-Funktionen oder Amazon Elastic Container Service (Amazon ECS)-Services.
  - Verwendung Sie Kubernetes Cluster Autoscaler zur automatischen Skalierung von Kubernetes-Clustern auf AWS.
- Prüfen Sie, ob die Metriken zum Hoch- oder Herunterskalieren für die jeweilige Art des bereitgestellten Workloads überprüft werden. Wenn Sie eine Anwendung zur Video-Transkodierung bereitstellen, wird eine CPU-Auslastung von 100 % erwartet, weshalb dies nicht die Hauptmetrik sein sollte. Sie können eine benutzerdefinierte Metrik (wie etwa die Speichernutzung) für Ihre Skalierungsrichtlinien verwenden, falls erforderlich. Beachten Sie zur Auswahl der geeigneten Metriken die folgenden Hinweise zu Amazon EC2:
  - Es sollte sich um eine gültige Nutzungsmetrik handeln, die beschreibt, wie stark eine Instance genutzt wird.
  - Der Metrikwert muss proportional zur Anzahl der Instances in der Auto Scaling-Gruppe steigen oder sinken.
- Verwendung Sie dynamische Skalierung anstelle von manueller Skalierung für Ihre Auto Scaling-Gruppe. Weiterhin empfehlen wir, dass Sie Zielverfolgungs-Skalierungsrichtlinien für Ihre dynamische Skalierung verwenden.

- Prüfen Sie, ob Workload-Bereitstellungen sowohl mit Hoch- als auch mit Herunterskalierungen umgehen können. Erstellen Sie Testszenarien für Herunterskalierungen, damit sich die Workload wie erwartet verhält. Sie können den Aktivitätsverlauf verwenden, um eine Skalierungsaktivität für eine Auto Scaling-Gruppe zu testen und zu verifizieren.
- Evaluieren Sie Ihren Workload auf vorhersagbare Muster und skalieren Sie proaktiv, wenn Sie vorhergesagte und geplante Änderungen der Nachfrage erwarten. Verwendung Sie die vorausschauende Skalierung mit Amazon EC2 Auto Scaling, um Kapazitäten nicht übermäßig testen zu müssen.

Ressourcen

Zugehörige Dokumente:

- Erste Schritte mit Amazon EC2 Auto Scaling
- Prädiktive Skalierung für EC2, unterstützt von Machine Learning
- Analyse des Benutzerverhaltens mit Amazon OpenSearch Service, Amazon Data Firehose und Kibana
- Was ist Amazon CloudWatch?
- Was ist AWS X-Ray?
- VPC Flow Logs
- Überwachen der DB-Last mit Performance Insights auf Amazon RDS
- Native Unterstützung für die vorausschauende Skalierung mit Amazon EC2 Auto Scaling
- Erstellen einer Amazon EC2 Auto Scaling-Richtlinie auf der Grundlage einer Speichernutzungsmetrik (Linux)
- Vorstellung von Karpenter – Open-Source-Kubernetes-Cluster-Autoscaler mit hoher Leistung

Zugehörige Videos:

- Bessere, schnellere und kostengünstigere Datenverarbeitung: Kostenoptimierung bei Amazon EC2 (CMP202-R1

Zugehörige Beispiele:

- Lab: Beispiele für Amazon EC2 Auto Scaling-Gruppen
- Lab: Implementierung von Autoscaling mit Karpenter

SUS02-BP02 Ausrichten von SLAs an Nachhaltigkeitszielen

Definieren und aktualisieren Sie Service Level Agreements (SLAs), darunter die Zeiträume für Verfügbarkeit und Datenaufbewahrung, um den Ressourcenaufwand für Ihre Workloads zu minimieren und gleichzeitig geschäftliche Anforderungen weiter erfüllen zu können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Definieren Sie SLAs, die Ihre Nachhaltigkeitsziele unterstützen und gleichzeitig Ihre geschäftlichen Anforderungen erfüllen.
- Definieren Sie SLAs neu, so dass sie die geschäftlichen Erwartungen erfüllen und nicht übertreffen.
- Gehen Sie Kompromisse ein, indem Sie Service Level in akzeptabler Weise verringern, um Auswirkungen auf die Nachhaltigkeit zu reduzieren.
- Nutzen Sie Entwurfsmuster, die geschäftskritische Funktionen priorisieren, und lassen Sie für nicht kritische Funktionen niedrigere Service Level zu (z. B. für Reaktions- und Wiederherstellungszeiten).

Ressourcen

Ähnliche Dokumente:

- Service Level Agreements (SLAs) für AWS
- Bedeutung von Dienstleistungsvereinbarungen für SaaS-Anbieter

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

SUS02-BP03 Beenden der Erstellung und Wartung nicht verwendeter Komponenten

Analysieren Sie Anwendungskomponenten (wie vorab kompilierte Berichte, Datensätze und statische Bilder) sowie Zugriffsmuster für Komponenten, um Redundanzen, eine zu geringe Auslastung und mögliche Kandidaten für die Außerbetriebnahme zu identifizieren. Konsolidieren Sie generierte Komponenten mit redundanten Inhalten (z. B. monatliche Berichte mit sich überschneidenden oder gemeinsam genutzten Datensätzen und Ausgaben), um für duplizierte Ausgaben genutzte

Ressourcen zu entfernen. Deaktivieren Sie nicht verwendete Komponenten (z. B. Bilder von Produkten, die nicht mehr verkauft werden), um genutzte Ressourcen freizugeben und die Zahl der Ressourcen zu reduzieren, die zur Unterstützung von Workloads verwendet werden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwalten Sie statische Komponenten und entfernen Sie Komponenten, die nicht mehr benötigt werden.
- Verwalten Sie die Generierung von Komponenten und beenden Sie die Generierung, wenn die Komponenten nicht mehr benötigt werden, oder entfernen Sie diese.
- Konsolidieren Sie sich überschneidende generierte Komponenten, um eine redundante Verarbeitung zu entfernen.
- Weisen Sie Dritte an, die Erstellung und Speicherung von Komponenten einzustellen, die in Ihrem Auftrag verwaltet und nicht mehr benötigt werden.
- Weisen Sie Dritte an, in Ihrem Auftrag erstellte redundante Komponenten zu konsolidieren.

Ressourcen

Ähnliche Dokumente:

- Optimizing your AWS Infrastructure for Sustainability, Part II: Storage (Ihre AWS-Infrastruktur für Nachhaltigkeit optimieren, Teil II: Speicher)

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

SUS02-BP04 Optimieren der geografischen Platzierung von Workloads für Benutzerstandorte

Analysieren Sie Netzwerkzugriffsmuster, um zu erkennen, aus welchen geographischen Regionen Ihre Kunden Verbindungen herstellen. Wählen Sie Regionen und Services, die die Entfernungen reduzieren, über die Netzwerkdatenverkehr übertragen werden muss, um die Zahl der Netzwerkressourcen zu verringern, die zur Unterstützung Ihres Workloads erforderlich sind.

