Operationalisierung der Agenten-KI am AWS

AWS Präskriptive Leitlinien

AWS Präskriptive Leitlinien: Operationalisierung der Agenten-KI am AWS

Table of Contents

Operationalisierung agentischer KI auf AWS

Aaron Sempf, Brad Ryan, Bhargs Srivathsan und Akhil Bhaskar, Amazon Web Services

August 2025 (Geschichte des Dokuments)

Agentic AI ist kein Feature, sondern ein neues betriebliches Paradigma. Organizations, die in eine disziplinierte Architektur, vertrauenswürdige Frameworks und geschäftsorientierte Bereitstellungsmodelle investieren, werden die nächste Generation adaptiver, intelligenter Unternehmen anführen.

Agentic AI steht für die Konvergenz von autonomen Softwareagenten und generativer KI. Sie verbindet die Entscheidungsfindung und das zielgerichtete Verhalten von Agenten mit den Fähigkeiten großer Sprachmodelle zum Sprachverständnis und zur Sprachgenerierung (). LLMs Diese Agenten können in dynamischen Unternehmensumgebungen vernünftig denken, handeln, sich anpassen und zusammenarbeiten. Um dieses Potenzial zu nutzen, müssen Unternehmen ihre Denkweise von der Modellbereitstellung hin zur Agenteninfrastruktur ändern.

Dieser Leitfaden bietet eine Unternehmensstrategie zur Umwandlung von künstlicher Intelligenz von isolierten Experimenten hin zu einer wertschöpfenden Infrastruktur auf Unternehmensebene. Er kann Ihnen dabei helfen, intelligente Agenten in Ihre Workflows einzubinden und dabei Governance, Skalierbarkeit und Geschäftsausrichtung zu gewährleisten.

Wichtigste Schwerpunktbereiche und Empfehlungen

Dieser Leitfaden konzentriert sich auf die folgenden grundlegenden Bereiche bei der Operationalisierung agentischer KI. Organisatorische und geschäftliche Empfehlungen werden für jeden Schwerpunktbereich bereitgestellt:

- Schwerpunktbereich 1: Erläutern Sie die Absicht und den Umfang des Agenten— Stimmen Sie die Mitarbeiter auf die Geschäftsprioritäten und kognitiven Engpässe ab. Behandeln Sie Agenten als digitale Teamkollegen, nicht nur als Tools.
- Schwerpunktbereich 2: Design für Kombinierbarkeit und Zusammenarbeit— Nutzen Sie Multi-Agenten-Systeme mit modularer Architektur, semantischen Protokollen und dynamischer Delegierung durch Arbiter-Agenten.
- Schwerpunktbereich 3: Architekt für Mehrmandantenfähigkeit und Kontrolle— Aufbau einer skalierbaren, mandantenorientierten Infrastruktur mit gemeinsam genutzten Agentendiensten, zentraler Verwaltung und rollenbasiertem Zugriff.

- Schwerpunktbereich 4: Vertrauen durch Identität, Leitplanken und Beobachtbarkeit aufbauen— Sorgen Sie für Rückverfolgbarkeit, Laufzeitkontrollen und Erklärbarkeit, um das Vertrauen der Stakeholder zu gewinnen.
- Schwerpunktbereich 5: Verwaltung des Lebenszyklus— Einrichtung von CI/CD-Pipelines (Continuous Integration and Continuous Deployment), umgehende Versionierung, Telemetrie und kontinuierliche Weiterbildung zur Unterstützung der Leistung und Effizienz der KI von Behörden.
- Schwerpunktbereich 6: Abstimmung der Agentenmodelle mit den Geschäftsmodellen— Monetarisieren Sie die Fähigkeiten Ihrer Agenten mithilfe nutzungsbasierter Modelle, interner ROI-Metriken und kommerzieller Angebote.

Sie können die Empfehlungen in diesem Leitfaden nutzen, um Ihr Unternehmen auf agentische KI in großem Maßstab vorzubereiten. Darin wird dargelegt, wie Unternehmen ihre Umstrukturierung auf agentischer KI ausrichten müssen. Dazu gehören der Aufbau von Teams DevOps für Agenten (AgentOps), interoperable Systeme und Change-Management-Strategien, die eine breite Akzeptanz ermöglichen. Es betont das Denken bei der Entscheidungsfindung und die Ausrichtung auf das AWS Well-Architected Framework.

Zielgruppe

Dieser Leitfaden richtet sich an Unternehmensarchitekten, AI/ML technische Leiter und Strategen für die digitale Transformation, die Agentensysteme entwerfen und skalieren, KI in Kerngeschäftsabläufe einbetten und Agenten in Produktionsumgebungen LLMs operationalisieren und autonome Agenten einsetzen. Um die Konzepte und Empfehlungen in diesem Leitfaden zu verstehen, sollten Sie mit modernen Cloud-nativen Architekturen und verteilten Systemen, großen Sprachmodellen, grundlegenden Modellfunktionen und den Prinzipien der KI-Governance und der Plattformtechnik vertraut sein. DevOps

Ziele

Durch die Umsetzung der Empfehlungen in diesem Leitfaden kann Ihr Unternehmen die folgenden Geschäftsergebnisse erzielen:

- Schnellere Entscheidungsfindung und Workflow-Ausführung durch autonome, zielorientierte Agenten, die menschliche Engpässe und kognitive Belastung reduzieren.
- Skalierbare, kosteneffiziente Bereitstellung intelligenter Funktionen in allen Geschäftsbereichen über wiederverwendbare, mandantenfähige Agentenplattformen.

- Höhere Widerstandsfähigkeit, mehr Vertrauen und bessere Steuerung von KI-Systemen, was eine sichere Einführung in regulierten, geschäftskritischen oder kundenorientierten Umgebungen ermöglicht.

Über diese Inhaltsserie

Dieser Leitfaden ist Teil einer Reihe über agentic AI on. AWS Weitere Informationen und die anderen Leitfäden dieser Reihe finden Sie unter Agentic AI auf der Prescriptive Guidance-Website. AWS

Strategische Grundlagen für agentische KI

Agentensysteme sind nicht neu. Softwareagenten, darunter Robotic Process Automation (RPA) und Decision Engines, gibt es schon seit Jahrzehnten. Sie waren jedoch einfach und deterministisch und darauf ausgelegt, vordefinierten Regeln und symbolischer Logik zu folgen, um sich wiederholende Aufgaben mit geringer Variation auszuführen. Mit dem Aufkommen der generativen KI hat sich das Spiel verändert. Große Sprachmodelle (LLMs) können jetzt komplexe Eingaben interpretieren, dynamisch Antworten generieren und Wissen schnell synthetisieren. Sie können Ihre Entscheidungsfreiheit jetzt ohne spröde oder fest programmierte Logik skalieren. Jetzt können Agenten Argumente liefern, Entscheidungen treffen, Tools aufrufen, sich an den Kontext anpassen und sich mit anderen Agenten über Workflows hinweg abstimmen. Sie können autonom auf Ziele hinarbeiten, ihr Gedächtnis behalten und über Ergebnisse nachdenken.

Die reine Leistungsfähigkeit reicht jedoch nicht aus. Intelligenz ohne Integration führt zu Neuheit, nicht zu Wirkung. Um das Potenzial leistungsfähiger Produkte voll auszuschöpfen LLMs, müssen Unternehmen von isolierten Experimenten abrücken und auf technisch ausgereifte Ökosysteme umstellen. Agenten müssen wie Dienstleistungen in Produktionsqualität behandelt werden, die derselben Disziplin unterliegen wie jedes andere Unternehmenssystem. Dazu gehören Unternehmensführung, Beobachtbarkeit, sichere Identitätsmodelle und Lebenszyklusmanagement. Sie müssen auch zu realen Geschäftsergebnissen und nicht zu spekulativem Potenzial führen. Diese Systeme sollten mit klaren Grenzen für Entscheidungsfindung und Fehlertoleranz konzipiert sein. Es ist wichtig, automatisierte Wiederherstellungsmechanismen, Leistungsüberwachung in Echtzeit und skalierbares Ressourcenmanagement zu integrieren. Auf diese Weise können Sie den dynamischen, nicht deterministischen Charakter von Agenteninteraktionen bewältigen und gleichzeitig konsistente Serviceniveaus in allen Unternehmensabläufen aufrechterhalten.

Auf grundlegender Ebene müssen Unternehmen überdenken, wie Intelligenz in die Betriebsstruktur eingebettet wird. Agenten müssen so konzipiert sein, dass sie sich in die Kernsysteme integrieren lassen, Unternehmensrichtlinien einhalten und messbaren Nutzen bieten. Sie müssen in großem Umfang, abteilungs-, domänen- und benutzerkontextübergreifend arbeiten können. Bei der Operationalisierung agentischer KI geht es letztlich um den Nutzen. Es ist der Unterschied zwischen dem Einsatz von KI, die isolierte Aufgaben ausführt, und dem Einsatz von Agenten, die Ihr Geschäftsmodell weiterentwickeln.

Agentic AI steht für eine neue Betriebsphilosophie, die einen grundlegenden Wandel in der Art und Weise erfordert, wie wir Systeme, Prozesse und Mitarbeiter angehen, um Intelligenz unternehmensweit zu skalieren. Agenten werden zu strategischen Ressourcen, die die menschlichen

Fähigkeiten erweitern. Durch die Integration von KI in ihre Abläufe können Unternehmen Erkenntnisse gewinnen, die den Geschäftswert steigern, die menschlichen Fähigkeiten erweitern und komplexe Arbeitsabläufe optimieren.

Strategische Schwerpunktbereiche für agentische KI

Um von frühen Prototypen zu serienreifen und wertschöpfenden Systemen überzugehen, benötigen Teams eine kohärente Strategie, die Architektur-, Prozess- und Produktdenken miteinander verbindet.

Viele Unternehmen gehen KI immer noch mit einer werkzeugsorientierten oder modellorientierten Denkweise an. Generative KI hat das Experimentieren verstärkt, aber oft ohne klare Ausrichtung auf die Geschäftsstrategie oder messbare Ergebnisse. Ohne eine definierte strategische Rolle laufen Agenten Gefahr, zu neuartigen Experimenten zu werden, die Ressourcen verbrauchen, anstatt einen skalierbaren Wert zu liefern. Um die strategische Rolle der agentischen KI zu etablieren, müssen Unternehmen mit Geschäftsprioritäten beginnen. Identifizieren Sie Bereiche mit kognitiver Überlastung, Entscheidungsengpässen oder fragmentierten Arbeitsabläufen, in denen Autonomie Abhilfe schaffen kann. Verwenden Sie domänenspezifische Problemstellungen, um die Verantwortlichkeiten der Agenten zu definieren. Behandeln Sie Agenten als digitale Teamkollegen — nicht als Tools —, die vernünftig denken, delegieren und sich anpassen können.

Entscheidungswissenschaften sind die Disziplin, in der Datenwissenschaft, Analytik und Verhaltensmodellierung kombiniert werden, um die Entscheidungsfindung zu verbessern. Sie sollte früh in den Prozess der Agentenarchitektur integriert werden, um das Design an den Geschäftsergebnissen auszurichten. Durch die Identifizierung von Entscheidungsmustern, die Simulation von Kompromissen und die Quantifizierung der Auswirkungen auf den Wert können Sie anhand von Entscheidungswissenschaften genau bestimmen, wo die Autonomie der Behörden den größten Nutzen bringen kann. Entscheidungswissenschaften können Entscheidungen beschleunigen, Fehler reduzieren und Anpassungen in Echtzeit ermöglichen. Diese datengestützte Grundlage stützt das Agentendesign auf messbaren Erkenntnissen und ermöglicht eine engere Integration mit bestehenden Unternehmenstechnologien wie Regelmodulen, Analyseplattformen und Vorhersagemodellen.

Um die strategische Rolle der Agenten besser zu verstehen, werden in diesem Abschnitt grundlegende Schwerpunktbereiche vorgestellt, die das Rückgrat für die Operationalisierung der Agenten-KI bilden. Jeder ist einer Kernaufgabe zugeordnet, die aus der Sicht eines technischen Leiters, Architekten oder Produkteigentümers erledigt werden muss, der dafür verantwortlich ist, wie Agenten konzipiert und gestaltet werden. Bei diesen Schwerpunktbereichen handelt es sich nicht um aufeinanderfolgende Schritte. Es lohnt sich, sie während des gesamten Systemlebenszyklus erneut zu überprüfen, um widerstandsfähige, skalierbare und monetarisierbare Agentenökosysteme zu schaffen.

Dieser Abschnitt enthält die folgenden Schwerpunktbereiche:

Schwerpunktbereich 1: Erläutern Sie die Absicht und den Umfang des Agenten

Zu erledigende Job: „Helfen Sie mir sicherzustellen, dass jeder Agent ein echtes Problem mit klaren Grenzen löst, nicht nur mit einer coolen Demo.“

Bei Agentic AI geht es nicht nur darum, Fähigkeiten aufzubauen. Es geht darum, das richtige Problem auf die richtige Art und Weise zu lösen, um das richtige Ergebnis zu erzielen. Das beginnt damit, dass Sie sich über die Absicht der agentischen KI-Lösung völlig im Klaren sind.

