



Ökonomie für agentische KI am AWS

AWS Präskriptive Leitlinien



AWS Präskriptive Leitlinien: Ökonomie für agentische KI am AWS

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

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

Table of Contents

Einführung	1
Zielgruppe	1
Ziele	2
Über diese Inhaltsserie	2
Die agentische KI-Ökonomie verstehen	3
Bewertung der Aufgabe	3
Bewertung der Auswirkungen auf Risiken	4
Kapitalrendite	6
Messung von Erfolg und ROI	8
Nutze deine Stiftung	8
Setzen Sie sich Ziele	8
Verfolgen Sie Kennzahlen	8
Benutzen AgentOps	9
Bewertung der Kosten für menschliche Prozesse	9
Arbeitskosten	10
Kosten für Leistung	10
Technologiekosten	11
Opportunitätskosten	12
Kosten für Risiken und Fehler	12
Implementierung agentischer KI-Systeme	13
Einbindung von menschlichem Feedback	14
Verhaltensorientiertes Lernen	14
Kontinuierliches Lernen	15
Zusammenarbeit zwischen Mensch und KI	16
Ergebnisorientierte Preisgestaltung	16
Herkömmliches Modell, das direkt im Voraus erfolgt	17
Ergebnisorientiertes Modell	17
Verwenden AWS Marketplace	18
Fallstudie: Rekrutierungsmaßnahmen	20
Szenario A	20
Grundkostenstruktur	21
Operationelle Metriken	22
Volumenbasierte Kostenanalyse	23
ROI-Analyse	23

Vergleich der kumulierten Kosten	24
Zusätzliche Vorteile	24
Szenario B	24
Grundlegende Kostenstruktur	25
Operationelle Metriken	26
Volumenbasierte Kostenanalyse	26
ROI-Analyse	27
Vergleich der kumulierten Kosten	27
Vergleich von Szenarien	28
Fazit und Ressourcen	30
Ressourcen	30
Dokumentverlauf	32
Glossar	33
#	33
A	34
B	37
C	39
D	42
E	47
F	49
G	51
H	52
I	54
L	56
M	57
O	62
P	65
Q	68
R	68
S	71
T	75
U	77
V	77
W	78
Z	79
.....	lxxx

Ökonomie für agentische KI auf AWS

Hans Schabert und Prasanta Roy, Amazon Web Services

Januar 2026 ([Geschichte der Dokumente](#))

Organizations, die KI-gesteuerte Automatisierung und agentische KI-Systeme einsetzen, müssen fundierte wirtschaftliche Entscheidungen zwischen menschlichen Arbeitskräften und intelligenten Agenten treffen. Dies wird für einen nachhaltigen Cloud-Betrieb von entscheidender Bedeutung. Dieser Leitfaden hilft Ihnen dabei, die wirtschaftlichen Kompromisse zwischen menschlicher Belegschaft und behördlichen KI-Systemen zu bewerten, zu implementieren und zu optimieren. AWS Sie können Ihre Kapitalrendite (ROI) maximieren und gleichzeitig Ihre betriebliche Exzellenz aufrechterhalten.

Kein System ist zu 100% richtig. Dieses grundlegende Prinzip bestimmt die wirtschaftliche Analyse menschlicher und agentischer KI-Systeme. Organizations müssen über einfache Kostenvergleiche hinausgehen, um die gesamten wirtschaftlichen Auswirkungen, Risikoprofile, Anforderungen an die Entscheidungsqualität und die langfristige strategische Wertschöpfung zu bewerten.

Das Kundenverhalten verändert sich dramatisch von traditionellen Technologieinvestitionen im Vorfeld hin zu pay-per-outcome Modellen, die die Kosten an den Geschäftsergebnissen ausrichten. Diese Transformation erfordert neue Ansätze für die Bewertung, Implementierung und Optimierung der Zusammenarbeit zwischen Mensch und Agent.

Der Weg zum Erfolg folgt einem klaren Muster: Beginnen Sie mit geeigneten Jobs, messen Sie alles und skalieren Sie, was funktioniert. Organizations, die diesen Ansatz verfolgen, erzielen durch intelligente Ressourcenzuweisung und ergebnisorientierte Automatisierung einen nachhaltigen Wettbewerbsvorteil.