Typische Anti-Muster:

- Sie wählen die Region des Workloads auf der Grundlage Ihres eigenen Standorts aus.

Vorteile der Nutzung dieser bewährten Methode: Die Platzierung von Workloads in der Nähe der jeweiligen Kunden bietet die geringstmögliche Latenz und verringert gleichzeitig die Bewegung der Daten durch das Netzwerk und damit die Umweltauswirkungen.

Risikostufe bei fehlender Befolgung dieser Best Practice: Mittel

Implementierungsleitfaden

- Wählen Sie die Regionen für Ihre Workload-Bereitstellung auf der Grundlage der folgenden zentralen Elemente aus:
  - Ihr Nachhaltigkeitsziel: wie unter Auswahl von Regionen erläutert.
  - Standort Ihrer Daten: Für datenintensive Anwendungen (wie etwa Big Data oder Machine Learning) sollte der Anwendungscode so nahe wie möglich zu den Daten ausgeführt werden.
  - Standort Ihrer Benutzer: Für benutzerorientierte Anwendungen sollten Sie eine Region wählen, die nahe der Kundenbasis des Workloads liegt.
  - Weitere Einschränkungen: Berücksichtigen Sie auch Einschränkungen wie die Sicherheit und Compliance., wie unter „Relevante Aspekte bei der Wahl einer Region für Ihre Workloads“ erläutert.
- Verwendung Sie AWS Local Zones für Workloads wie Video-Rendering oder grafikintensive virtuelle Desktop-Anwendungen. Mit Local Zones können Sie von allen Vorteilen profitieren, die sich durch die Platzierung der Datenverarbeitungs- und Speicherressourcen in der Nähe Ihrer Endbenutzer ergeben.
- Verwenden Sie lokale Zwischenspeicherung oder AWS-Caching-Lösungen für häufig genutzte Ressourcen zur Verbesserung der Leistung, zur Verringerung der Datenbewegung und zur Reduzierung der Umweltauswirkungen.
  - Verwendung Sie Amazon CloudFront für die Zwischenspeicherung statischer Inhalte wie Bilder, Skripts und Videos sowie dynamischer Inhalte wie API-Antworten oder Webanwendungen.
  - Verwendung Sie Amazon ElastiCache für die Zwischenspeicherung von Inhalten für Webanwendungen.
  - Verwendung Sie DynamoDB Accelerator für die Add-in-Speicher-Beschleunigung für Ihre DynamoDB-Tabellen.
- Nutzen Sie Services, die Ihnen dabei helfen können, Code näher an den Nutzern Ihres Workloads auszuführen:

  - Verwendung Sie Lambda@Edge für rechenintensive Anwendungen, die ausgeführt werden, wenn sich Objekte nicht im Zwischenspeicher befinden.
  - Verwendung Sie Amazon CloudFront-Funktionen für einfache Anwendungsfälle wie HTTP(s)-Anfragen oder Antwortmanipulationen, die von kurzlebigen Funktionen ausgeführt werden können.
  - Verwendung Sie AWS IoT Greengrass für die Ausführung lokaler Rechenoperationen, Messaging sowie die Datenzwischenspeicherung für verbundene Geräte.
- Nutzen Sie Verbindungspooling, um die erneute Nutzung von Verbindungen zu ermöglichen und die Zahl der erforderlichen Ressourcen zu reduzieren.
- Verwenden Sie verteilte Datenspeicher, die nicht auf persistente Verbindungen und synchrone Updates angewiesen sind, um regionale Benutzergruppen zu unterstützen.
- Ersetzen Sie vorab bereitgestellte statische Netzwerkkapazität durch geteilte dynamische Kapazitäten und teilen Sie die Auswirkungen von Netzwerkkapazitäten auf die Nachhaltigkeit mit anderen Abonnenten.

Ressourcen

Zugehörige Dokumente:

- Optimieren Ihrer AWS-Infrastruktur für Nachhaltigkeit, Teil III: Netzwerke
- Amazon ElastiCache-Dokumentation
- Was ist Amazon CloudFront?
- Wichtigste Amazon CloudFront-Funktionen
- Lambda@Edge
- CloudFront-Funktionen
- AWS IoT Greengrass

Zugehörige Videos:

- Nachhaltig entwickeln mit AWS

Zugehörige Beispiele:

- Workshops zu AWS-Netzwerken

SUS02-BP05 Optimieren von Ressourcen für Teammitglieder im Hinblick auf die ausgeführten Aktivitäten

Optimieren Sie die Ressourcen, die Teammitgliedern zur Verfügung gestellt werden, um negative Auswirkungen auf die Nachhaltigkeit zu minimieren und gleichzeitig ihre Anforderungen zu erfüllen. Beispielsweise können Sie komplexe Vorgänge wie Rendering und Kompilierung auf intensiv genutzten, geteilten Cloud-Desktops statt auf weniger ausgelasteten Einzelbenutzersystemen mit hohem Energieverbrauch ausführen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Stellen Sie Workstations und andere Geräte entsprechend ihrer Verwendung bereit.
- Verwenden Sie virtuelle Desktops und Anwendungs-Streaming, um Upgrade- und Geräteanforderungen zu begrenzen.
- Verschieben Sie prozessor- oder arbeitsspeicherintensive Aufgaben in die Cloud.
- Evaluieren Sie die Auswirkungen von Prozessen und Systemen auf die Lebenszyklen von Geräten. Wählen Sie Lösungen aus, die den Bedarf für Geräteaussetzungen minimieren und gleichzeitig die geschäftlichen Anforderungen erfüllen.
- Implementieren Sie eine Remote-Verwaltung für Geräte, um die Zahl der Geschäftsreisen zu reduzieren.

Ressourcen

Ähnliche Dokumente:

- Was ist Amazon WorkSpaces?
- Amazon AppStream 2.0-Dokumentation
- NICE DCV
- AWS Systems Manager Fleet Manager

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

Software- und Architekturmuster

Frage

- SUS 3 Wie können Sie Software- und Architekturmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

SUS 3 Wie können Sie Software- und Architekturmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Implementieren Sie Muster für den Lastausgleich und die Wahrung einer konsistent hohen Nutzung der bereitgestellten Ressourcen, um die Zahl der genutzten Ressourcen zu minimieren. Komponenten werden möglicherweise aufgrund von Änderungen des Benutzerverhaltens über die Zeit nicht mehr genutzt. Prüfen Sie Muster und Architekturen, um nicht ausreichend genutzte Komponenten zu konsolidieren und so die Nutzung insgesamt zu erhöhen. Nehmen Sie Komponenten außer Betrieb, die nicht mehr benötigt werden. Identifizieren Sie die Leistung Ihrer Workload-Komponenten und optimieren Sie die Komponenten, die die meisten Ressourcen verbrauchen. Achten Sie auf die Geräte, mit denen Ihre Kunden auf Ihre Services zugreifen, und implementieren Sie Muster, um den Bedarf für Geräte-Upgrades zu minimieren.