Strategie

Allzu oft beginnen Unternehmen mit dem, was das Modell leisten kann (z. B. anrufen APIs, Fragen beantworten oder Zusammenfassungen erstellen) und passen dann einen entsprechenden Anwendungsfall an. Dies führt dazu, dass der Umfang ständig erweitert wird, die Integration schlecht ist und Agenten technisch beeindruckend, aber betrieblich nutzlos sind. Definieren Sie stattdessen zunächst die Rolle des Agenten anhand konkreter Fragen wie den folgenden:

- Für welches konkrete Ergebnis ist der Agent verantwortlich?
- Für wen handelt er?
- Wer profitiert davon?
- Wo beginnt und endet die Autonomie des Agenten?
- Was passiert, wenn sie versagt?

Ein gut ausgebildeter Agent hat eine klare Aufgabe, definierte Verantwortlichkeiten und messbare Erfolgskriterien. Stellen Sie sich den Agenten nicht als Assistenten oder Chatbot vor. Geben Sie

ihm stattdessen eine Berufsbezeichnung. Stellen Sie sich das als Kundenberater, als Mitarbeiter für Produktrücksendungen oder als Compliance-Monitor vor.

Betonen Sie bei der Einbindung von Stakeholdern oder Kunden die Skalierbarkeit und Anpassungsfähigkeit agentischer KI-Systeme. Diese Agenten entwickeln sich mit dem Unternehmen weiter und verbessern sich kontinuierlich durch Lernen und Feedback. Um Widerstände zu verringern und die Einführung zu beschleunigen, sollten Sie hervorheben, wie die Tools von Agentic unter Berücksichtigung des Einfühlungsvermögens der Mitarbeiter konzipiert wurden. Sie bieten Transparenz, Kontrolle und optionale Übersteuerungsmechanismen, die Vertrauen schaffen. Anstatt Mitarbeiter zu ersetzen, verbessern Agenten die menschlichen Fähigkeiten und die Entscheidungsfindung und helfen den Mitarbeitern, auf dem Laufenden zu bleiben und sich auf wichtige Aufgaben zu konzentrieren.

Der Schlüssel zu einer erfolgreichen Implementierung liegt darin, die KI der Agenturen auf spezifische, wirkungsvolle Geschäftsergebnisse auszurichten. Ermutigen Sie Teams und Partner, mit gezielten Pilotprojekten zu beginnen, die sichtbare Probleme lösen. Schnelle Erfolge sorgen für eine messbare Kapitalrendite (ROI), sorgen für interne Zustimmung und schaffen Dynamik für eine breitere Akzeptanz.

Um die Akzeptanz und den Reifegrad zu fördern, können Unternehmen das Agentendesign anhand eines Evolutionsmodells gestalten. Die Autonomie der Agenten, die Komplexität und die Auswirkungen auf das Geschäft nehmen zunehmend zu. Dieses Modell besteht aus folgenden Phasen:

- Beobachteragenten gewinnen Erkenntnisse aus Geräuschen. Ein Beispiel ist ein Marktsentiment-Agent, der die Markenwahrnehmung auf allen digitalen Kanälen verfolgt.
- Assistenzagenten unterstützen die menschliche Entscheidungsfindung. Ein Beispiel ist ein Geschäftsberater, der Wettbewerbsdaten und Marktbedingungen für Vertriebsteams zusammenfasst.
- Autonome Agenten agieren unabhängig innerhalb definierter Grenzen. Ein Beispiel ist ein Agent für die Ressourcenzuweisung, der die Cloud-Infrastruktur dynamisch an den Bedarf anpasst.
- Orchestrator-Agenten koordinieren Workflows mit mehreren Agenten. Ein Beispiel ist ein Supply-Chain-Optimierungsagent, der Interaktionen zwischen Inventar-, Logistik- und Prognoseagenten verwaltet.
- Innovative Agenten eröffnen neue strategische Möglichkeiten. Ein Beispiel ist ein Innovationsagent für Geschäftsmodelle, der Markttrends analysiert und neue Einnahmequellen empfiehlt.

Wenn sich die Akteure an diesen strategischen Ergebnissen und Reifegraden orientieren, wird der Fokus erhöht, die Akzeptanz beschleunigt und das Vertrauen der Stakeholder gestärkt.

Um die Abstimmung in diesem Schwerpunktbereich zu unterstützen AWS-Services, kann Amazon Quick beispielsweise wichtige Leistungsindikatoren (KPIs) visualisieren, die mit agentengesteuerten Ergebnissen verknüpft sind. Sie können Amazon verwenden, CloudWatch um das Verhalten von Agenten, Leistungskennzahlen und den Systemzustand nahezu in Echtzeit zu überwachen. Nutzen Sie das betriebliche Feedback, um die Interaktionen der Agenten und die Ressourcennutzung zu optimieren. AWS CloudTrailkann in frühen Experimentierungs- und Verfeinerungsphasen Einblick in die Aktivitäten und Integrationsmuster der Agenten geben.

Geschäftlicher Nutzen der Definition von Absicht und Umfang

Die Einführung agentischer KI stellt einen entscheidenden Wandel in der Art und Weise dar, wie Unternehmen mit digitaler Transformation und operativer Exzellenz umgehen. Dabei geht es nicht nur um Automatisierung. Es geht darum, intelligente Autonomie zu ermöglichen, die Entscheidungsfindung und Wertschöpfung beschleunigt.

Zu den wichtigsten Geschäftstreibern gehören:

- Wettbewerbsvorteil — Early Adopters verschaffen sich strategische Vorteile durch schnellere Einblicke, besseren Service und anpassungsfähige Abläufe.
- Verbesserung des Kundenerlebnisses — Agenten bieten personalisierten, stets verfügbaren Support in Echtzeit, der die Zufriedenheit und Loyalität steigert.
- Betriebseffizienz — Agentic AI reduziert die kognitive Belastung des Menschen erheblich, indem komplexe, sich wiederholende Entscheidungsaufgaben automatisiert werden. Dies ermöglicht es den Mitarbeitern, sich auf höherwertige Aktivitäten zu konzentrieren, und die Kosten können gesenkt werden.

Zu den realen Anwendungsfällen in allen Branchen gehören:

- Finanzdienstleistungen — KI-Agenten könnten personalisierte Finanzberatung anbieten und Betrug aufdecken.
- Gesundheitswesen — Agenten für Triage und Behandlungspläne könnten den klinischen Durchsatz verbessern.
- Einzelhandel — Agenten könnten als intelligente Einkaufsassistenten agieren oder das Inventar in Echtzeit optimieren.

- Fertigung — Agenten könnten vorausschauende Wartungsarbeiten durchführen oder Lieferketten koordinieren.

Schwerpunktbereich 2: Design für Kombinierbarkeit und Zusammenarbeit

Zu erledigende Job: „Lassen Sie mich Agenten so erstellen, wie ich Dienste baue — modular und testbar, sodass sie nach Bedarf zusammengestellt und orchestriert werden können.“

Viele KI-Bemühungen beginnen als monolithische, modellzentrierte Pilotprojekte. Sie sind nützlich, aber es ist schwierig, sie domänenübergreifend zu skalieren oder an komplexe Probleme anzupassen. Nutzen Sie Verbindungen, wenn diese Wirkstoffe so konzipiert sind, dass sie zusammenarbeiten. In der Technologie bezeichnet Zusammensetzbarkeit die Kombination modularer Komponenten zu einer flexiblen, skalierbaren Lösung, die sich an Veränderungen anpassen kann. Ohne Zusammensetzbarkeit ist Intelligenz an bestimmte Arbeitsabläufe gebunden. Darüber hinaus führt die Zusammenarbeit zwischen Agenten zu komplexen Orchestrierung, Statusverwaltung und Protokollverhandlungen, für die herkömmliche Automatisierungsteams möglicherweise nicht gewappnet sind.

Strategie

Machen Sie sich das Multi-Agent-Paradigma zunutze. Modellieren Sie Agenten wie Unternehmensabteilungen: modular, spezialisiert und interoperabel. Definieren Sie klare Schnittstellen, gemeinsam genutzte Kontextformate und Standardkommunikationsprotokolle wie Model Context Protocol (MCP) oder Agent2Agent (A2A). Verwenden Sie Orchestrierungsmuster für mehrere Agenten wie Schwarm-, Graph- oder hierarchische Koordination. Diese Muster helfen Agenten dabei, Funktionen zu erkennen und Dienste dynamisch voneinander anzufordern, entweder in parallel, sequentiellen oder konsensbasierten Workflows, je nach Aufgabenstruktur und Vertrauensstufe.

Verwenden Sie einen Arbiter-Agent, um eine skalierbare und kontrollierte Zusammenarbeit zu fördern. Bei dieser Art von Agent handelt es sich um eine neutrale Behörde, die die Delegierung von Aufgaben auf der Grundlage bekannter Fähigkeiten und Ausweichstrategien erleichtert. Ein Schiedsrichter ist zwar kein zentraler Kontrolleur, spielt aber eine entscheidende Rolle in Bezug auf Vertrauen und Einhaltung von Vorschriften. Er stellt sicher, dass sensible oder regulierte Aufgaben nur an Agenten weitergeleitet werden, die die Identitäts- und Richtlinienanforderungen erfüllen. Es fungiert als Gatekeeper für richtliniengebundene Workflows. Es erzwingt Isolation und ermöglicht

eine nachvollziehbare Delegierung. Entscheidend ist, dass ein Schiedsrichter kein Engpass darstellt, sondern dass er mit sich selbst koordinierenden Stellen koexistiert, die horizontal agieren. peer-to-peer Diese Agenten delegieren Unteraufgaben, teilen sich den Kontext und lösen Abhängigkeiten direkt auf.

Dieses hybride Modell unterstützt sowohl deterministische Zuweisungen (durch den Arbiter-Agent) als auch emergente Zusammenarbeit. Es verbindet Struktur mit Flexibilität. Innerhalb dieser Architektur können Agenten in die folgenden speziellen Rollen eingeteilt werden:

- Entscheidungsträger, wie z. B. Personen, die Richtlinien durchsetzen, Ressourcen zuweisen und Risikobewerter
- Wissensagenten wie Kontext-Aggregatoren, Mustererkennungsprogramme und Anomaliedetektoren
- Ausführungsagenten wie Aufgabenausführende, Qualitätskontrolleure und Integrationsmanager

Für eine effektive Koordination müssen Systeme mit mehreren Agenten robuste Interaktionsprotokolle für die Statusverwaltung, die Fehlerbehebung und die Konfliktlösung unterstützen. Dies fördert Stabilität und Rechenschaftspflicht, auch wenn die Agenten unabhängig voneinander arbeiten.

Legen Sie klare Regeln für die Skalierung fest, z. B. für die lastbasierte Instanziierung von Agenten, die kontextsensitive Ressourcenzuweisung und die automatische Erkennung und Registrierung von Funktionen. Diese Maßnahmen tragen dazu bei, dass das System je nach Bedarf oder Komplexität dynamisch wächst.

Entwerfen Sie Agenten so, dass sie ready-to-use Module innerhalb eines verteilten Messaging-Substrats sind. Beispielsweise könnten Sie Amazon EventBridge mit A2A oder MCP anstelle von isolierten Diensten verwenden. Führen Sie Versionierung, CI/CD Pipelines und Agentenvorlagen ein, um die Systemstabilität zu unterstützen und gleichzeitig die interne Einführung und die Entwicklung des Lebenszyklus zu beschleunigen. Fördern Sie die Wiederverwendung und Standardisierung von Code, um Integrationsprobleme zu verringern und ein robustes Ökosystem zu fördern.

Zusammenarbeit ist ein Multiplikator. Sie ermöglicht Skalierbarkeit, Spezialisierung und Widerstandsfähigkeit in Umgebungen mit mehreren Agenten. Um diese dynamische Zusammenarbeit zu unterstützen, sollten Unternehmen eine einfache Kontrollebene für die Agentenkoordination einrichten. Diese Kontrollebene umfasst Folgendes:

- Funktionsregister, die definieren, was jeder Agent tun kann, und die versionierte Metadaten für die Peer-Discovery unterstützen
- Logik der Aufgabenverteilung, die Arbiter- oder Supervisor-Agenten verwendet, um Aufgaben auf der Grundlage von Kontext, Verfügbarkeit und Richtlinien weiterzuleiten
- Lebenszyklus- und Statusverfolgung, die Entscheidungskontext in Echtzeit und sichere Übergaben ermöglicht

Kontrollebenen stellen sicher, dass Systeme mit mehreren Agenten erweiterbar, richtlinienkonform und fehlertolerant bleiben, ohne dass die Autorität zentralisiert oder der Betrieb verlangsamt wird.

Umgebungen mit mehreren Agenten bringen jedoch auch betriebliche Herausforderungen mit sich. Die Aufrechterhaltung des Kontextes zwischen den Interaktionen der Agenten, die Verwaltung des gemeinsamen Zustands und die Koordination von Aktionen können die Komplexität erhöhen und die Kosten erhöhen. Die Kosten können steigen, wenn Sie LLMs diese verbrauchenden Tokens bei der Kommunikation zwischen Agenten verwenden. Diese Kosten müssen gegen die zusätzlichen Geschäftsvorteile einer intelligenten, skalierbaren Autonomie abgewogen werden.