Zielgruppe

Dieser Leitfaden richtet sich an folgende Themen:

- Führungskräfte (CEOs, CTOs, CFOs), die strategische Anlageentscheidungen treffen
- Unternehmensarchitekten, die Strategien zur organisatorischen Automatisierung entwerfen
- Praktiker im Finanzbereich, die das Cloud-Finanzmanagement optimieren
- Technologieführer, die Ansätze zur KI-Implementierung evaluieren

- Leiter von Geschäftsbereichen, die den ROI der Automatisierung verstehen möchten
- Einkaufsexperten, die sich mit neuen KI-Preismodellen auseinandersetzen

Um die Konzepte in diesem Leitfaden zu verstehen, empfehlen wir Ihnen, die [Grundlagen der KI für Agenturen unter zu lesen. AWS](#)

Ziele

Dieser Leitfaden hilft Ihnen dabei, Folgendes zu verstehen:

- Wie bewertet man Jobs im Hinblick auf ihr automatisiertes Potenzial
- Ökonomische Modelle für den Vergleich menschlicher Arbeitskosten mit Investitionen in künstliche KI-Systeme
- Pay-per-outcome Preismodelle und ihre Auswirkungen auf die Wirtschaftlichkeit von KI-Projekten
- Messtechniken zum Nachweis des ROI und zum Risikomanagement
- Skalierungsstrategien, die Fixkosten in variable Ergebnisse umwandeln

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

Die agentische KI-Ökonomie verstehen am AWS

Eines der wichtigsten Prinzipien besteht darin, zu bestimmen, wann KI-Agenten und wann traditionelle deterministische Methoden eingesetzt werden sollten. Organizations müssen systematisch bewerten, für welche Aufgaben eine Automatisierung durch Behörden gerechtfertigt ist und bei welchen die traditionelle Automatisierung oder der kontinuierliche Betrieb durch den Menschen zum Einsatz kommen sollte. Diese Entscheidung setzt voraus, dass der Zusammenhang zwischen Aufgabenmerkmalen, Risikobereitschaft und operativem Ansatz verstanden wird.

Bevor Sie sich für die Implementierung von agentischer KI entscheiden, sollten Sie den Entscheidungsrahmen nutzen, um die wirtschaftlichen Auswirkungen zu verstehen. Der Entscheidungsrahmen umfasst die folgenden drei Schlüsselfragen:

1. [Aufgabenbeurteilung](#) — Ist diese Aufgabe für einen KI-Agenten geeignet?
2. [Risikofolgenabschätzung](#) — Was sind die damit verbundenen Risiken?
3. [Kapitalrendite](#) — Wird es kosteneffektiv sein?

Bewertung der Aufgabe

Aufgaben mit hochkomplexen, standardisierten Entscheidungsregeln können von agentischen KI-Ansätzen profitieren. Hochgradig standardisierte, einfache Aufgaben lassen sich besser mit herkömmlicher Automatisierung oder robotergestützter Prozessautomatisierung bewältigen. Agentische KI-Systeme zeichnen sich dadurch aus, dass sie argumentieren, den Kontext verstehen oder adaptiv Entscheidungen treffen können. Sie bieten einen Mehrwert, der über die regelbasierte Verarbeitung hinausgeht. Erfolgreiche KI-Implementierungen von Agenturen erfordern Systeme, die lernen und sich anpassen können.

Berücksichtigen Sie bei der Bewertung einer Aufgabe die folgenden Faktoren:

- Komplexität — Grad der Argumentation und des Verständnisses des Kontextes erforderlich. Aufgaben, die kontextuelles Verständnis, nuancierte Interpretation oder anpassungsfähige Reaktionen auf sich ändernde Bedingungen erfordern, bevorzugen agentische Ansätze gegenüber herkömmlicher Automatisierung, während rein mechanische oder rechnerische Aufgaben möglicherweise keine agentische Intelligenz erfordern.

- **Standardisierung** — Vorhandensein klarer Muster und Regeln. Agentic AI wird empfohlen, wenn die Aufgabe ein kontextuelles Verständnis erfordert. Wenn keine Anpassung oder kein Lernen erforderlich ist, sollten Sie die traditionelle Automatisierung in Betracht ziehen.
- **Volumen** — Häufigkeit der Aufgabenausführung. Agentic AI wird für autonome Aktivitäten empfohlen. Herkömmliche Automatisierung wird für umfangreiche, konsistente Aufgaben empfohlen. Das Volumen allein bestimmt jedoch nicht den Ansatz. Entscheidungen mit geringem Volumen und hohem Wert könnten eher eine behördliche Unterstützung zur Verbesserung der Entscheidungsqualität als zur Kostensenkung rechtfertigen.
- **Wert** — Geschäftsauswirkung pro erledigter Aufgabe. Ziehen Sie künstliche Intelligenz in Betracht, wenn Sie hochwertige Ergebnisse erzielen möchten, für die menschenähnliche autonome Fähigkeiten erforderlich sind. Ziehen Sie traditionelle Automatisierung für sich wiederholende, konsistente Aufgaben in Betracht, die deterministisch ausgeführt werden können.