Bewährte Methoden:

SUS03-BP01 Optimieren von Software und Architektur für asynchrone und geplante Aufträge

Verwenden Sie effiziente Softwaredesigns und Architekturen, um die Zahl der für einzelne Arbeitseinheiten im Durchschnitt benötigten Ressourcen zu minimieren. Implementieren Sie Mechanismen für die gleichmäßige Nutzung von Komponenten, um die Zahl der Ressourcen zu reduzieren, die zwischen Aufgaben nicht genutzt werden, und die Auswirkungen von Lastspitzen zu minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verschieben Sie Anforderungen, die nicht sofort verarbeitet werden müssen, in Warteschlangen.
- Intensivieren Sie die Serialisierung, um die Nutzung über Ihre Pipeline hinweg gleichmäßig zu gestalten.
- Modifizieren Sie die Kapazität einzelner Komponenten, um zu vermeiden, dass ungenutzte Ressourcen auf Eingaben warten.

- Richten Sie Puffer ein und legen Sie Ratenbegrenzungen fest, um die Nutzung externer Services zu optimieren.
- Verwenden Sie die jeweils effizienteste verfügbare Hardware für Ihre Software-Optimierungen.
- Verwenden Sie warteschlangenbasierte Architekturen, Pipeline-Verwaltung und On-Demand-Instance-Worker, um die Nutzung für Batch-Verarbeitungen zu maximieren.
- Planen Sie Aufgaben, um Lastspitzen und das Konkurrieren um Ressourcen bei gleichzeitiger Ausführung zu vermeiden.
- Planen Sie Aufträge für Tageszeiten ein, an denen die Kohlendioxidemissionen am geringsten sind.

Ressourcen

Ähnliche Dokumente:

- Was ist Amazon Simple Queue Service?
- Was ist Amazon MQ?
- Skalierung auf Basis von Amazon SQS
- Was ist AWS Step Functions?
- Was ist AWS Lambda?
- Verwenden von AWS Lambda mit Amazon SQS
- Was ist Amazon EventBridge?

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)
- Moving to event-driven architectures (Umstieg auf ereignisgesteuerte Architekturen)

SUS03-BP02 Entfernen von Workload-Komponenten mit geringer oder keiner Nutzung oder Faktorwechsel

Überwachen Sie die Workload-Aktivität, um Änderungen bei der Nutzung einzelner Komponenten über die Zeit zu erkennen. Entfernen Sie ungenutzte Komponenten, die nicht mehr benötigt werden. Setzen Sie wenig genutzte Ressourcen neu ein, um die Verschwendung von Ressourcen zu begrenzen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Analysieren Sie (mithilfe von Indikatoren wie Transaktionsfluss und API-Aufrufen) die Last für funktionale Komponenten, um nicht oder nicht ausreichend genutzte Komponenten zu identifizieren.
- Nehmen Sie Komponenten außer Betrieb, die nicht mehr benötigt werden.
- Führen Sie einen Faktorwechsel für nicht ausreichend genutzte Komponenten durch.
- Konsolidieren Sie nicht ausreichend genutzte Ressourcen mit anderen Ressourcen, um die Nutzungseffizienz zu verbessern.

Ressourcen

Ähnliche Dokumente:

- Was ist AWS X-Ray?
- Was ist Amazon CloudWatch?
- Verwenden von ServiceLens zur Überwachung des Zustands Ihrer Anwendungen
- Automated Cleanup of Unused Images in Amazon ECR (Automatische Bereinigung von nicht verwendeten Images in Amazon ECR)

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

SUS03-BP03 Optimieren von Codebereichen, die die meiste Zeit oder die meisten Ressourcen verbrauchen

Überwachen Sie die Workload-Aktivität, um die Anwendungskomponenten zu identifizieren, die die meisten Ressourcen verbrauchen. Optimieren Sie den Code, der innerhalb dieser Komponenten ausgeführt wird, um die Ressourcennutzung zu minimieren und die Leistung zu maximieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Überwachen Sie die Leistung als Funktion der Ressourcennutzung, um Komponenten mit einem hohen Ressourcenbedarf pro Arbeitseinheit als Ziele für Optimierungen zu identifizieren.
- Verwenden Sie einen Code-Profiler, um die Codebereiche zu identifizieren, die die meiste Zeit oder die meisten Ressourcen verwenden.
- Ersetzen Sie Algorithmen durch effizientere Versionen, die dasselbe Ergebnis erzielen.
- Verwenden Sie Hardwarebeschleunigung, um die Effizienz von Codeblöcken mit langen Ausführungszeiten zu verbessern.
- Verwenden Sie das jeweils effizienteste Betriebssystem und die optimale Programmiersprache für den Workload.
- Entfernen Sie unnötige Sortierungen und Formatierungen.
- Verwenden Sie Datenübertragungsmuster, die die Ressourcennutzung basierend auf der Häufigkeit der Änderung von Daten und ihrer Nutzung minimieren. Sie können z. B. Statusänderungsinformationen zu einem Client übertragen. So werden keine Ressourcen für Abfragen verbraucht, die wertlose Meldungen mit „Keine Änderung“ zurückgeben.

Ressourcen

Ähnliche Dokumente:

- Was ist Amazon CloudWatch?
- Was ist Amazon CodeGuru Profiler?
- FPGA-Instances
- Die AWS SDKs für die Entwicklung in AWS

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

SUS03-BP04 Optimieren der Auswirkungen auf Geräte und Ausrüstung von Kunden

Identifizieren Sie die Geräte und Einrichtungen, mit denen Ihre Kunden Ihre Services nutzen, ihren voraussichtlichen Lebenszyklus und die finanziellen und nachhaltigkeitsbezogenen Auswirkungen der Ersetzung dieser Komponenten. Implementieren Sie Softwaremuster und Architekturen, die es für

Kunden unnötig machen, Geräte zu ersetzen oder ihre Ausrüstung zu aktualisieren. Implementieren Sie beispielsweise neue Funktionen, die Code verwenden, der mit älterer Hardware und älteren Betriebssystemversionen abwärtskompatibel ist, oder gestalten Sie die Größe von Nutzlasten so, dass sie die Speicherkapazitäten der Zielgeräte nicht überschreiten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Inventarisieren Sie die Geräte, die Ihre Kunden verwenden.
- Führen Sie Tests mithilfe verwalteter Gerätefarmen mit repräsentativer Hardware durch, um die Auswirkungen von Änderungen zu verstehen. Iterieren Sie Entwicklungsschritte, um die Zahl der unterstützten Geräte zu maximieren.
- Berücksichtigen Sie beim Erstellen von Nutzlasten Netzwerkbandbreite und Latenz und implementieren Sie Funktionen, mit denen Ihre Anwendungen auch über Verbindungen mit geringer Bandbreite und hoher Latenz gut funktionieren.
- Verarbeiten Sie Datennutzlasten vorab, um die Anforderungen an lokale Verarbeitung und Datenübertragung zu reduzieren.
- Führen Sie rechenintensive Aktivitäten (z. B. das Rendern von Bildern) serverseitig aus oder nutzen Sie Anwendungs-Streaming, um das Benutzererlebnis auf älteren Geräten zu verbessern.
- Segmentieren und paginieren Sie Ausgaben, besonders für interaktive Sitzungen, um Nutzlasten zu verwalten und lokale Speicheranforderungen zu begrenzen.