Um diesen Herausforderungen zu begegnen, sollten Sie Agentenplattformen in Betracht ziehen, die zentrale Anliegen abstrahieren, wie z. B. die folgenden:

- Standardisierte Kommunikationsprotokolle und semantische Formate
- Integrierte Orchestrierungslogik und dynamisches Routing
- Gemeinsames Kontext- und Speichermanagement zwischen Agenten
- Fallback-Behandlung und graziöse Degradation bei Ausfällen

Für Teams, die Strategien mit mehreren Agenten anwenden, ist es am besten, klein anzufangen und skalierbar zu planen. Beginnen Sie mit gezielten Einzelagentenlösungen, die echte Probleme lösen. Stellen Sie diese Agenten dann schrittweise zu einem kooperativen System zusammen, in dem jeder auf der Grundlage gemeinsamer Ziele und des systemweiten Kontextes Informationen ermitteln, koordinieren und delegieren kann.

Wichtig ist, dass eine robuste Fehlerbehandlung und eine angemessene Degradation die wichtigsten Entwurfsprinzipien sein müssen. Systeme mit mehreren Agenten sollten in der Lage sein, Teilabläufe fortzusetzen oder Backup-Logik zu initiieren, wenn Agenten nicht verfügbar sind oder ausfallen. Dies fördert die Zuverlässigkeit ohne starre Kopplung.

AWS-Services bieten robuste Funktionen zur Unterstützung dieser Architektur in großem Maßstab. Amazon EventBridge und EventBridge Pipes bieten das strukturierte, ereignisgesteuerte Backbone für Multi-Agent-Messaging. Ermöglicht die Verwaltung des modularen Verhaltens und AWS AppConfigermöglicht eine sichere, dynamische Konfigurationsumschaltung zwischen Agent-Instances. Um die gemeinsame Kontext- und Speicherverwaltung zu unterstützen, verwenden Sie Amazon DynamoDB für einfache, mandantenorientierte Statuspersistenz und schnellen Kontextabruf zwischen Agenten. Sie können Amazon Simple Storage Service (Amazon S3) verwenden, um strukturierte Prompt-Historien, gemeinsam genutzte Artefakte oder von Agenten generierte Ausgaben zu speichern. Für komplexere Workflows, die eine zustandsorientierte Koordination erfordern, AWS Step Functionskann es lang andauernde Prozesse mit Checkpoints und Fehlerwiederherstellungslogik orchestrieren. Zusammen helfen Ihnen diese Services dabei, zusammensetzbare, belastbare und semantisch verbundene Multi-Agenten-Systeme zu erstellen, die mit den Unternehmensanforderungen skaliert werden können.

Der geschäftliche Nutzen von Systemen mit mehreren Agenten

Während viele Unternehmen ihre KI-Reise mit Single-Agent-Lösungen beginnen, wird das volle Potenzial der agentischen KI durch skalierbare Multi-Agenten-Systeme ausgeschöpft. Diese Systeme sind der Schlüssel zur Lösung komplexer, verteilter Probleme und zur Schaffung robuster, flexibler KI-Ökosysteme, die sich mit den Geschäftsanforderungen weiterentwickeln.

Zu den wichtigsten Geschäftsvorteilen von Systemen mit mehreren Agenten gehören:

- Skalierbarkeit — Aufgaben und Workloads können auf spezialisierte Agenten verteilt werden, um Kapazität und Leistung zu erhöhen.
- Flexibilität — Agenten können mit minimaler Unterbrechung hinzugefügt, ersetzt oder geändert werden, was Flexibilität in dynamischen Umgebungen ermöglicht.
- Ausfallsicherheit — Dank redundanter Rollen und intelligentem Failover bleibt die Systemstabilität auch dann erhalten, wenn einzelne Agenten ausfallen.
- Spezialisierung — Speziell entwickelte Agenten führen Aufgaben effizienter und präziser aus.
- Kosteneffizienz — Wiederverwendbare Agentenkomponenten beschleunigen die Entwicklung und reduzieren die Kosten für die Bereitstellung neuer Funktionen.

Systeme mit mehreren Agenten erfordern zwar mehr Planung im Voraus, bieten aber langfristige Agilität, Geschwindigkeit und Innovationskapazität. Unternehmen, die in flexible Architekturen für die Zusammenarbeit mit Agenten investieren, sind in der Lage, neue KI-Funktionen schnell

bereitzustellen, sich an sich ändernde Anforderungen anzupassen und in einem zunehmend agentenorientierten Wettbewerbsumfeld eine Vorreiterrolle einzunehmen.

Schwerpunktbereich 3: Architekt für Mehrmandantenfähigkeit und Kontrolle

Zu erledigende Aufgabe: „Helfen Sie mir, die Agentennutzung auf mehrere Kunden zu skalieren, ohne die Kontrolle, Verantwortlichkeit oder Transparenz zu verlieren.“

Frühe Prototypen eignen sich gut, um den Wert isoliert nachzuweisen, aber die meisten Unternehmen müssen mehrere Kunden, Abteilungen oder Workflows gleichzeitig unterstützen. Das bedeutet, dass jeder Mitarbeiter innerhalb klar definierter Richtlinien-, Daten- und Identitätsgrenzen arbeiten muss. Ohne Mehrmandantenfähigkeit wird der Betrieb spröde und kostspielig, und die Verwaltung wird zu einem Flickenteppich.

Strategie

Folgen Sie den Prinzipien von Software-as-a-Service (SaaS) -Architekturen. Denken Sie beispielsweise an die Isolierung von Mandanten, die Durchsetzung von Richtlinien und die Ressourcenkontrolle. Entwickeln Sie Agenten und Orchestrierungsplattformen mit mandantenorientiertem Speicher, Konfiguration und Identität. Verwenden Sie Tagging, rollenbasierte Zugriffskontrolle (RBAC) und Scoping für Identitäts- und Zugriffsmanagement, um Grenzen durchzusetzen.

Führen Sie eine einheitliche Beobachtungsebene ein, in der die Agententelemetrie nach Mandantenkontext aggregiert wird. Implementieren Sie zentralisierte Policy-Engines und konfigurationsbasierte Funktionen zur Durchsetzung dynamischer Verhaltensregeln.

Entwickeln Sie die Agentenbereitstellung als Service. Ermöglichen Sie internen Teams oder Kunden, die Funktionen der Agenten skalierbar und gesteuert zu nutzen APIs. AWS bietet eine solide Grundlage für diese Muster. Sie können Amazon Cognito für die Verwaltung der Benutzer- und Mandantenidentität AWS Organizationssowie für Richtlinien zur Servicesteuerung (SCPs) für die kontoübergreifende Steuerung und AWS Resource Access Manager (AWS RAM) für die sichere gemeinsame Nutzung von Funktionen verwenden. Darüber hinaus AWS AppConfigkann das Verhalten der Agenten dynamisch nach Mandanten oder Umgebung verwaltet werden. Diese Dienste helfen bei der Durchsetzung von Grenzen und Richtlinien und unterstützen gleichzeitig die gemeinsame Infrastruktur.

Dieser Übergang von der statischen Bereitstellung zur dynamischen Bereitstellung macht aus agentischer KI eine unternehmensweite Plattform.

Der geschäftliche Nutzen von Multi-Tenant-Agent-Plattformen

Mehrmandantenfähigkeit ist mehr als nur ein architektonischer Vorteil — sie beschleunigt das Geschäft. Da sich intelligente Agenten in allen Abteilungen und Teams immer weiter ausbreiten, müssen Unternehmen das Wachstum unterstützen, ohne die Infrastruktur zu duplizieren oder die Unternehmensführung zu fragmentieren.

Zu den wichtigsten Geschäftsvorteilen von Mehrmandantensystemen gehören:

- Skalierbarkeit — Eine Agentenplattform mit mehreren Mandanten ermöglicht es internen Teams, Geschäftseinheiten oder Kunden, KI-Funktionen schneller zu integrieren, ohne dass maßgeschneiderte Umgebungen erforderlich sind.
- Kosteneffizienz — Eine gemeinsam genutzte Infrastruktur minimiert redundante Bereitstellungen, konsolidiert die Betriebskosten und vereinfacht die Wartung in allen Umgebungen.
- Steuerung und Risikominderung — Zentralisierte Richtlinienkontrollen, Identitätsmodelle und Beobachtbarkeit helfen den Mitarbeitern, sicherer und richtlinienkonformer zu arbeiten — und das bei allen Mandanten.
- Wiederverwendbarkeit von Diensten — Um die Wiederverwendung zu fördern und Doppelarbeit zu vermeiden, können Kundenbetreuer als interne Dienste angeboten werden, z. B. zur Erweiterung, Einhaltung von Vorschriften oder zur Zusammenfassung.

Zu den Anwendungsfällen für Systeme mit mehreren Mandanten gehören unter anderem die folgenden:

- Ein Compliance-Agent, der in allen Niederlassungen eingesetzt wird, passt seine Logik durch eine mandantenspezifische Konfiguration an die lokalen Vorschriften an. Dadurch entfällt die Notwendigkeit, separate Agenten für jede Region einzurichten.
- Ein interner Agent zur Workflow-Automatisierung bedient mehrere Abteilungen mit unterschiedlichen Datengrenzen und Berechtigungen. Er sorgt für Isolation und beschleunigt gleichzeitig die Aufgabenerfüllung.

Indem Unternehmen Agenten als multi-tenant-aware Dienste konzipieren, vermeiden sie den Aufwand isolierter KI-Initiativen. Stattdessen fördern sie eine einheitliche Informationsplattform. Diese

Architektur ermöglicht einen skalierbaren Rollout, betriebliche Konsistenz und einen besseren ROI. Sie macht es auch einfacher, die Einführung von KI im gesamten Unternehmen auszuweiten.

Schwerpunktbereich 4: Vertrauen durch Identität, Leitplanken und Beobachtbarkeit aufbauen

Zu erledigende Job: „Geben Sie mir die Gewissheit, dass die Agenten sicher und vorhersehbar handeln, insbesondere wenn niemand zuschaut.“

Autonome Agenten stellen traditionelle Kontrollmodelle in Frage. Ihre Fähigkeit, unabhängig zu denken und zu handeln, birgt Risiken, wenn sie nicht ordnungsgemäß verwaltet werden. Ohne klare Verantwortlichkeit, Überprüfbarkeit oder politische Einschränkungen können sie von ihrem beabsichtigten Verhalten abweichen. Um organisatorisches Vertrauen aufzubauen, ist mehr als nur technische Zuverlässigkeit erforderlich. Es erfordert Erklärbarkeit, Rechenschaftspflicht und Konsistenz.

Strategie

Bauen Sie ein Kontrollsystem auf, bei dem Identität an erster Stelle steht, als Rückgrat vertrauenswürdiger Autonomie. Jeder Agent muss mit einer überprüfbaren Identität, begrenzten Berechtigungen und einer nachvollziehbaren Ausführungshistorie arbeiten. Agenten sollten in ein Zero-Trust-Framework eingebettet werden, das Mandantenbindung, kontextbezogene Zugriffsvererbung und Laufzeitdurchsetzung durch Guardrails und Policy-Engines umfasst. Auf diese Weise können Sie die Aktionen der Agenten auf der Grundlage der Unternehmensregeln und der Risikosituation überprüfen, rückgängig machen oder einschränken.

Integrieren Sie die Vertrauensdurchsetzung während der Laufzeit durch intelligente Leitplanken. Dazu gehören Ratenkontrollen und Drosselungen auf der Grundlage von Verhaltensmustern oder Arbeitslastbedingungen, die Durchsetzung von Ressourcengrenzen sowie die auto-scaling und die Bewertung von Entscheidungen zur Risikobewertung. Erstellen Sie Auslöser, um human-in-the-loop Workflows zu aktivieren, wenn Schwellenwerte überschritten werden.

Jeder Mitarbeiter muss außerdem transparent und erklärbar sein. Integrieren Sie strukturierte Telemetrie mithilfe von Protokollierung, Ablaufverfolgung und Zusammenfassungen der Argumentation, um die Entscheidungslogik aufzuzeigen. Unterstützen Sie Entscheidungswege und die Nachverfolgung von Auswirkungen. Auf diese Weise können Sie die Aktionen Ihrer Agenten wieder mit wichtigen Kennzahlen oder Ergebnissen verknüpfen. Implementieren Sie Mechanismen

zur Erkennung von Abweichungen, die Abweichungen vom erwarteten Verhalten oder von den Richtlinien überwachen.

Führen Sie reflektierende Agenten ein, die das Verhalten und die Systemmuster der Agenten kontinuierlich beobachten. Sie sollten Anomalien oder Inkonsistenzen in Echtzeit melden. Diese Agenten tragen zu Feedbackschleifen bei der Unternehmensführung bei, die eine erneute Validierung, Anpassung oder Außerbetriebnahme von Kapazitäten einleiten können.

Richten Sie Aufsichtsgremien ein, die die Richtlinien der Agenten überprüfen, Änderungen an den Fähigkeiten genehmigen und die Protokolle zur Reaktion auf Vorfälle überwachen. Vertrauen muss verdient, gemessen und kontinuierlich gestärkt werden.

AWS bietet eine solide Grundlage für die Umsetzung dieses Vertrauensrahmens:

- AWS Identity and Access Management (IAM) erzwingt rollenbasierte Ausführungs- und Berechtigungsgrenzen
- Amazon CloudWatch und AWS X-Rayunterstützen vollständige Transparenz und Rückverfolgbarkeit.
- Amazon GuardDuty und AWS Configerkennen Sicherheitsanomalien oder Richtlinienabweichungen.

Zusammen ermöglichen diese Services Identitätsdurchsetzung, Runtime-Sicherheit und vertrauensbasierte Governance in großem Maßstab. Sie können dazu beitragen, autonome Systeme sowohl leistungsstark als auch zuverlässig zu machen.

Vertrauensvoller Autonomie für Unternehmen

Da Agenten immer autonomer werden, wird Vertrauen zu einem entscheidenden Faktor für die Akzeptanz, Steuerung und betriebliche Leistung von Unternehmen. Die Schaffung einer Grundlage für Identität, Beobachtbarkeit und Leitplanken hilft Unternehmen dabei, agentische KI auf sensible Bereiche auszudehnen, ohne dabei auf Governance oder Kontrolle zu verzichten.

Zu den wichtigsten Geschäftstreibern gehören die folgenden:

- Sicherstellung der Unternehmensführung — Starke Identitätsmodelle, Prüfpfade und Genehmigungsgrenzen reduzieren das Compliance-Risiko und unterstützen die Angleichung gesetzlicher Vorschriften.

- Betriebskontinuität — Runtime-Leitplanken und Anomalieerkennung tragen dazu bei, unbeabsichtigtes Verhalten zu verhindern und unterstützen die automatische Wiederherstellung nach Ausfällen im Extremfall.
- Vertrauen der Stakeholder — Entscheidungserklärbarkeit und Telemetrie schaffen Vertrauen bei internen Stakeholdern, Risikomanagern und externen Prüfern.
- Resilienz bei Vorfällen — Integrierte Beobachtbarkeit beschleunigt die Ursachenanalyse und beschleunigt die Reaktionszeit bei Problemen.

Zu den beispielhaften Anwendungsfällen gehören:

- Im Finanzdienstleistungssektor müssen Betrugserkennungsbeamte ihre Argumentation offenlegen, jede Aktion mit nachvollziehbarer Identität protokollieren und im Rahmen eng gestaffelter IAM-Rollen agieren.
- Im Gesundheitswesen müssen autonome Triage-Agenten Sicherheitsüberprüfungen zur Laufzeit durchsetzen, bei Erreichen der Schwellenwerte eine Überprüfung durch einen Mitarbeiter vornehmen und vollständige Protokolle für die klinische Überwachung bereitstellen.

Durch die Integration von Vertrauensmechanismen in den Lebenszyklus der Agenten können Unternehmen dafür sorgen, dass ihre Systeme autonom und verantwortungsbewusst arbeiten. Diese Grundlage reduziert Risiken und ermöglicht es den Mitarbeitern, transparent und integer im Namen des Unternehmens zu handeln.

Letztlich beschleunigt vertrauenswürdige Autonomie die Akzeptanz, da sie sowohl Benutzern als auch Führungskräften die nötige Sicherheit gibt, intelligente Agenten für alle Kernoperationen zu skalieren.

Schwerpunktbereich 5: Verwaltung des Lebenszyklus

Zu erledigende Aufgabe: „Stellen Sie sicher, dass mein Team die Agenten im Laufe der Zeit verbessern kann, ohne Chaos oder Heldentaten.“

Im Gegensatz zu herkömmlichen Anwendungen, die nur durch Code geprägt sind, wird das Verhalten der Agenten auch durch Eingabeaufforderungen, Speicher, Tools und den Schulungskontext bestimmt. Diese Faktoren verändern sich im Laufe der Zeit. Drift beeinträchtigt die Zuverlässigkeit, treibt die Kosten in die Höhe und macht das Debuggen nahezu unmöglich. Ohne Lebenszykluskontrollen bieten Agenten keinen Mehrwert mehr und beginnen, Risiken anzuhäufen.

Strategie

Etablieren Sie DevOps für Agenten (AgentOps) als Praxis. Integrieren Sie CI/CD Pipelines, die auf Agenten zugeschnitten sind. Verwenden Sie diese Pipelines, um Eingabeaufforderungsausgaben zu testen, Tool-Integrationen zu validieren und das Preis-Leistungs-Verhältnis zu profilieren. Pflegen Sie Versionsverläufe von Eingabeaufforderungen, Richtlinien und Modellinteraktionen.

Nutzen Sie Rückkopplungsschleifen aus Observability-Daten, um Umschulungen, schnelle Anpassungen oder die Einstellung von Agenten einzuleiten. Integrieren Sie systemweite Reflexionsmechanismen, wie z. B. ein Verbesserungsregister, um das Lernen zu institutionalisieren.

Erstellen Sie ein leistungsstarkes Telemetrie-Dashboard, das Entscheidungsgenauigkeit, Latenz, Kosten und Zuverlässigkeit anzeigt. Um das Lebenszyklusmanagement mithilfe der AWS Infrastruktur zu rationalisieren und zu beschleunigen, können Teams Agenten-Toolkits verwenden. Ein Beispiel ist das Strands Agents SDK, das strukturierte Tools für die schnelle Versionierung, die Registrierung von Tools und die CI/CD-Integration mit z. B. AWS-Services, und bietet. AWS CodePipelineAWS Cloud Development Kit (AWS CDK)AWS Lambda Verwenden Sie außerdem Amazon S3 und Amazon Elastic File System (Amazon EFS) zum Speichern von Agentenartefakten und Trainingsdaten. Wird verwendet AWS Step Functions, um komplexe Umschulungs- oder Validierungsabläufe zu automatisieren. Sie können Amazon SageMaker AI verwenden, wenn Agenten benutzerdefinierte Modelloptimierungen oder Feinabstimmungen von Workflows benötigen, die über die LLM-Orchestrierung hinausgehen. Die Disziplin im Lebenszyklus verwandelt Agenten von Experimenten in langlebige, sich weiterentwickelnde Ressourcen.

Im Laufe der Zeit bildet dieses Lebenszyklussystem das Rückgrat der Innovation. Es hilft Ihnen, Funktionen agil neu zusammenzustellen, neu zu schulen und neu einzusetzen. Dadurch wird die Agentenebene in ein lebendiges System umgewandelt, das sich als Reaktion auf Feedback und Chancen weiterentwickeln kann.

Der geschäftliche Nutzen des Lebenszyklusmanagements

Ein effektives Lebenszyklusmanagement ist ein wichtiger Faktor für die Leistung und Kosteneffizienz der Agenten. Es stellt sicher, dass intelligente Agenten auch weiterhin genaue, zuverlässige und wertorientierte Ergebnisse liefern, während sie sich weiterentwickeln. Agenten bleiben nicht standardmäßig wertvoll. Sie müssen sich synchron mit den sich ändernden Geschäftsanforderungen, Workflows und Datenumgebungen weiterentwickeln. Ein diszipliniertes AgentOps Team hilft den Mitarbeitern dabei, präzise und effizient zu bleiben und sich im Laufe der Zeit an den Unternehmenszielen auszurichten.

Zu den wichtigsten Geschäftstreibern gehören die folgenden:

- Leistungskonstanz — Kontinuierliche Tests, zeitnahe Validierung und Umschulung helfen den Mitarbeitern dabei, die Entscheidungsqualität unter sich ändernden Bedingungen und Datensätzen aufrechtzuerhalten.
- Kostenoptimierung — Telemetriegestützte Profilerstellung identifiziert ineffiziente Tools, häufig verwendete Eingabeaufforderungen oder unnötige Ausführungen. Anschließend können Sie Anpassungen vornehmen, um die Betriebskosten zu senken.
- Schnellere Iteration — Die Lebenszyklusautomatisierung CI/CD beschleunigt die Entwicklungszyklen und hilft den Teams, Agenten vertrauensvoll zu testen, einzusetzen und zu verbessern.
- Risikominderung — Schnelle Versionierung, Rollback-Support und strukturierte Bewertungsmechanismen tragen dazu bei, Regressionen zu verhindern und ein sicheres und zuverlässiges Änderungsmanagement zu unterstützen.

Zu den beispielhaften Anwendungsfällen gehören die folgenden:

- Ein Mitarbeiter des Kundensupports wird im Hinblick auf Latenz, Modellkosten und Benutzerfeedback überwacht. Die Beobachtung zeigt, dass die Kosten in die Höhe geschnellt sind, was zu einer erneuten Optimierung der eingebetteten Eingabeaufforderungen und der Logik des Fallback-Modells führt.
- Ein Agent für die Vertragszusammenfassung wird auf der Grundlage des Feedbacks der Rechtsteams aktualisiert. Versionierte Eingabeaufforderungen werden vor der Produktionsfreigabe in Sandbox-Umgebungen getestet, um Sicherheit und Qualität zu gewährleisten.

Mit strukturiertem Lebenszyklusmanagement gehen Unternehmen über die reaktive Wartung hinaus und setzen auf proaktive, kontinuierliche Verbesserung. Agenten werden zu adaptiven digitalen Ressourcen, die anhand der Geschäftsziele gemessen, verfeinert und erneut validiert werden. Diese Praxis verwandelt die Agenten-Ökosysteme in leistungsstarke, kostenbewusste und belastbare Systeme, die einen dauerhaften Mehrwert bieten und gleichzeitig mit dem Wandel Schritt halten.

Schwerpunktbereich 6: Abstimmung der Agentenmodelle mit den Geschäftsmodellen

Zu erledigende Job: „Zeigen Sie mir die Auswirkungen, damit ich weitere Investitionen rechtfertigen kann.“

Selbst technisch fähige Agenten werden zu Verbindlichkeiten, wenn sie nicht an Geschäftsergebnisse gebunden sind. Agenten müssen entweder der Effizienz, der Monetarisierung oder der strategischen Differenzierung dienen. Dennoch fällt es den meisten Unternehmen schwer, zu definieren, wie Agenten in Preis-, Verpackungs- oder Nutzungsmodelle passen. Ohne eine klare Ausrichtung auf den Geschäftswert ist es schwierig, eine Skalierung oder gar Beibehaltung der Investition zu rechtfertigen.

Strategie

Wenden Sie Produktmanagement-Praktiken an. Behandeln Sie Agenten als monetarisierbare Dienste mit einem messbaren ROI. Definieren Sie Preisstrategien auf der Grundlage von Entscheidungen, Sitzungen oder Ergebnissen. Anschließend bündeln Sie die Funktionen der Agenten in abgestufte Angebote, die auf Kundensegmente oder interne Geschäftsbereiche zugeschnitten sind.

Um Nachhaltigkeit zu fördern, müssen Unternehmen durch den Einsatz von Agenten sowohl direkten Mehrwert als auch Wachstumsmultiplikatoren nutzen. Erwägen Sie, die folgenden ROI-Metriken zu verwenden, um den unmittelbaren Nutzen zu messen:

- Kosten pro Entscheidung — Vergleichen Sie die Bearbeitungskosten der Mitarbeiter mit den entsprechenden Kosten für Mitarbeiter.
- Zeitkomprimierung — Quantifizieren Sie den Wert beschleunigter Zyklen, z. B. schnellerer Verkäufe oder Genehmigungen.
- Fehlerreduzierung — Messen Sie die Einsparungen, die sich aus verbesserter Genauigkeit, Konsistenz und Konformität ergeben.

Neben diesen unmittelbaren Vorteilen können Agenten die folgenden langfristigen Wachstumschancen nutzen:

- Capability Stacking — Kombinieren Sie Agentendienste, um domänenspezifische vertikale Lösungen zu entwickeln.
- Netzwerkeffekte — Steigern Sie den Wert durch Ökosysteme mit mehreren Agenten, in denen Koordination den Nutzen verstärkt.

- Markterweiterung — Generierung neuer Einnahmequellen durch extern nutzbare, agentengestützte Dienste.

Erstellen Sie Feedbackschleifen anhand von Geschäftskennzahlen (wie Kosteneinsparungen, Steigerung der Konversionsrate oder time-to-resolution), um die kontinuierliche Weiterentwicklung der Mitarbeiter voranzutreiben. Analysieren Sie Nutzungstelemetrie und Nutzerzufriedenheitswerte, um Ihre Werteausrichtung und Ihre Roadmap-Prioritäten zu verfeinern. Durch die direkte Verknüpfung von Agentenfunktionen mit Geschäftsmodellen positionieren sich Unternehmen so, dass sie nicht nur technische Ergebnisse erzielen, sondern auch einen nachhaltigen Mehrwert erzielen können.

Die folgenden Maßnahmen AWS-Services unterstützen diese Ausrichtung, indem sie robuste Rahmenbedingungen für Nachverfolgung und Monetarisierung bereitstellen:

- AWS Cost Explorerund Amazon CloudWatch bieten Einblicke in die Kosten pro Mitarbeiter und die betriebliche Effizienz.
- Amazon API Gateway ermöglicht kostenpflichtigen Zugriff, Ratenbegrenzung und gestaffelte Preisgestaltung für Agenten-Endgeräte.
- AWS Marketplacebietet einen Kanal für die Veröffentlichung von Agenten und Agenturlösungen als kommerzielle Produkte.