Ressourcen

Ähnliche Dokumente:

- Was ist AWS Device Farm?
- Amazon AppStream 2.0-Dokumentation
- NICE DCV
- Amazon Elastic Transcoder-Dokumentation

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

SUS03-BP05 Verwenden von Softwaremustern und Architekturen, die Datenzugriffs- und Speichermuster optimal unterstützen

Identifizieren Sie, wie Daten in Ihrem Workload verwendet, von Benutzern genutzt, übertragen und gespeichert werden. Wählen Sie Technologien aus, die die Anforderungen an Datenverarbeitung und -speicherung minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Analysieren Sie Ihre Datenzugriffs- und -speichermuster.
- Speichern Sie Datendateien in effizienten Dateiformaten wie Parquet, um eine unnötige Verarbeitung (z. B. beim Ausführen von Analysen) zu verhindern und den insgesamt bereitgestellten Speicher zu reduzieren.
- Nutzen Sie Technologien, die nativ mit komprimierten Daten funktionieren.
- Verwenden Sie die Datenbank-Engine, die das dominierende Abfragemuster jeweils am besten unterstützt.
- Verwalten Sie Ihre Datenbankindizes so, dass die Indexdesigns die effiziente Ausführung von Abfragen unterstützen.
- Wählen Sie Netzwerkprotokolle aus, die die Menge der genutzten Netzwerkkapazitäten reduzieren.

Ressourcen

Ähnliche Dokumente:

- Athena-Compression-Support-Dateiformate
- KOPIEREN aus Spaltendatenformaten mit Amazon Redshift
- Umwandeln Ihres Eingangsdatensatzformats in Firehose
- Formatierungsoptionen für ETL-Eingaben und -Ausgaben in AWS Glue
- Verbessern der Abfrageleistung in Amazon Athena durch Umwandlung in Spaltenformate
- Laden komprimierter Datendateien aus Amazon S3 in Amazon Redshift
- Überwachen der DB-Last mit Performance Insights in Amazon Aurora
- Überwachen der DB-Last mit Performance Insights in Amazon RDS
- AWS IoT FleetWise

Ähnliche Videos:

- Building Sustainably on AWS (Nachhaltig entwickeln mit AWS)

Datenmuster

Frage

- SUS 4 Wie können Sie Datenzugriffs- und -nutzungsmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

SUS 4 Wie können Sie Datenzugriffs- und -nutzungsmuster zur Unterstützung Ihrer Nachhaltigkeitsziele nutzen?

Implementieren Sie Verfahren für die Datenverwaltung, die den zur Unterstützung Ihres Workloads bereitgestellten Speicher und die für dessen Nutzung erforderlichen Ressourcen reduzieren. Identifizieren Sie Ihre Daten und verwenden Sie Speichertechnologien und Konfigurationen, die den Unternehmenswert und die Nutzung der Daten optimal unterstützen. Verschieben Sie die Daten während des Lebenszyklus zu effizienteren Speichern mit geringerer Leistung, wenn die Anforderungen abnehmen. Löschen Sie Daten, die nicht mehr benötigt werden.

Bewährte Methoden:

SUS04-BP01 Implementieren einer Richtlinie für die Klassifizierung von Daten

Klassifizieren Sie Daten, um ihre Bedeutung für geschäftliche Ergebnisse zu verstehen. Nutzen Sie diese Informationen, um festzulegen, wann Daten in einen energieeffizienteren Speicher übertragen oder auf sichere Weise gelöscht werden können.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Legen Sie Anforderungen für die Verteilung, Aufbewahrung und Löschung Ihrer Daten fest.
- Markieren Sie Volumes und Objekte, um die Metadaten zur Angabe ihrer Verwaltung aufzuzeichnen, einschließlich Datenklassifizierung.
- Prüfen Sie die Umgebung regelmäßig auf nicht markierte und nicht klassifizierte Daten und klassifizieren und markieren Sie die Daten entsprechend.

Ressourcen

Ähnliche Dokumente:

- Datenklassifizierungsprozess
- Nutzung der AWS Cloud zur Unterstützung der Datenklassifizierung
- Tag-Richtlinien von AWS Organizations

SUS04-BP02 Verwenden von Technologien, die Datenzugriff und Speichermuster unterstützen

Nutzen Sie einen Speicher, der den Zugriff auf Ihre Daten und ihre Speicherung jeweils optimal unterstützt, um die Zahl der bereitgestellten Ressourcen zu minimieren und gleichzeitig den Workload zu unterstützen. Beispielsweise verbrauchen SSD-Laufwerke mehr Energie als magnetische Laufwerke und sollten nur für aktive Datenanwendungsfälle eingesetzt werden. Verwenden Sie für Daten, auf die nicht häufig zugegriffen wird, einen energieeffizienten Archivierungsspeicher.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Überwachen Sie die Datenzugriffsmuster.
- Migrieren Sie Daten auf der Basis der Zugriffsmuster zur jeweils optimal geeigneten Technologie.
- Migrieren Sie Archivdaten zu Speichern, die für diesen Zweck vorgesehen sind.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS Volume-Typen
- Amazon EC2-Instance-Speicher
- Amazon S3-Intelligent-Tiering
- Verwenden von Amazon S3-Speicherklassen
- Was ist Amazon CloudWatch?
- Was ist Amazon S3 Glacier?

Ähnliche Videos:

- Architectural Patterns for Data Lakes on AWS (Architekturmodelle für Data Lakes in AWS)

SUS04-BP03 Verwenden von Lebenszyklusrichtlinien zum Löschen nicht notwendiger Daten

Verwalten Sie den Lebenszyklus aller Daten und setzen Sie automatisch Löschfristen durch, um die Speicheranforderungen Ihres Workloads insgesamt zu minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Definieren Sie Lebenszyklusrichtlinien für alle Arten von Datenklassifizierungen.
- Legen Sie automatisierte Lebenszyklusrichtlinien zur Durchsetzung von Lebenszyklusregeln fest.
- Löschen Sie nicht verwendete Volumes und Snapshots.
- Aggregieren Sie Daten auf der Basis von Lebenszyklusregeln. wenn möglich.

Ressourcen

Ähnliche Dokumente:

- Amazon ECR-Lebenszyklusrichtlinien
- Amazon EFS-Lebenszyklusmanagement
- Amazon S3-Intelligent-Tiering
- Bewerten von Ressourcen mit AWS-Config-Regeln
- Verwalten des Speicherlebenszyklus in Amazon S3
- Objektlebenszyklus-Richtlinien in AWS Elemental MediaStore

Ähnliche Videos:

- Amazon S3 Lifecycle (Amazon-S3-Lebenszyklus)

SUS04-BP04 Minimieren übermäßiger Bereitstellungen im Blockspeicher

Erstellen Sie zur Minimierung des insgesamt bereitgestellten Speichers Blockspeicher mit Größenzuweisungen entsprechend dem jeweiligen Workload. Verwenden Sie elastische Volumes, um den Speicher bei wachsenden Datenmengen erweitern zu können, ohne die Größe des an

Computing-Ressourcen angefügten Speichers ändern zu müssen. Überprüfen Sie elastische Volumes regelmäßig und verkleinern Sie zu große Volumes, um sie an den aktuellen Datenumfang anzupassen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Überwachen Sie die Nutzung Ihrer Daten-Volumes.
- Verwenden Sie elastische Volumes und verwaltete Blockdaten-Services, um automatisch zusätzlichen Speicher zuzuweisen, wenn die Menge der persistenten Daten wächst.
- Legen Sie Zielstufen für die Nutzung Ihrer Daten-Volumes fest und passen Sie die Größe von Volumes an, die außerhalb der erwarteten Bereiche liegen.
- Passen Sie die Größe schreibgeschützter Volumes an die Datenmenge an.
- Migrieren Sie Daten zu Objektspeichern, um zu vermeiden, dass die überschüssige Kapazität aus Volumes mit fester Größe im Blockspeicher bereitgestellt wird.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS Elastic Volumes
- Amazon FSx-Dokumentation
- Was ist Amazon CloudWatch?
- Was ist Amazon Elastic File System?

SUS04-BP05 Entfernen nicht benötigter oder redundanter Daten

Duplizieren Sie Daten nur wie notwendig, um den insgesamt genutzten Speicher zu minimieren. Verwenden Sie Backup-Technologien, die Daten auf Datei- und Blockebene deduplizieren. Verwenden Sie Konfigurationen mit Redundant Array of Independent Drives (RAID) nur, wenn dies zur Erfüllung von SLAs (Service Level Agreements) notwendig ist.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwenden Sie Mechanismen, die Daten auf Block- und Objektebene deduplizieren können.

- Verwenden Sie eine Backup-Technologie, die inkrementelle Backups erstellen und Daten auf Block-, Datei- und Objektebene deduplizieren kann.
- Verwenden Sie RAID nur, wenn dies zur Erfüllung von SLAs notwendig ist.
- Zentralisieren Sie Protokoll- und Nachverfolgungsdaten, deduplizieren Sie identische Protokolleinträge und richten Sie Mechanismen für die Anpassung der Ausführlichkeit ein, wenn notwendig.
- Füllen Sie Zwischenspeicher nur vorab aus, wenn dies begründet werden kann.
- Richten Sie Überwachung und Automatisierung für den Zwischenspeicher ein, um die Größe des Zwischenspeichers entsprechend anzupassen.
- Entfernen Sie veraltete Bereitstellungen und Komponenten aus Objektspeichern und Edge-Zwischenspeichern, wenn Sie neue Versionen Ihres Workloads veröffentlichen.

Ressourcen

Ähnliche Dokumente:

- Amazon EBS-Snapshots
- Ändern der Aufbewahrung von Protokolldaten in CloudWatch Logs
- Datendeduplizierung in Amazon FSx for Windows File Server
- Funktionen von Amazon FSx for ONTAP einschließlich Datendeduplizierung
- Invalidieren von Dateien auf Amazon CloudFront
- Using AWS Backup to back up and restore Amazon EFS file systems (Verwenden von AWS Backup, um Amazon EFS-Dateisysteme zu sichern und wiederherzustellen)
- Was ist Amazon CloudWatch Logs?
- Arbeiten mit Backups in Amazon RDS

Ähnliche Beispiele:

- Übung: Optimierung von Datenmustern mithilfe von Amazon Redshift Data Sharing

SUS04-BP06 Verwenden geteilter Dateisysteme oder Objektspeicher für den Zugriff auf allgemeine Daten

Nutzen Sie geteilten Speicher und zentrale Datenquellen, um Datenduplizierungen zu vermeiden und den Gesamtspeicherbedarf des Workloads zu reduzieren. Rufen Sie Daten nur wie notwendig

aus dem geteilten Speicher ab. Trennen Sie die Verbindung mit nicht verwendeten Datenträgern, um mehr Ressourcen verfügbar zu machen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Migrieren Sie Daten in einen geteilten Speicher, wenn die Daten mehrfach genutzt werden.
- Rufen Sie Daten nur wie notwendig aus dem geteilten Speicher ab.
- Löschen Sie Daten gemäß Ihren Nutzungsmustern und implementieren Sie eine Gültigkeitsdauer-Funktionalität (Time to Live, TTL) zur Verwaltung zwischengespeicherter Daten.
- Trennen Sie Volumes von Clients, die sie nicht aktiv verwenden.

Ressourcen

Ähnliche Dokumente:

- Amazon FSx
- Strategien für Zwischenspeicher
- Was ist Amazon Elastic File System?
- Was ist Amazon S3?

SUS04-BP07 Minimieren von Datenübertragungen zwischen Netzwerken

Verwenden Sie einen geteilten Speicher und greifen Sie über regionale Datenspeicher auf Daten zu, um den Gesamtbedarf an Netzwerkressourcen zur Unterstützung von Datenübertragungen für den Workload zu minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Speichern Sie Daten so nahe an den Nutzern wie möglich.
- Partitionieren Sie regional genutzte Services so, dass regionsspezifische Daten in der Region gespeichert werden, in der sie genutzt werden.
- Verwenden Sie eine Duplizierung auf Blockebene statt auf Datei- oder Objektebene, wenn Änderungen über das Netzwerk kopiert werden.
- Komprimieren Sie Daten, bevor sie über das Netzwerk übertragen werden.