Diese Services helfen Ihnen dabei, Agentenfunktionen in skalierbare, wertorientierte digitale Angebote umzuwandeln, die auf Unternehmenswachstums- und Monetarisierungsstrategien abgestimmt sind.

Weiterentwicklung der Softwarebereitstellung für agentische KI

Die moderne Softwarebereitstellung basiert auf einer einfachen Annahme: Sie haben die Kontrolle über die Systeme, die Sie ausliefern. Sie definieren Anforderungen, schreiben Logik, testen anhand erwarteter Ergebnisse und stellen vorhersehbare Dienste bereit. Selbst agile DevOps Methoden und Ansätze basieren immer noch auf dem Prinzip, dass jeder Sprint etwas Deterministisches, Überprüfbares liefert und größtenteils unter menschlicher Aufsicht liegt.

Agentische KI stellt dieses Fundament auf den Kopf. Agentische Systeme interpretieren, begründen und passen sich an, anstatt Skripten zu folgen. Ihr Verhalten hängt vom Code ab, den Sie schreiben, vom Kontext, in dem sie arbeiten, von den Eingaben, die sie erhalten, von den Tools, auf die sie zugreifen können, und von den Zielen, die ihnen zugewiesen wurden. Kurz gesagt, sie befolgen keine Befehle, sie verfolgen Ergebnisse.

Dadurch geht es bei der Umsetzung weniger um Kontrolle als vielmehr um Abstimmung. Anstatt Anweisungen zu geben, müssen Sie festlegen, wie es sich verhält. Das bedeutet, dass der herkömmliche Software Development Lifecycle (SDLC) nicht mehr passt, weil er für logikbasierte, von Menschen gesteuerte Systeme konzipiert wurde.

In diesem Abschnitt werden folgende Themen behandelt:

- Zielzonen für agentische KI
- Weiterentwicklung des Bereitstellungslebenszyklus für agentische KI
- Teams auf agentische KI vorbereiten

Zielzonen für agentische KI

Statt starrer Phasen wie Definition, Aufbau, Test und Freigabe benötigen wir ein Modell, das Autonomie, Unsicherheit und Entfaltung berücksichtigt. Stattdessen verwenden Sie Zonen of Intent. Eine Absichtszone definiert einen begrenzten Raum, in dem ein Agent innerhalb von Einschränkungen autonom agieren kann. Das Ziel besteht darin, vom Mikromanagement jeder Aufgabe zur Gestaltung von Umgebungen überzugehen, in denen Agenten sicher handeln, lernen und zusammenarbeiten können. Sie geben das Was (das gewünschte Ergebnis), das Warum (die Absicht) und die Leitplanken (die Einschränkungen, Richtlinien und Vertrauensgrenzen) an. Anhand dieser Grenzen und dieser Informationen findet der Agent heraus, wie.

Stellen Sie sich die Umwelt nicht als Fließband vor, sondern als Luftraum. Sie kontrollieren, wer Zutritt hat, was sie tun können und wohin sie gehen können. Sobald sie jedoch drinnen sind, können sie sich nach Bedarf frei bewegen. So skalieren Agentensysteme ohne Chaos.

Das ist nicht nur ein philosophischer, sondern auch ein praktischer Wandel. Das nichtdeterministische Ergebnis agentenbasierter Systeme kann nicht vollständig durch Komponententests getestet werden. Es kann nicht wie statische Binärdateien versioniert werden. Agenten ändern sich im Laufe der Zeit, passen sich neuen Daten an und interagieren auf unvorhersehbare Weise mit anderen Systemen. Der Versuch, sie mithilfe herkömmlicher Modelle bereitzustellen, führt zu fragilen, nicht skalierbaren Architekturen. Im schlimmsten Fall führt dies zu falschem Vertrauen in Systeme, die Sie nicht wirklich steuern können.

Wenn Teams auf absichtsbasierte Bereitstellung setzen, haben sie zwei Vorteile:

- Kontrollieren Sie, wo es am wichtigsten ist — Sie definieren Grenzen statt Outputs.
- Skalierbarkeit durch Delegierung — Sie ermöglichen es Agenten, mit Komplexität umzugehen, die Menschen nicht fest programmieren können.

Auf diese Weise gelangen Sie von isolierten Prototypen zu echten Agentensystemen in Produktionsqualität, die wiederholt und zuverlässig Mehrwert bieten können.

Weiterentwicklung des Bereitstellungslebenszyklus für agentische KI

Um intelligentes, adaptives Verhalten zu unterstützen, muss das SDLC von deterministischer Steuerung auf adaptive Absicht umgestellt werden. Im Folgenden sind die Änderungen aufgeführt, die erforderlich sind, um das traditionelle SDLC für agentische KI weiterzuentwickeln:

- Aus Planung wird Intent Design. Teams definieren Ziele, Einschränkungen und erwartetes Verhalten der Agenten. Richtlinien und Erfolgskriterien orientieren sich an der Abstimmung, nicht an der Logik.
- Architektur wird zu Gerüsten. Teams konzentrieren sich darauf, Rollen, Schnittstellen, Leitplanken, Ausweichmechanismen und Beobachtbarkeit zu definieren, anstatt für jeden Entscheidungspfad ein Skript zu erstellen.
- Aus Testen wird Verhaltensevaluierung. Anstatt bestimmte Ergebnisse zu bestätigen, überprüfen die Teams, ob die Agenten die akzeptablen Grenzen einhalten und ihre Absichten unter verschiedenen Inputs erfüllen.

- Aus der Bereitstellung wird eine kontinuierliche Orchestrierung. Agentensysteme werden mit Laufzeitsteuerungen, Live-Überwachung und Feedback-Kanälen bereitgestellt, die eine Optimierung in Echtzeit ermöglichen.
- Aus Iteration wird Feedback und Anpassung. Statt herkömmlicher Patch-Zyklen zur Codeänderung beobachten Teams, wie sich Agenten weiterentwickeln, wo sie erfolgreich sind oder wann sie abdriften. Bei Bedarf greifen die Teams ein, indem sie die Einschränkungen aktualisieren, sie umschulen und Kontrollmechanismen hinzufügen oder ändern.

Bestehende Praktiken, die sich auf Iteration, Experimente und schnelles Feedback konzentrieren, haben die Hälfte erreicht. Die Umstellung auf agentische Systeme ist keine Ablehnung agiler Prinzipien. Tatsächlich ist es eine natürliche Weiterentwicklung von ihnen. Agiles Denken betont Anpassungsfähigkeit, Feedback und funktionierende Lösungen gegenüber starren Plänen. Das passt perfekt zur Natur agentischer Systeme, die in Echtzeit lernen, sich anpassen und auf den Kontext reagieren. Wenn Sie bereits kurze Zyklen haben, Annahmen schnell validieren und Unsicherheiten durch kontinuierliche Bereitstellung bewältigen, sind Sie gut gerüstet, um diesen Übergang zu leiten.

Es gibt jedoch wichtige Unterschiede. Der traditionelle agile Ansatz geht davon aus, dass das, was geliefert wird, deterministisch ist. Es wird davon ausgegangen, dass sich das Ding nach der Erstellung konsistent und vorhersehbar verhält und wiederholbare Ergebnisse für dieselben Eingaben liefert. Diese Wiederholbarkeit hilft Ihnen beim Debuggen, Testen und Iterieren mit Zuversicht. Agentensysteme machen dieses Modell kaputt. Sie sind probabilistisch, kontextsensitiv und können sich unabhängig entwickeln. Das bedeutet, dass einige Agile-Methoden weniger nützlich sind, wie z. B. die Geschwindigkeitsverfolgung auf der Grundlage der Fertigstellung von Storys, strenge Akzeptanzkriterien oder deterministische Sprint-Planung.

Die folgenden Aspekte des traditionellen SDLC gelten für agentische KI:

- Iterative Entwicklung und Bereitstellung
- Kundenfeedback als primäres Signal
- Funktionsübergreifende Zusammenarbeit
- Kontinuierliche Integration und Bereitstellung

Die folgenden Aspekte des traditionellen SDLC müssen sich für agentische KI weiterentwickeln:

- Definieren Sie „Erledigt“ im Einklang mit der Absicht neu. Konzentrieren Sie sich darauf, ob das Verhalten des Agenten das angestrebte Ziel innerhalb der definierten Einschränkungen erfüllt.

- Wechseln Sie von Akzeptanzkriterien zu Verhaltensleitplanken.
- Erweitern Sie die Definition von erledigt um die Runtime Readiness. Dazu gehören Beobachtbarkeit, Erklärbarkeit und Feedback-Mechanismen, die kontinuierliches Lernen und Vertrauen fördern.
- Priorisieren Sie Feedback-Schleifen und Verhaltensverfolgung in Echtzeit gegenüber der Vorausplanung

Die gute Nachricht ist, dass Sie das SDLC-Playbook nicht wegwerfen müssen. Sie müssen es nur von der Verwaltung des Codes zur Verhaltensgestaltung weiterentwickeln. In Agentensystemen hängt Erfolg nicht nur davon ab, ob Software ausgeführt wird, sondern auch davon, wie sie sich verhält.

Teams auf agentische KI vorbereiten

Die Softwareentwicklung wird nicht verschwinden. Es entwickelt sich. Die Arbeit verlagert sich vom Schreiben von Funktionen hin zur Gestaltung von Rahmenbedingungen und Kontrollmechanismen für intelligentes Verhalten. In der Welt der agentischen KI ist das Bauen nicht mehr der schwierige Teil — der Umgang mit neuen Entwicklungen schon. Für die meisten Entwicklungsteams fühlt sich die Entwicklung eher wie eine Änderung der Denkweise als wie ein technischer Sprung an. Anstatt zu fragen: „Was wird das System tun?“ Die Frage lautet: „Wozu haben wir es ermächtigt, und wie werden wir wissen, ob es auf Kurs bleibt?“

Für Entwicklungsteams erfordert die Entwicklung hin zu Agenten-KI die folgenden Änderungen:

- Ein kultureller Wandel — Teams müssen sich mit der Unsicherheit und Autonomie in Systemen, die sie nicht vollständig kontrollieren, vertraut machen.
- Neue Rollen — Intent-Designer, Verhaltenstester und Observability-Techniker werden zu einem zentralen Bestandteil der Umsetzung.
- Gemeinsame Sprache — Teams benötigen ein klares, gemeinsames Verständnis von Zielen, Leitplanken und Erfolgssignalen, genau wie früher Spezifikationen und Testfälle.

Mit zunehmender Reife der generativen KI werden wir mehr Agentensysteme sehen, die mit Kunden, Produkten und Abläufen interagieren. Die Organisationen, die erfolgreich sein werden, werden nicht die mit den besten Modellen sein. Es werden diejenigen sein, die Agenten sicher, kontrolliert und schnell in reale Arbeitsabläufe integrieren können. Das bedeutet, dass sich die Bereitstellungsmodelle und die Entwicklungsteams gemeinsam weiterentwickeln müssen. Zonen of

Intent geben Ihnen die nötige Abstraktion dafür. Sie helfen Ihnen dabei, Autonomie zu verwirklichen, ohne die Rechenschaftspflicht aufzugeben. Sie bieten auch ein gemeinsames Framework für alle Teams, um Systeme zu verwalten, die nicht fest codiert werden können.

Weitere Informationen zur Vorbereitung von Teams auf Agenten-KI finden Sie im Abschnitt Vorbereitung des Unternehmens auf agentische KI in großem Maßstab dieses Leitfadens.

Vorbereitung des Unternehmens auf agentische KI in großem Maßstab

Im Zuge der Konvergenz der in diesem Leitfaden beschriebenen Schwerpunktbereiche verlagert sich die agentische KI von isolierten Funktionen hin zu einer einheitlichen Informationsebene, die als Fähigkeitsplattform verstanden werden kann. Diese Plattform führt nicht nur Aufgaben aus. Sie entwickelt sich, passt sich an und koordiniert domänenübergreifend. Agenten werden zu modularen, wiederverwendbaren und auffindbaren Diensten, die Innovationen beschleunigen, die kognitive Belastung reduzieren und zu messbaren Ergebnissen im gesamten Unternehmen führen. Diese Plattformansicht schafft die Voraussetzungen für skalierbare Intelligenz, die in das gesamte Betriebsmodell integriert ist.

Die Operationalisierung der Agenten-KI erfordert mehr als den Einsatz intelligenter Agenten. Dies erfordert eine grundlegende Veränderung der Art und Weise, wie Unternehmen Teams organisieren, Prozesse entwerfen und Technologie steuern. Genau wie die Umstellung auf die Cloud oder DevOps neu definierte Betriebsmodelle läutet die agentische KI eine neue Ära der Entscheidungsautomatisierung, des kontinuierlichen Lernens und der autonomen Koordination ein. Der Erfolg hängt davon ab, ob die Systeme, Mitarbeiter und Prozesse auf diese neue Betriebsphilosophie abgestimmt sind.

In diesem Abschnitt werden folgende Themen behandelt:

- Abstimmung von Teams und Eigentümermodellen
- Bewältigung von Veränderungen und organisatorische Bereitschaft
- Architektur für Interoperabilität und Zusammenarbeit
- Integration von Governance in ein Agentengefüge
- Einführung einer betriebswirtschaftlichen Denkweise, bei der die Entscheidung an erster Stelle steht
- Zielgerichtet und zielgerichtet skalieren

Abstimmung von Teams und Eigentümermodellen

Der erste Schritt zur Reife ist die funktionsübergreifende Ausrichtung. Unternehmen müssen AgentOps Teams zusammenstellen, denen AI/ML Praktiker und Fachspezialisten wie Architekten für verteilte Systeme, Softwareingenieure, Produktbesitzer, Compliance-Verantwortliche und

Plattformanten angehören. Diese Teams sind gemeinsam für den gesamten Lebenszyklus eines Agenten verantwortlich, von der Planung und Bereitstellung bis hin zur Umschulung und Überwachung.

Bei der Bereitstellung und Veröffentlichung von Agenten sollten cloudnative Verfahren angewendet werden, wie z. B. die Verwendung von AWS Cloud Development Kit (AWS CDK)und AWS CodePipelinefür die Infrastruktur als Code und die automatisierte Bereitstellung. Diese Struktur fördert die gemeinsame Rechenschaftspflicht und beschleunigt die Iteration. So wie sie Entwicklung und Betrieb DevOps vereinheitlicht, AgentOps verbindet sie Intelligenz mit Steuerung und Ausführung.

Um effektiv zu sein, benötigen diese Teams auch eine gemeinsame Sprache. Geschäftsbeteiligte müssen verstehen, was Agenten sind, wie sie arbeiten und welche Ergebnisse sie erzielen. Schulung und interne Unterstützung sind unerlässlich. Durch die Entmystifizierung der Akteure und die Einbettung dieses mentalen Modells in alltägliche Gespräche ermöglichen Unternehmen eine breitere Teilhabe und zielgerichtetere Innovationen.

Um die Entwicklung und Integration von Agenten mithilfe von Agenten zu beschleunigen AWS-Services, können Teams Frameworks wie das Strands Agents SDK einsetzen, das CLI-basierte Tools für das Gerüsten, Konfigurieren und Paketieren von Agenten bietet. Strands Agents ist so konzipiert, dass es nahtlos mit AWS Infrastrukturen wie Amazon Bedrock, Amazon AWS Lambda EventBridge AWS CDK, The und AWS CodePipeline zusammenarbeitet. Es ermöglicht eine schnelle Prototypenerstellung und Implementierung unter Beibehaltung produktionsgerechter Standards.

Struktur und Werkzeuge allein reichen jedoch nicht aus. Die Skalierung agentischer KI erfordert eine bewusste Bereitschaft zu Kultur, Bildung und Führung, um sicherzustellen, dass die Akzeptanz im gesamten Unternehmen Fuß fasst.

Bewältigung von Veränderungen und organisatorische Bereitschaft

Die erfolgreiche Skalierung der KI von Agenturen erfordert mehr als den Einsatz von Infrastruktur oder intelligenten Agenten. Es erfordert einen strukturierten Ansatz für organisatorische Veränderungen. Dazu gehören kulturelle Bereitschaft, die Entwicklung von Fähigkeiten, auf Kennzahlen basierende Feedback-Schleifen und die Ausrichtung der Führungskräfte, um sicherzustellen, dass die Einführung sowohl gewollt als auch nachhaltig erfolgt.

Fördern Sie die kulturelle Entwicklung

- Stellen Sie Agenten als Teamkollegen und nicht als Ersatz ein, um Widerstände abzubauen und Vertrauen aufzubauen.
- Kommunizieren Sie transparent über die Fähigkeiten und Grenzen der Agenten, um realistische Erwartungen zu setzen.
- Richten Sie klare Übergabeprotokolle ein, die festlegen, wann Agenten Entscheidungen an eine höhere Behörde weiterleiten oder Teile des Prozesses an einen menschlichen Mitarbeiter delegieren sollten.

Richten Sie einen Rahmen für die Entwicklung von Fähigkeiten ein

- Bieten Sie rollenbasierte Schulungen an, die auf Ingenieure, Produktmanager, Bereichsleiter und Compliance-Beauftragte zugeschnitten sind.
- Richten Sie Kompetenzzentren ein, um bewährte Verfahren, Werkzeugmuster und wiederverwendbare Ressourcen auszutauschen.
- Kombinieren Sie KI-Spezialisten mit Fachexperten im Rahmen von Mentorenprogrammen, um Wissenslücken zu schließen.

Definieren Sie Metriken und Feedback-Schleifen

- Verknüpfen Sie technische und geschäftliche KPIs Aspekte mit strategischen Werten, um die Auswirkungen zu bewerten. Zu den Vorteilen zählen beispielsweise die Latenz bei der Entscheidungsfindung, die Genauigkeit der Problemlösung und Kosteneinsparungen.
- Erfassen Sie systematisch und kontinuierlich das Feedback der Benutzer, um Reibungspunkte und Herausforderungen bei der Einführung zu ermitteln.
- Führen Sie regelmäßige Retrospektiven durch, um die Leistung der Agenten, Nutzungstrends und Verbesserungsmöglichkeiten zu bewerten.

Stimmen Sie die Führung von oben ab ab

- Gewinnen Sie Unterstützung durch Führungskräfte, indem Sie die Initiativen Ihrer Agenten mit strategischen Ergebnissen und dem ROI verknüpfen.
- Bilden Sie funktionsübergreifende Verwaltungsausschüsse, denen sowohl technische als auch geschäftliche Führungskräfte angehören.

- Passen Sie Kommunikationsstrategien an, um Klarheit und Engagement auf allen Organisationsebenen zu gewährleisten.

Dieser systematische Ansatz für das Change-Management stellt sicher, dass die Implementierung der Technologie mit der organisatorischen Reife einhergeht. Es schafft eine Grundlage für Vertrauen, Akzeptanz und langfristigen Geschäftswert.

Architektur für Interoperabilität und Zusammenarbeit

Isolierte Agentenbereitstellungen sorgen für Vorteile auf lokaler Ebene. Der Unternehmenswert wird jedoch erst dann sichtbar, wenn Agenten sich gegenseitig erkennen, sich gegenseitig aufrufen und mit ihnen zusammenarbeiten können. Das bedeutet, dass Standards für die Registrierung, Authentifizierung und den Austausch von Fähigkeiten von Agenten definiert werden müssen. Architektonisch spiegelt dies den Wandel von Monolithen hin zu Microservices wider, bei denen es sich um zusammensetzbare, wiederverwendbare und lose gekoppelte Einheiten handelt, die komplexe Probleme gemeinsam lösen.

Neue Protokolle wie A2A und MCP sind grundlegend. Sie ermöglichen die semantische Interoperabilität zwischen Agenten, Tools und Speichersystemen. A2A unterstützt die Interaktion auf Peer-Ebene, sodass Agenten die Verantwortung für Aufgaben aushandeln, den Kontext teilen und Arbeitsabläufe koordinieren können. MCP ergänzt dies durch gemeinsame Schemas für den Austausch von Kontextdaten zwischen Agenten und ihren Umgebungen. Es standardisiert, wie Funktionen aufgerufen werden, wie auf sie zugegriffen wird und wie Zustände verwaltet APIs werden. Zusammen sorgen diese Protokolle für Erweiterbarkeit, Konsistenz und langfristige Wartbarkeit im gesamten Agenten-Ökosystem.

Die Steuerung bleibt von entscheidender Bedeutung. Kontrollebenen, wie z. B. Arbiter-Agenten, ermöglichen eine richtlinienorientierte Delegierung, ohne dass es zu zentralen Engpässen kommt. Diese Agenten agieren als Vertrauensvermittler. Sie setzen Grenzen durch und ermöglichen es anderen Agenten gleichzeitig, sich selbst zu organisieren. Die Zusammenarbeit mit Agenten hilft Unternehmen dabei, ihre agentischen KI-Ökosysteme sowohl agil als auch vertrauensvoll zu skalieren.

Integration von Governance in ein Agentengefüge

Mit größerer Autonomie geht ein höheres Risiko einher. Die Steuerung muss vom ersten Tag an in die Agentenarchitektur integriert werden. Dazu gehören die Festlegung von politischen Grenzen,

die festlegen, was Agenten tun dürfen, die Durchsetzung von Identitätsmodellen, die festlegen, für wen sie handeln, und die Implementierung von Erklärbarkeit und Rückverfolgbarkeit. Observability-Systeme müssen Telemetriedaten zum Verhalten der Agenten mithilfe von Diensten wie Amazon CloudWatch und erfassen AWS X-Ray, die eine zentrale Protokollierung und verteilte Nachverfolgung über die Workflows der Agenten hinweg ermöglichen. Reflective Agents können die Leistung auf der Grundlage dieser Telemetrie-Feeds kontinuierlich prüfen und bewerten.

Auch die Unternehmensführung muss sich im Zuge der Weiterentwicklung des Agenten-Ökosystems weiterentwickeln. Da die Akteure immer fähiger und autonomer werden, müssen die Aufsichtsmechanismen anpassungsfähiger werden. Richtlinienaktualisierungen, Capability Gating und Verhaltenseinschränkungen zur Laufzeit müssen dynamisch und in großem Umfang durchsetzbar sein. Vertrauen ist keine Zusatzfunktion. Es wird kontinuierlich durch Architektur, Verhalten und Prozesse gestärkt. AWS Identity and Access Management (IAM) und AWS AppConfigspielen eine entscheidende Rolle bei der Durchsetzung sicherer Identitäten, Laufzeitberechtigungsgrenzen und umgebungsspezifischer Verhaltensänderungen zwischen den Agenten.

Einführung einer betriebswirtschaftlichen Denkweise, bei der die Entscheidung an erster Stelle steht

Herkömmliche Automatisierung konzentriert sich auf die Prozesseffizienz, d. h. die schnellere und zuverlässigere Ausführung vordefinierter Skripts oder Workflows. Agentic AI hingegen führt eine Automatisierung ein, bei der die Entscheidung an erster Stelle steht. Agenten bewerten den Kontext, wägen Optionen ab und passen das Verhalten in Echtzeit an. Dieser Übergang von einer Denkweise, bei der die Ausführung an erster Stelle steht, erfordert ein neues Denken über Erfolgskennzahlen und Ergebnisse. Anstatt den Erfolg ausschließlich an der Erledigung von Aufgaben zu messen, wird der Erfolg agentischer KI daran gemessen, wie gut die Entscheidung mit der Absicht, den Richtlinien und den sich entwickelnden Bedingungen abgestimmt ist.

Anstatt nur die Erledigung von Aufgaben oder die Zykluszeit zu messen, müssen Unternehmen die Qualität der Entscheidungen und die Reaktionsfähigkeit auf Veränderungen bewerten. time-to-action KPIs sollte Kennzahlen wie die folgenden beinhalten:

- Entscheidungsqualität — Wie gut hat der Agent seine Antwort auf den jeweiligen Benutzer oder das jeweilige Szenario personalisiert? Hat er nuancierte Entscheidungen getroffen, die auf die Geschäftsziele und den Benutzerkontext abgestimmt sind?

- T ime-to-action — Wie schnell und intelligent hat der Mitarbeiter eine Situation eingeschätzt und darauf reagiert? War die Latenz niedrig genug, um sich anpassungsfähig und menschenähnlich zu fühlen?
- Kognitive Entlastung — Wie viele manuelle Analysen, Triage oder routinemäßige Entscheidungen konnte der Mitarbeiter im Namen eines Menschen bewältigen? Hat es den Aufwand reduziert oder ihn einfach verschoben?

Unternehmen, die die Entscheidung an erste Stelle setzen, können widerstandsfähiger und anpassungsfähiger werden und in der Lage sein, auf einer neuen Ebene der Komplexität zu agieren.

Zielgerichtet und zielgerichtet skalieren

Bei der erfolgreichen Skalierung agentischer KI geht es nicht darum, mit mehr Tools zu experimentieren. Es geht darum, eine dauerhafte Ebene von Unternehmensinformationen aufzubauen. Dies erfordert Investitionen in die Plattforminfrastruktur, die Betriebskultur, die Rahmenbedingungen für die Unternehmensführung und die strategische Ausrichtung. Unternehmen müssen einen bewussten Ansatz verfolgen. Sie müssen Agenten nicht als Experimente, sondern als Kernkomponenten ihres digitalen Betriebsmodells behandeln.

Durch die Ausrichtung auf das AWS Well-Architected Framework können Ihre Systeme die Unternehmensstandards für Zuverlässigkeit, Sicherheit, Leistungseffizienz und Kostenoptimierung erfüllen. Tools wie das Strands Agents SDK können diesen Prozess beschleunigen, indem sie strukturierte Eingabeaufforderungen, die Registrierung von Tools und die CI/CD-Bereitschaft bereitstellen. Dies hilft Teams, mithilfe vertrauter Workflows vom Experimentieren zur skalierbaren Bereitstellung überzugehen. AWS

Agentic AI ist kein Tool, sondern eine Veränderung der Art und Weise, wie Intelligenz in Abläufe integriert wird. Organizations, die sich entsprechend vorbereiten, können mehr automatisieren, intelligenter arbeiten, sich schneller anpassen und sich in einer zunehmend komplexen Welt dauerhafte Vorteile verschaffen.