Ressourcen

Ähnliche Dokumente:

- Optimieren Ihrer AWS-Infrastruktur für Nachhaltigkeit, Teil III: Netzwerke
- Globale AWS-Infrastruktur
- Hauptfunktionen von Amazon CloudFront einschließlich CloudFront Globales Edge-Netzwerk
- Compressing HTTP requests in Amazon OpenSearch Service (Komprimieren von HTTP-Anforderungen in Amazon OpenSearch Service)
- Zwischenkomprimierung von Daten mit Amazon EMR
- Laden komprimierter Datendateien aus Amazon S3 in Amazon Redshift
- Bereitstellen komprimierter Dateien mit Amazon CloudFront

SUS04-BP08 Sichern von Daten nur in dem Fall, dass ihre erneute Erstellung schwierig ist

Sichern Sie zur Minimierung der Speichernutzung nur Daten, die einen Unternehmenswert besitzen oder zur Erfüllung von Compliance-Anforderungen benötigt werden. Prüfen Sie Backup-Richtlinien und vermeiden Sie einen flüchtigen Speicher, der in einem Wiederherstellungsszenario keinen Wert bietet.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwenden Sie Ihre Datenklassifizierung, um festzulegen, welche Daten gesichert werden müssen.
- Schließen Sie Daten aus, die Sie auf einfache Weise neu erstellen können.
- Schließen Sie flüchtige Daten von Backups aus.
- Schließen Sie sekundäre Kopien von Daten aus, es sei denn, die für die Wiederherstellung dieser Daten von einem gemeinsamen Standort benötigte Zeit überschreitet Ihre Service Level Agreements (SLAs).

Ressourcen

Ähnliche Dokumente:

- Using AWS Backup to back up and restore Amazon EFS file systems (Verwenden von AWS Backup, um Amazon EFS-Dateisysteme zu sichern und wiederherzustellen)

- Amazon EBS-Snapshots
- Arbeiten mit Backups in Amazon Relational Database Service

Hardwaremuster

Frage

- SUS 5 Wie können Hardwareverwaltung und Nutzungsverfahren Ihre Nachhaltigkeitsziele unterstützen?

SUS 5 Wie können Hardwareverwaltung und Nutzungsverfahren Ihre Nachhaltigkeitsziele unterstützen?

Suchen Sie nach Möglichkeiten, die Auswirkungen auf die Nachhaltigkeit Ihrer Workloads durch Änderungen der Methoden für die Hardwareverwaltung zu reduzieren. Minimieren Sie den Umfang der für die Bereitstellung erforderlichen Hardware und wählen Sie die jeweils effizienteste Hardware für den jeweiligen Workload aus.

Bewährte Methoden:

SUS05-BP01 Verwenden der geringstmöglichen Menge an Hardware zur Erfüllung Ihrer Anforderungen

Mit den Möglichkeiten der Cloud können Sie häufige Änderungen für Ihre Workload-Implementierungen ausführen. Aktualisieren Sie bereitgestellte Komponenten, wenn sich Ihre Anforderungen ändern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Unterstützen Sie die horizontale Skalierung und nutzen Sie Automatisierung, um bei steigender Nutzung die Bereitstellung aufzuskalieren und bei sinkender Nutzung die Bereitstellung abzuskalieren.
- Skalieren Sie für variable Workloads in kleinen Schritten.
- Passen Sie die Skalierung an zyklische Nutzungsmuster an (z. B. Gehaltsbuchhaltungssysteme mit intensiven Verarbeitungsaktivitäten alle zwei Wochen), wenn die Last über Tage, Wochen, Monate oder Jahre unterschiedlich ist.

- Verhandeln Sie SLAs (Service Level Agreements), die eine vorübergehende Reduzierung von Kapazitäten zulassen, während die Bereitstellung von Ersatzressourcen automatisiert wird.

Ressourcen

Ähnliche Dokumente:

- AWS Compute Optimizer-Dokumentation
- Ausführen von Lambda: Leistungsoptimierung
- Auto Scaling-Dokumentation

SUS05-BP02 Verwenden von Instance-Typen mit den geringsten Auswirkungen

Überwachen Sie kontinuierlich die Einführung neuer Instance-Typen und nutzen Sie Verbesserungen bei der Energieeffizienz, einschließlich Instance-Typen, die zur Unterstützung spezifischer Workloads bestimmt sind, wie z. B. Machine-Learning-Trainings und -Inferenzen und Videotranskodierung.

Typische Anti-Muster:

- Sie verwenden lediglich eine Familie von Instances.
- Sie verwenden nur x86-Instances.
- Sie geben einen Instance-Typ in Ihrer Amazon EC2 Auto Scaling-Konfiguration an.
- Sie verwenden AWS-Instances in einer Weise, für die sie nicht gedacht sind (beispielsweise Computing-optimierte Instances für speicherintensive Workloads).
- Sie evaluieren nicht regelmäßig die Instance-Typen.
- Sie prüfen nicht die Empfehlungen von AWS-Dimensionierungstools wie etwa AWS Compute Optimizer.

Vorteile der Nutzung dieser bewährten Methode: Durch die Verwendung energieeffizienter und korrekt dimensionierter Instances können Sie die Umweltauswirkungen und die Kosten Ihrer Workloads deutlich reduzieren.

Risikostufe bei fehlender Befolgung dieser Best Practice: Niedrig

Implementierungsleitfaden

- Informieren Sie sich über Instance-Typen, die die Umweltauswirkungen Ihrer Workloads reduzieren können.
  - Abonnieren Sie Neuerungen bei AWS, um bei den neuesten AWS-Technologien und -Instances auf dem Laufenden zu bleiben.
  - Informieren Sie sich über die verschiedenen AWS-Instance-Typen.
  - Informieren Sie sich über auf AWS Graviton basierende Instances, die die höchste Leistung pro Watt in Amazon EC2 bieten; sehen Sie sich dazu Folgendes an: re:Invent 2020 - Deep dive on AWS Graviton2 processor-powered Amazon EC2 instances (Ein tiefer Einblick in vom AWS-Graviton2-Prozessor unterstütze EC2-Instances) und Deep dive into AWS Graviton3 and Amazon EC2 C7g instances (Ein tiefer Einblick in AWS-Graviton3- und EC2-C7g-Instances).
- Planen und übertragen Sie Ihre Workloads auf Instance-Typen mit den geringsten Auswirkungen.
  - Definieren Sie einen Prozess zur Evaluierung neuer Funktionen oder Instances für Ihre Workloads. Nutzen Sie die Agilität in der Cloud, um schnell zu testen, wie neue Instance-Typen die ökologische Nachhaltigkeit Ihrer Workloads verbessern können. Nutzen Sie Proxy-Metriken, um zu messen, wie viele Ressourcen Sie für eine Arbeitseinheit benötigen.
  - Modifizieren Sie Ihren Workload nach Möglichkeit so, dass er mit unterschiedlichen Zahlen von CPUs und Arbeitsspeichergrößen kompatibel ist, um die größtmögliche Auswahl an Instance-Typen zu erhalten.
  - Erwägen Sie die Übertragung Ihres Workloads zu auf Graviton basierenden Instances, um die Leistungseffizienz Ihres Workloads zu verbessern (siehe AWS Graviton-Schnellstart und AWS Graviton2 für ISVs). Denken Sie an diese Überlegungen bei der Übertragung von Workloads zu auf AWS Graviton basierenden Amazon Elastic Compute Cloud-Instances.
  - Erwägen Sie die Auswahl der AWS-Graviton-Option bei Ihrer Verwendung der verwalteten AWS-Services.
  - Migrieren Sie Ihren Workload zu Regionen mit Instances, die die geringsten nachhaltigkeitsbezogenen Auswirkungen bieten und dennoch Ihre geschäftlichen Anforderungen erfüllen.
  - Verwenden Sie für Machine-Learning-Workloads Amazon EC2-Instances, die auf benutzerdefinierten Amazon Machine Learning-Chips basieren, wie etwa AWS Trainium, AWS Inferentiaund Amazon EC2 DL1.
  - Verwendung Sie Amazon SageMaker Inference Recommender für die Dimensionierung des ML-Inferenz-Endpunkts.