Fazit zur Operationalisierung agentischer KI

Agentic AI ist mehr als ein technologischer Wandel. Es markiert die Entstehung eines neuen Betriebssystems für Unternehmen. Organizations, die diese Transformation begrüßen, gehen über enge Anwendungsfälle der Automatisierung hinaus und integrieren Intelligenz in die Grundlage ihrer Betriebsabläufe. Bei diesem Wandel geht es darum, die Art und Weise, wie Entscheidungen getroffen werden, wie sich Systeme anpassen und wie Ergebnisse in großem Maßstab erzielt werden, neu zu gestalten.

In einer Zeit, die von wachsender Komplexität, Echtzeitnachfrage und Informationsflut geprägt ist, ist das traditionelle Modell der skriptgesteuerten Automatisierung an seine Grenzen gestoßen. Der Erfolg hängt heute von der Fähigkeit ab, Intelligenz direkt in Arbeitsabläufe zu integrieren, um Systeme zu entwickeln, die wahrnehmen, argumentieren, handeln und sich weiterentwickeln. Agentische KI kann Autonomie mit Zielsetzung, Entscheidungsfindung mit Steuerung und Anpassungsfähigkeit mit Rechenschaftspflicht in Einklang bringen.

Dieser Übergang erfordert einen Übergang vom Denken, bei dem die Ausführung an erster Stelle steht, hin zu einem Denken, das zuerst Entscheidungen trifft. Agentensysteme folgen nicht einfach Anweisungen. Sie interpretieren Ziele, wägen Kompromisse ab und verfolgen Ergebnisse innerhalb definierter Grenzen. In diesem Zusammenhang wird Erfolg nicht nur an der Erledigung von Aufgaben gemessen. Er wird auch an der Qualität, Agilität und Erklärbarkeit von Entscheidungen gemessen, die in Echtzeit getroffen werden. Organizations müssen Metriken, Anreize und Systemdesign überdenken, um Agenten zu unterstützen, die unter Ungewissheiten intelligent agieren.

Die Operationalisierung der Agenten-KI ist kein Upgrade. plug-and-play Es ist eine architektonische und kulturelle Transformation. Es erfordert disziplinierte Verfahren in den Bereichen Lebenszyklusmanagement, Durchsetzung von Vertrauen, Interoperabilität und Anpassung an Geschäftsmodelle. Es erfordert auch die Weiterentwicklung von Bereitstellungsmodellen, wie etwa die Festlegung von Absichtszonen, die Einbettung von Laufzeitplanken und die kontinuierliche Abstimmung des Verhaltens der Agenten an strategischen Ergebnissen. Die Teams müssen eine gemeinsame Sprache, gemeinsame Verantwortung und gemeinsame Rechenschaftspflicht für die Leistung und Sicherheit der Agenten einführen.

Die Bereitschaft der Unternehmen kann darüber entscheiden, wer in dieser neuen Umgebung erfolgreich ist. Organizations müssen in interne Unterstützung, AgentOps Fähigkeiten und Governance-Rahmenbedingungen investieren, die skalieren und langfristigen Wert schaffen. Diejenigen, die erfolgreich sind, können intelligentere Systeme aufbauen, und sie können auch anpassungsfähigere, widerstandsfähigere und erkenntnisorientiertere Unternehmen aufbauen.

Dieser Leitfaden legt den Grundstein. Es verbindet Strategie mit Umsetzung und bereitet Unternehmen darauf vor, skalierbare Plattformen mit intelligenten Agenten aufzubauen. Die umfassendere Inhaltsreihe über agentic AI on AWS bietet ergänzende Anleitungen. Die anderen Leitfäden dieser Reihe finden Sie unter Agentic AI auf der Prescriptive Guidance-Website. AWS Diese Inhaltsreihe bietet eine Roadmap zur disziplinierten und zielgerichteten Operationalisierung von Autonomie.

Identifizieren Sie zunächst einen Entscheidungsbereich mit großer Wirkung, in dem Agenten messbare Verbesserungen in Bezug auf Geschwindigkeit, Genauigkeit oder Reaktionsfähigkeit erzielen können. Setzen Sie dann einen Testagenten ein, der gezielt eingesetzt werden kann und über Instrumentierung, Steuerung und Feedbackschleifen verfügt. Verwenden Sie dies, um die Werthypothese zu validieren, interne Dynamik zu erzeugen und Vertrauen in den Ansatz aufzubauen. Die Dynamik verstärkt sich durch Lernen.

Agentic AI ist kein Ziel, sondern eine Fähigkeitsebene, die sich mit Ihrem Unternehmen weiterentwickelt. Sie steht für einen langfristigen Wandel hin zu Intelligenz als Infrastruktur. Organizations, die in diesem Bereich führend sind, können mehr automatisieren, schneller reagieren, sich besser anpassen und Betriebsmodelle entwickeln, die in der Lage sind, die Komplexität auf Unternehmensebene zu bewältigen.

Ressourcen für die Operationalisierung agentischer KI

AWS-Services

Die folgenden Funktionen können Ihnen beim Aufbau AWS-Services und der Operationalisierung agentischer KI-Systeme in folgenden Bereichen helfen: AWS Cloud

- Amazon API Gateway kann die Funktionen der Agenten skalierbar machen und bietet nutzungsabhängige Preise.
- AWS AppConfigbietet Laufzeitkonfigurationsmanagement und Funktionsumschaltung für Agenten in verschiedenen Mandanten oder Umgebungen.
- Amazon Bedrock ist ein Service mit Basismodell, den Agenten zur Argumentation, Generierung und sofortigen Ausführung verwenden können.
- AWS Cloud Development Kit (AWS CDK)ist ein Infrastructure-as-Code-Service, mit dem Sie Agenten-Stacks bereitstellen und verwalten können.
- AWS CloudTrailzeichnet den Ereignisverlauf auf, sodass Sie Agentenaktivitäten, Prüfpfade und Integrationsverhalten verfolgen können.
- Amazon CloudWatch kann Protokolle, Metriken und Alarme verwalten, um die Leistung der Agenten und das Kooperationsverhalten mehrerer Agenten zu überwachen.
- AWS CodePipelinebietet CI/CD Automatisierung, mit der Sie Agentencode testen, validieren und bereitstellen können.
- Amazon Cognito ist ein Identitätsdienst, mit dem Sie die Benutzer- und Mandantenauthentifizierung in Systemen mit mehreren Agenten verwalten können.
- AWS Configbietet Konformität und Erkennung von Abweichungen in Bezug auf Agentenrichtlinien und Umgebungskonfiguration.
- AWS Cost Explorerkann die Nutzung auf Agentenebene verfolgen und dabei helfen, die Kosten so aufeinander abzustimmen, dass Ihr ROI maximiert wird.
- Amazon DynamoDB ist ein Speicherservice, den Sie für Agentenspeicher, Verbesserungsprotokolle und den Kontextstatus verwenden können.
- Amazon Elastic File System (Amazon EFS) ist ein gemeinsam genutztes Dateisystem, das Sie für die Zusammenarbeit mit Agenten oder die zwischengeschaltete Verarbeitung zwischen Workflows verwenden können.

- Amazon EventBridge ist ein zentraler Event-Bus, mit dem Sie Aufgaben weiterleiten und die Kommunikation in der Agent-Fabric orchestrieren können.
- Amazon EventBridge Pipes kann die Erfassung und Weiterleitung von Ereignissen für die Verbindung von Agenten und Services optimieren.
- Amazon GuardDuty bietet Bedrohungserkennung und Anomalieüberwachung, die die sichere Ausführung von Agenten unterstützen können.
- AWS Identity and Access Management (IAM) unterstützt Sie bei der Definition detaillierter Berechtigungen für die Ausführung von Agenten und den Datenzugriff.
- AWS Lambdaist ein zustandsloser Rechendienst, der Agentenlogik ausführen und Drohnen anlocken kann.
- AWS Marketplaceist eine externe Vertriebsplattform, mit der Sie Agentenfunktionen als kommerzielle Produkte anbieten können.
- AWS Organizationsist ein kontenübergreifender Service zur Verwaltung und Durchsetzung von Richtlinien, der Sie bei der Verwaltung der Agenteninfrastruktur für mehrere Mandanten unterstützen kann.
- AWS Organizations Richtlinien zur Dienstkontrolle dienen als Leitplanken für die Kontrolle von Berechtigungen auf Konto- oder Organisationseinheitsebene.
- Amazon Quick ist eine generative KI-gestützte Business Intelligence (BI) -Plattform, mit der Sie Daten analysieren, Visualisierungen erstellen, Workflows automatisieren und mit anderen in Ihrem Unternehmen zusammenarbeiten können.
- AWS Resource Access Manager (AWS RAM) kann Ihnen dabei helfen, Funktionen zwischen Konten und Agentendiensten gemeinsam zu nutzen.
- Amazon SageMaker AI ist ein Service, den Sie für Modelltraining, Feinabstimmung und Inferenz verwenden können, die über grundlegende Modelle hinausgehen.
- Amazon Simple Storage Service (Amazon S3) bietet Objektspeicher für Prompt-Bibliotheken, Modellartefakte und von Agenten generierte Daten.
- AWS Step Functionsist eine Workflow-Engine, die Ihnen helfen kann, Abläufe mit mehreren Agenten und Umschulungspipelines zu koordinieren.
- AWS X-Raybietet verteiltes Tracing, mit dem Sie Entscheidungsabläufe von Agenten und Serviceabhängigkeiten verfolgen können.

Andere Ressourcen AWS

- Grundlagen der Agenten-KI auf AWS
- Agentengestützte KI-Muster und Workflows auf AWS
- Agentische KI-Frameworks, Protokolle und Tools auf AWS
- Aufbau serverloser Architekturen für agentische KI auf AWS
- Aufbau von Mehrmandantenarchitekturen für agentische KI auf AWS

Dokumentverlauf

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

Änderung	Beschreibung	Datum
Erste Veröffentlichung	—	12. August 2025

AWS Glossar zu präskriptiven Leitlinien

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

Zahlen

7 Rs

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

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

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

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

A

ABAC

Siehe attributbasierte Zugriffskontrolle.

abstrahierte Dienste

Siehe Managed Services.

ACID

Siehe Atomarität, Konsistenz, Isolierung und Haltbarkeit.

Aktiv-Aktiv-Migration

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

Aktiv-Passiv-Migration

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

Aggregatfunktion

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

AI

Siehe künstliche Intelligenz.

AIOps

Siehe Operationen im Bereich künstliche Intelligenz.

Anonymisierung

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

Anti-Muster

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

Anwendungssteuerung

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

Anwendungsportfolio

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

künstliche Intelligenz (KI)

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

Operationen mit künstlicher Intelligenz (AIOps)

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

Asymmetrische Verschlüsselung

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

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

Atomizität, Konsistenz, Isolierung, Haltbarkeit (ACID)

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

Attributbasierte Zugriffskontrolle (ABAC)

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

maßgebliche Datenquelle

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

Availability Zone

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

AWS Framework für die Einführung der Cloud (AWS CAF)

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

AWS Workload-Qualifizierungsrahmen (AWS WQF)

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

B

schlechter Bot

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

BCP

Siehe Planung der Geschäftskontinuität.

Verhaltensdiagramm

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

Big-Endian-System

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

Binäre Klassifikation

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

Bloom-Filter

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

Blau/Grün-Bereitstellung

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

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

Bot

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

Botnetz

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

branch

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

Zugang durch Glasbruch

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

Brownfield-Strategie

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

Puffer-Cache

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

Geschäftsfähigkeit

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

Planung der Geschäftskontinuität (BCP)

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

C

CAF

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

Bereitstellung auf Kanaren

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

CCoE

Weitere Informationen finden Sie im Cloud Center of Excellence.

CDC

Siehe Erfassung von Änderungsdaten.

Erfassung von Datenänderungen (CDC)

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

Chaos-Technik

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

CI/CD

Siehe Continuous Integration und Continuous Delivery.

Klassifizierung

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

clientseitige Verschlüsselung

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

Cloud-Exzellenzzentrum (CCoE)

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

Cloud Computing

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

Cloud-Betriebsmodell

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

Phasen der Einführung der Cloud

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

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

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

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

CMDB

Siehe Datenbank für das Konfigurationsmanagement.

Code-Repository

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

Kalter Cache

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

Kalte Daten

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

Computer Vision (CV)

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

Drift in der Konfiguration

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

Verwaltung der Datenbankkonfiguration (CMDB)

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

Konformitätspaket

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

Kontinuierliche Bereitstellung und kontinuierliche Integration (CI/CD)

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

CV

Siehe Computer Vision.

D

Daten im Ruhezustand

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

Datenklassifizierung

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

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

Datendrift

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

Daten während der Übertragung

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

Datennetz

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

Datenminimierung

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

Datenperimeter

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

Vorverarbeitung der Daten

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

Herkunft der Daten

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

betroffene Person

Eine Person, deren Daten gesammelt und verarbeitet werden.

Data Warehouse

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

Datenbankdefinitionssprache (DDL)

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

Datenbankmanipulationssprache (DML)

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

DDL

Siehe Datenbankdefinitionssprache.

Deep-Ensemble

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

Deep Learning

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

defense-in-depth

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

delegierter Administrator

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

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

Einsatz

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

Entwicklungsumgebung

Siehe Umgebung.

Detektivische Kontrolle

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

Abbildung des Wertstroms in der Entwicklung (DVSM)

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

digitaler Zwilling

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

Maßtabelle

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

Katastrophe

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

Notfallwiederherstellung (DR)

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

DML

Siehe Sprache zur Datenbankmanipulation.

Domainorientiertes Design

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

DR

Siehe Disaster Recovery.

Erkennung von Driften

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

DVSM

Siehe Abbildung des Wertstroms in der Entwicklung.

E

EDA

Siehe explorative Datenanalyse.

EDI

Siehe elektronischer Datenaustausch.

Edge-Computing

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

elektronischer Datenaustausch (EDI)

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

Verschlüsselung

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

Verschlüsselungsschlüssel

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

Endianismus

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

Endpunkt

Siehe Service-Endpunkt.

Endpunkt-Services

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

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

Unternehmensressourcenplanung (ERP)

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

Envelope-Verschlüsselung

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

Umgebung

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

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

Epics

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

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

ERP

Siehe Enterprise Resource Planning.

Explorative Datenanalyse (EDA)

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

F

Faktentabelle

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

schnell scheitern

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

Grenze zur Fehlerisolierung

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

Feature-Zweig

Siehe Zweig.

Features

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

Bedeutung der Feature

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

Featuretransformation

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

Eingabeaufforderung mit wenigen Klicks

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

FGAC

Siehe detaillierte Zugriffskontrolle.

Feinkörnige Zugriffskontrolle (FGAC)

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

Flash-Cut-Migration

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

FM

Siehe Fundamentmodell.

Fundamentmodell (FM)

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

G

Generative KI

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

Geoblocking

Siehe geografische Einschränkungen.

Geografische Einschränkungen (Geoblocking)

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

Gitflow-Workflow

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

goldenes Bild

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

Greenfield-Strategie

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

Integritätsschutz

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

H

HEKTAR

Siehe Hochverfügbarkeit.

Heterogene Datenbankmigration

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

hohe Verfügbarkeit (HA)

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

historische Modernisierung

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

Daten zurückhalten

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

Homogene Datenbankmigration

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

heiße Daten

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

Hotfix

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

Hypercare-Phase

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

I

IaC

Sehen Sie Infrastruktur als Code.

Identitätsbasierte Richtlinie

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

Leerlaufanwendung

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

IIoT

Siehe Industrielles Internet der Dinge.

unveränderliche Infrastruktur

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

Eingehende (ingress) VPC

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

Inkrementelle Migration

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

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

Industrie 4.0

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

Infrastruktur

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

Infrastructure as Code (IaC)

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

industrielles Internet der Dinge (T) IIo

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

Inspektions-VPC

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

Internet of Things (IoT)

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

Interpretierbarkeit

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

IoT

Siehe Internet der Dinge.

IT information library (ITIL, IT-Informationsbibliothek)

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

T service management (ITSM, IT-Servicemanagement)

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

BIS

Siehe IT-Informationsbibliothek.

ITSM

Siehe IT-Servicemanagement.

L

Labelbasierte Zugangskontrolle (LBAC)

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

Landing Zone

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

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

großes Sprachmodell (LLM)

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

Große Migration

Eine Migration von 300 oder mehr Servern.

SCHWARZ

Siehe Labelbasierte Zugriffskontrolle.

Geringste Berechtigung

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

Lift and Shift

Siehe 7 Rs.

Little-Endian-System

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

LLM

Siehe großes Sprachmodell.

Niedrigere Umgebungen

Siehe Umgebung.

M

Machine Learning (ML)

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

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

Hauptzweig

Siehe Filiale.

Malware

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

verwaltete Dienste

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

Manufacturing Execution System (MES)

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

MAP

Siehe Migration Acceleration Program.

Mechanismus

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

Mitgliedskonto

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

MES

Siehe Manufacturing Execution System.

Message Queuing-Telemetrietransport (MQTT)

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

Microservice

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

Microservices-Architekturen

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

Migration Acceleration Program (MAP)

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

Migration in großem Maßstab

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

Migrationsfabrik

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

Migrationsmetadaten

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

Migrationsmuster

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

Migration Portfolio Assessment (MPA)

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

Migration Readiness Assessment (MRA)

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

Migrationsstrategie

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

ML

Siehe maschinelles Lernen.

Modernisierung

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

Bewertung der Modernisierungsfähigkeit

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

Monolithische Anwendungen (Monolithen)

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

MPA

Siehe Bewertung des Migrationsportfolios.

MQTT

Siehe Message Queuing-Telemetrietransport.

Mehrklassen-Klassifizierung

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

veränderbare Infrastruktur

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

O

OAC

Siehe Origin Access Control.

EICHE

Siehe Zugriffsidentität von Origin.

COM

Siehe organisatorisches Change-Management.

Offline-Migration

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

OI

Siehe Betriebsintegration.

OLA

Siehe Vereinbarung auf operativer Ebene.

Online-Migration

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

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

OPC-UA

Siehe Open Process Communications — Unified Architecture.

Offene Prozesskommunikation — Einheitliche Architektur (OPC-UA)

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

Vereinbarung auf Betriebsebene (OLA)

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

Überprüfung der Betriebsbereitschaft (ORR)

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

Betriebstechnologie (OT)

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

Betriebsintegration (OI)

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

Organisationspfad

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

Organisatorisches Veränderungsmanagement (OCM)

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

Ursprungszugriffskontrolle (OAC)

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

Ursprungszugriffsidentität (OAI)

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

ORR

Weitere Informationen finden Sie unter Überprüfung der Betriebsbereitschaft.

NICHT

Siehe Betriebstechnologie.

Ausgehende (egress) VPC

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

P

Berechtigungsgrenze

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

persönlich identifizierbare Informationen (PII)

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

Personenbezogene Daten

Siehe persönlich identifizierbare Informationen.

Playbook

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

PLC

Siehe programmierbare Logiksteuerung.

PLM

Siehe Produktlebenszyklusmanagement.

policy

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

Polyglotte Beharrlichkeit

Unabhängige Auswahl der Datenspeichertechnologie eines Microservices auf der Grundlage von Datenzugriffsmustern und anderen Anforderungen. Wenn Ihre Microservices über dieselbe Datenspeichertechnologie verfügen, kann dies zu Implementierungsproblemen oder zu

Leistungseinbußen führen. Microservices lassen sich leichter implementieren und erzielen eine bessere Leistung und Skalierbarkeit, wenn sie den Datenspeicher verwenden, der ihren Anforderungen am besten entspricht.

Portfoliobewertung

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

predicate

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

Prädikat Pushdown

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

Präventive Kontrolle

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

Prinzipal

Eine Entität AWS , die Aktionen ausführen und auf Ressourcen zugreifen kann. Diese Entität ist in der Regel ein Root-Benutzer für eine AWS-Konto, eine IAM-Rolle oder einen Benutzer. Weitere Informationen finden Sie unter Prinzipal in Rollenbegriffe und -konzepte in der IAM-Dokumentation.

Datenschutz von Natur aus

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

Privat gehostete Zonen

Ein Container, der Informationen darüber enthält, wie Amazon Route 53 auf DNS-Abfragen für eine Domain und deren Subdomains innerhalb einer oder mehrerer VPCs Domains antworten soll. Weitere Informationen finden Sie unter Arbeiten mit privat gehosteten Zonen in der Route-53-Dokumentation.

proaktive Steuerung

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

Produktlebenszyklusmanagement (PLM)

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

Produktionsumgebung

Siehe Umgebung.

Speicherprogrammierbare Steuerung (SPS)

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

schnelle Verkettung

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

Pseudonymisierung

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

publish/subscribe (pub/sub)

Ein Muster, das asynchrone Kommunikation zwischen Microservices ermöglicht, um die Skalierbarkeit und Reaktionsfähigkeit zu verbessern. In einem auf Microservices basierenden MES kann ein Microservice beispielsweise Ereignismeldungen in einem Kanal veröffentlichen, den andere Microservices abonnieren können. Das System kann neue Microservices hinzufügen, ohne den Veröffentlichungsservice zu ändern.

Q

Abfrageplan

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

Abfrageplanregression

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

R

RACI-Matrix

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

RAG

Siehe Erweiterte Generierung beim Abrufen.

Ransomware

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

RASCI-Matrix

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

RCAC

Siehe Zugriffskontrolle für Zeilen und Spalten.

Read Replica

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

neu strukturieren

Siehe 7 Rs.

Recovery Point Objective (RPO)

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

Wiederherstellungszeitziel (RTO)

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

Refaktorierung

: Siehe 7 Rs.

Region

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

Regression

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

rehosten

: Siehe 7 Rs.

Veröffentlichung

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

umziehen

: Siehe 7 Rs.

neue Plattform

: Siehe 7 Rs.

Rückkauf

Siehe 7 Rs.

Ausfallsicherheit

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

Ressourcenbasierte Richtlinie

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

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

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

Reaktive Kontrolle

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

Beibehaltung

Siehe 7 Rs.

zurückziehen

Siehe 7 Rs.

Retrieval Augmented Generation (RAG)

Eine generative KI-Technologie, bei der ein LLM auf eine maßgebliche Datenquelle verweist, die sich außerhalb seiner Trainingsdatenquellen befindet, bevor eine Antwort generiert wird. Ein

RAG-Modell könnte beispielsweise eine semantische Suche in der Wissensdatenbank oder in benutzerdefinierten Daten einer Organisation durchführen. Weitere Informationen finden Sie unter Was ist RAG.

Drehung

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

Zugriffskontrolle für Zeilen und Spalten (RCAC)

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

RPO

Siehe Recovery Point Objective.

RTO

Siehe Ziel für die Erholungszeit.

Runbook

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

S

SAML 2.0

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

SCADA

Siehe Aufsichtskontrolle und Datenerfassung.

SCP

Siehe Richtlinie zur Dienstkontrolle.

Secret

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

Sicherheit durch Design

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

Sicherheitskontrolle

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

Härtung der Sicherheit

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

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

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

Automatisierung von Sicherheitsreaktionen

Eine vordefinierte und programmierte Aktion, die darauf ausgelegt ist, automatisch auf ein Sicherheitsereignis zu reagieren oder es zu beheben. Diese Automatisierungen dienen als detektive oder reaktionsschnelle Sicherheitskontrollen, die Sie bei der Implementierung bewährter

AWS Sicherheitsmethoden unterstützen. Beispiele für automatisierte Antwortaktionen sind das Ändern einer VPC-Sicherheitsgruppe, das Patchen einer Amazon EC2 EC2-Instance oder das Rotieren von Anmeldeinformationen.

Serverseitige Verschlüsselung

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

Service-Kontrollrichtlinie (SCP)

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

Service-Endpunkt

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

Service Level Agreement (SLA)

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

Service-Level-Indikator (SLI)

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

Service-Level-Ziel (SLO)

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

Modell der geteilten Verantwortung

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

SIEM

Siehe Sicherheitsinformations- und Event-Management-System.

Single Point of Failure (SPOF)

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

SLA

Siehe Service Level Agreement.

SLI

Siehe Service-Level-Indikator.

ALSO

Siehe Service-Level-Ziel.

split-and-seed Modell

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

SPOTTEN

Siehe Single Point of Failure.

Sternschema

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

Strangler-Fig-Muster

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

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

Subnetz

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

Aufsichtskontrolle und Datenerfassung (SCADA)

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

Symmetrische Verschlüsselung

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

synthetisches Testen

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

Systemaufforderung

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

T

tags

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

Zielvariable

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

Aufgabenliste

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

Testumgebungen

Siehe Umgebung.

Training

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

Transit-Gateway

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

Stammbasierter Workflow

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

Vertrauenswürdiger Zugriff

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

Optimieren

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

Zwei-Pizzen-Team

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

U

Unsicherheit

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

undifferenzierte Aufgaben

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

höhere Umgebungen

Siehe Umgebung.

V

Vacuuming

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

Versionskontrolle

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

VPC-Peering

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

Schwachstelle

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

W

Warmer Cache

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

warme Daten

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

Fensterfunktion

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

Workload

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

Workstream

Funktionsgruppen in einem Migrationsprojekt, die für eine bestimmte Reihe von Aufgaben verantwortlich sind. Jeder Workstream ist unabhängig, unterstützt aber die anderen Workstreams

im Projekt. Der Portfolio-Workstream ist beispielsweise für die Priorisierung von Anwendungen, die Wellenplanung und die Erfassung von Migrationsmetadaten verantwortlich. Der Portfolio-Workstream liefert diese Komponenten an den Migrations-Workstream, der dann die Server und Anwendungen migriert.

WURM

Sehen Sie einmal schreiben, viele lesen.

WQF

Siehe AWS Workload-Qualifizierungsrahmen.

einmal schreiben, viele lesen (WORM)

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

Z

Zero-Day-Exploit

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

Zero-Day-Sicherheitslücke

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

Eingabeaufforderung ohne Zwischenfälle

Bereitstellung von Anweisungen für die Ausführung einer Aufgabe an einen LLM, jedoch ohne Beispiele (Schnappschüsse), die ihm als Orientierungshilfe dienen könnten. Der LLM muss sein vortrainiertes Wissen einsetzen, um die Aufgabe zu bewältigen. Die Effektivität von Zero-Shot Prompting hängt von der Komplexität der Aufgabe und der Qualität der Aufforderung ab. Siehe auch Few-Shot-Prompting.

Zombie-Anwendung

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

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