  - Verwenden Sie für Workloads mit Echtzeit-Video-Transkodierung Amazon EC2-VT1-Instances.
  - Verwenden Sie für Workloads, bei denen es gelegentlich zu zusätzlichen Kapazitätsanforderungen kommt, Instances mit Spitzenlastleistung.
  - Verwenden Sie für zustandslose und fehlertolerante Workloads Amazon EC2 Spot-Instances , um die allgemeine Nutzung der Cloud zu verbessern und die nachhaltigkeitsbezogenen Auswirkungen nicht genutzter Ressourcen zu reduzieren.
- Betreiben und optimieren Sie Ihre Workload-Instance.
  - Prüfen Sie für kurz andauernde Workloads Amazon CloudWatch-Instance-Metriken wie die `CPU-Nutzung` , um festzustellen, ob die Instance eventuell zu wenig oder gar nicht genutzt wird.
  - Prüfen Sie für stabile Workloads AWS-Dimensionierungstools wie etwa AWS Compute Optimizer in regelmäßigen Intervallen, um Möglichkeiten zur Optimierung und zur korrekten Dimensionierung der Instances zu erkennen.

Ressourcen

Zugehörige Dokumente:

- Optimieren Ihrer AWS-Infrastruktur für Nachhaltigkeit, Teil I: Datenverarbeitung
- AWS Graviton Processor
- AWS Inferentia
- AWS Trainium
- Amazon EC2 DL1
- Amazon EC2-Instances mit Spitzenlastleistung
- Amazon EC2-Flotten zur Kapazitätsreservierung
- Amazon EC2-Spot-Flotte
- Amazon EC2 Spot-Instances
- Amazon EC2-VT1-Instances
- Amazon EC2-Instance-Typen
- AWS Compute Optimizer
- Funktionen: Lambda-Funktionskonfiguration

Zugehörige Videos:

- Deep dive on AWS Graviton2 processor-powered Amazon EC2 instances (Ein tiefer Einblick in vom Graviton2-Prozessor unterstütze Instances)
- Deep dive into AWS Graviton3 and Amazon EC2 C7g instances (Ein tiefer Einblick in AWS-Graviton3- und EC2-C7g-Instances)

Zugehörige Beispiele:

- Lab: Empfehlungen zur Dimensionierung
- Lab: Dimensionierung mit Compute Optimizer
- Lab: Optimieren von Hardwaremustern und Beobachtung von Nachhaltigkeits-KPIs

SUS05-BP03 Verwenden verwalteter Services

Mit verwalteten Services geht die Verantwortung für die Wahrung einer hohen durchschnittlichen Nutzung und die Optimierung der Nachhaltigkeit der bereitgestellten Hardware auf AWS über. Mit verwalteten Services können Sie die nachhaltigkeitsbezogenen Auswirkungen des Service über alle Mandanten des Service verteilen und so Ihren Beitrag verringern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Migrieren Sie selbst gehostete Services zu verwalteten Services. Verwenden Sie beispielsweise verwaltete Amazon Relational Database Service (Amazon RDS)- Instances, anstatt Ihre eigenen Amazon RDS-Instances auf Amazon Elastic Compute Cloud (Amazon EC2)zu verwalten, oder verwenden Sie verwaltete Container-Services wie AWS Fargate,anstatt Ihre eigene Container-Infrastruktur zu implementieren.

Ressourcen

Ähnliche Dokumente:

- AWS Fargate,
- Amazon DocumentDB
- Amazon Elastic Kubernetes Service (EKS)
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)

- Amazon Redshift
- Amazon Relational Database Service (RDS)

SUS05-BP04 Optimieren der GPU-Nutzung

Grafikverarbeitungseinheiten (Graphics Processing Units, GPUs) können sehr viel Energie verbrauchen. Zahlreiche GPU-Workloads sind hoch variabel, z. B. Rendern, Transkodieren sowie Machine-Learning-Trainings und -Modellierungen. Führen Sie GPU-Instances nur für die benötigte Zeit aus und automatisieren Sie ihre Außerbetriebnahme, wenn sie nicht benötigt werden, um den Ressourcenverbrauch zu minimieren.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Verwenden Sie GPUs nur für Aufgaben, bei denen Sie effizienter als CPU-basierte Alternativen sind.
- Automatisieren Sie die Freigabe nicht genutzter GPU-Instances.
- Verwenden Sie eine flexible Grafikbeschleunigung anstelle dedizierter GPU-Instances.
- Nutzen Sie speziell für Ihren Workload entwickelte Hardware.

Ressourcen

Ähnliche Dokumente:

- Accelerated Computing
- AWS Inferentia
- AWS Trainium
- Beschleunigtes Computing für EC2-Instances
- Amazon EC2-VT1-Instances
- Amazon Elastic Graphics

Entwicklungs- und Bereitstellungsprozess

Frage

SUS 6 Wie können Ihre Entwicklungs- und Bereitstellungsprozesse Ihre Nachhaltigkeitsziele unterstützen?

Reduzieren Sie nachhaltigkeitsbezogene Auswirkungen, indem Sie Ihre Entwicklungs-, Test- und Bereitstellungsmethoden ändern.

Bewährte Methoden:

SUS06-BP01 Einführen von Methoden, die schnelle Verbesserungen für die Nachhaltigkeit ermöglichen

Testen und validieren Sie potenzielle Verbesserungen, bevor Sie sie in der Produktion bereitstellen. Berücksichtigen Sie die Testkosten bei der Berechnung des potenziellen zukünftigen Nutzens einer Verbesserung. Entwickeln Sie kostengünstige Testmethoden, um kleine Verbesserungen einzuführen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

Implementierungsleitfaden

- Fügen Sie Ihrem Entwicklungsprozess Anforderungen an die Nachhaltigkeit hinzu.
- Erlauben Sie die parallele Ausführung von Ressourcen, um Verbesserungen für die Nachhaltigkeit zu entwickeln, zu testen und bereitzustellen.
- Testen und validieren Sie potenzielle Verbesserungen in Bezug auf Nachhaltigkeit, bevor Sie sie in der Produktion bereitstellen.
- Testen Sie mögliche Verbesserungen mit der geringstmöglichen Zahl repräsentativer Komponenten.
- Stellen Sie getestete Verbesserungen in Bezug auf Nachhaltigkeit in der Produktion bereit, sobald sie verfügbar sind.

Ressourcen

Ähnliche Dokumente:

- AWS unterstützt nachhaltige Lösungen

Ähnliche Beispiele:

- Übung: Verwandeln von Kosten- und Nutzungsberichten in Effizienzberichte

SUS06-BP02 Konstantes Aktualisieren Ihres Workloads

Aktuelle Betriebssysteme, Bibliotheken und Anwendungen können die Workload-Effizienz verbessern und die Nutzung effizienterer Technologien unterstützen. Eine aktuelle Software kann darüber hinaus Funktionen für eine genauere Messung der Auswirkungen Ihres Workloads bereitstellen, da die Anbieter mit ihrer Software ebenfalls Nachhaltigkeitsziele erfüllen müssen.

Typische Anti-Muster:

- Sie gehen davon aus, dass Ihre derzeitige Architektur unverändert bleibt und im Laufe der Zeit nicht aktualisiert wird.
- Sie haben keine Systeme oder regelmäßigen Besprechungen zur Prüfung, ob aktualisierte Software und Pakete mit Ihrem Workload kompatibel sind.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie begründen.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie einen Prozess einrichten, um Ihren Workload aktuell zu halten, können Sie neue Funktionen und Kapazitäten nutzen, Probleme lösen und die Workload-Effizienz verbessern.

Risikostufe bei fehlender Befolgung dieser Best Practice: Niedrig

Implementierungsleitfaden

- Definieren Sie einen Prozess und einen Zeitplan zur Evaluierung neuer Funktionen oder Instances für Ihre Workloads. Nutzen Sie die Agilität in der Cloud, um schnell zu testen, wie neue Funktionen Ihre Workloads auf den folgenden Gebieten verbessern können:
  - Reduzierung von Auswirkungen auf die Nachhaltigkeit.
  - Erzielen von Leistungseffizienzen.
  - Beseitigen von Hindernissen für geplante Verbesserungen.
  - Verbesserung Ihrer Fähigkeit für die Messung von und den Umgang mit Nachhaltigkeitsauswirkungen.
- Inventarisierung Ihrer Workload-Software und -Architektur sowie Identifizierung von Komponenten, die aktualisiert werden müssen. Sie können AWS Systems Manager Inventory verwenden, um

Betriebssystem (BS)-, Anwendungs- und Instance-Metadaten von Ihren Amazon EC2-Instances zu erfassen und so schnell zu verstehen, welche Instances die Software und die Konfigurationen ausführen, die Ihre Softwarerichtlinie erfordert, und welche Instances aktualisiert werden müssen.

- Verständnis der Aktualisierung der Komponenten Ihres Workloads.
  - Umgang mit Aktualisierungen von Amazon Machine Images (AMI) für Linux oder von Windows Server-Images mit EC2 Image Builder.
  - Sie sollten Amazon Elastic Container Registry (Amazon ECR) mit Ihrer vorhandenen Pipeline verwenden, um Amazon Elastic Container Service (Amazon ECS)-Images und Amazon Elastic Kubernetes Service-Images zu verwalten.
  - AWS Lambda beinhaltet Versionsmanagementfunktionen.
- Verwenden Sie Automatisierung für den Aktualisierungsvorgang, um den Aufwand für die Bereitstellung neuer Funktionen zu reduzieren und Fehler zu begrenzen, die durch manuelle Prozesse verursacht werden. Verwenden Sie Tools wie AWS Systems Manager Patch Manager zur Automatisierung des Systemaktualisierungsprozesses und zur Planung der Aktivität mithilfe von AWS Systems Manager Maintenance Windows.

Ressourcen

Zugehörige Dokumente:

- AWS-Architekturzentrum
- Neuerungen bei AWS
- AWS Developer Tools
- AWS Systems Manager Patch Manager

Zugehörige Beispiele:

- Well-Architected Labs: Bestands- und Patch-Verwaltung
- Lab: AWS Systems Manager

SUS06-BP03 Höhere Auslastung von Entwicklungsumgebungen

Verwenden Sie Automatisierung und „Infrastructure as Code“, um Vorproduktionsumgebungen in Betrieb zu nehmen, wenn notwendig, und bei Nichtverwendung zu deaktivieren. Eine typische Vorgehensweise besteht in der Planung von Verfügbarkeitszeiten, die mit den Arbeitszeiten der

Entwicklungsteams übereinstimmen. Der Ruhezustand ist ein nützliches Tool, um den aktuellen Status beizubehalten und Instances nur bei Bedarf schnell zu aktivieren.. Verwenden Sie Instance-Typen mit Burst-Kapazität, Spot-Instances, elastische Datenbank-Services, Container und andere Technologien, um Entwicklungs- und Testkapazitäten an die Nutzung anzupassen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

- Nutzen Sie die Automatisierung, um die Auslastung Ihrer Entwicklungs- und Testumgebungen zu maximieren.
- Nutzen Sie die Automatisierung, um den Lebenszyklus Ihrer Entwicklungs- und Testumgebungen zu verwalten.
- Verwenden Sie die geringstmögliche Zahl repräsentativer Umgebungen, um mögliche Verbesserungen zu entwickeln und zu testen.
- Verwenden Sie On-Demand-Instances, um Entwicklergeräte zu ergänzen.
- Nutzen Sie die Automatisierung, um die Effizienz Ihrer Entwicklungsressourcen zu maximieren.
- Verwenden Sie Instance-Typen mit Burst-Kapazität, Spot-Instances und andere Technologien, um die Entwicklungskapazität an der Nutzung auszurichten.
- Nutzen Sie native Cloud-Services für den sicheren Instance-Shell-Zugriff, statt Bastion-Host-Flotten bereitzustellen.

Ressourcen

Ähnliche Dokumente:

- AWS Systems Manager Session Manager
- Amazon EC2-Instances mit Spitzenlastleistung
- Was ist AWS CloudFormation?

SUS06-BP04 Verwenden verwalteter Gerätefarmen für Tests verwenden

Verwaltete Gerätefarmen verteilen die nachhaltigkeitsbezogenen Auswirkungen der Hardwarefertigung und der Ressourcennutzung über zahlreiche Beteiligte. Verwaltete Gerätefarmen stellen verschiedene Gerätetypen bereit, unterstützen auch ältere und weniger verbreitete Hardware und vermeiden nachhaltigkeitsbezogene Auswirkungen durch unnötige Geräte-Upgrades seitens Kunden.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Niedrig

Implementierungsleitfaden

Führen Sie Tests mithilfe verwalteter Gerätefarmen mit repräsentativer Hardware durch, um die Auswirkungen von Änderungen zu verstehen. Iterieren Sie Entwicklungsschritte, um die Zahl der unterstützten Geräte zu maximieren.

Ressourcen

Ähnliche Dokumente:

- Was ist AWS Device Farm?

Hinweise

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