Bewertung der Auswirkungen auf Risiken

Derzeit gibt es vier Ansätze für den Einsatz von KI durch Behörden: vollständig autonom, vom Menschen am laufenden Band, als Co-Pilot oder von Menschen geführter Einsatz mit Agentenunterstützung. Jeder hat sein eigenes Risikoprofil und seine eigene Fehlertoleranz, und alle beziehen in gewisser Weise Menschen mit ein. In der folgenden Tabelle werden die Risikodetails dieser Ansätze beschrieben.

Grad der Autonomie	Risikoprofil	Fehlertoleranz	Beispiele für Anwendungsfälle	Menschliches Engagement
Vollständig autonom	Geringes Risiko	1-2% akzeptabel	<ul style="list-style-type: none"> • Grundlegende Datenkategorisierung • Weiterleitung von Dokumenten • Generierung von Standardberichten 	<ul style="list-style-type: none"> • Minimaler Überblick • Regelmäßige Prüfungen

Der Mensch ist auf dem Laufenden	Mittleres Risiko	Unter 0,5%	<ul style="list-style-type: none"> • Entwürfe von Antworten • Moderation von Inhalten • Erste Bearbeitung von Reklamationen 	<ul style="list-style-type: none"> • Regelmäßige Überprüfung • Ausnahmev erarbeitung • Qualitäts sicherung
Co-Pilot	Hohes Risiko	Nahezu Null	<ul style="list-style-type: none"> • Input zur strategischen Planung • Risikobew ertungen • Anlageent scheidungen 	<ul style="list-style-type: none"> • Der Mensch trifft endgültig e Entscheidungen • Der Agent gibt Empfehlungen
Von Menschen geführt mit Agentenunterstützung	Kritisches Risiko	Null-Toleranz	<ul style="list-style-type: none"> • Rechtliche Entscheidungen • Medizinische Diagnose • Einhaltung gesetzlicher Vorschriften 	<ul style="list-style-type: none"> • Der Mensch treibt den Prozess an • Der Agent stellt nur Recherche n oder Analysen und unterstützende Informationen zur Verfügung

In der folgenden Tabelle werden die wichtigsten Überlegungen bei der Auswahl zwischen diesen Ansätzen beschrieben.

Überlegung	Vollständig autonom	Der Mensch ist auf dem Laufenden	Copilot	Von Menschen geführt
Kosteneffizienz	Am höchsten	Hoch	Mittel	Niedrig
Skalierbarkeit	Unbegrenzt	Hoch	Mittel	Begrenzt
Geschwindigkeit der Verarbeitung	Am schnellsten	Schnell	Mittel	Langsam
Risikomanagement	Basic	Verbessert	Stark	Am stärksten
Umgang mit Komplexität	Einfache Aufgaben	Mäßig komplexe Aufgaben	Komplexe Aufgaben	Kritische Aufgaben

Dieser Rahmen für Überlegungen hilft Unternehmen dabei, die Autonomiestufen an Risikoprofile anzupassen, Abläufe angemessen zu skalieren, Effizienz und Kontrolle in Einklang zu bringen, eine angemessene Unternehmensführung zu implementieren und die Ressourcenzuweisung zu optimieren.

Kapitalrendite

Die Berechnung der Kapitalrendite für agentische KI-Systeme beginnt mit einer umfassenden Kostenanalyse. Organizations müssen zunächst ihre aktuellen Personalkosten, einschließlich Gehalt, Sozialleistungen und Arbeitsplatzkosten, zusammen mit prozessspezifischen Ausgaben und versteckten Kosten wie Schulung, Versicherungsschutz und Ausfallzeiten berechnen.

Für eine Break-Even-Analyse sollten Unternehmen die Implementierungskosten, die laufenden Betriebskosten und das zur Rechtfertigung der Investition erforderliche Volumen berücksichtigen. Es ist auch wichtig, saisonale Schwankungen und die Vorteile der Lernkurve zu berücksichtigen, die sich ergeben, wenn Systeme immer ausgereifter werden und sich im Laufe der Zeit verbessern.

Bei der Bewertung von KI-Agenten sollten Unternehmen berücksichtigen, dass diese Systeme in der Regel höhere Vorlaufkosten, aber niedrigere Kosten pro Transaktion im Vergleich zu menschlichen Abläufen haben. Darüber hinaus weisen KI-Agenten im Laufe der Zeit eine Verbesserung der Leistung auf und bieten eine bessere Skalierbarkeit als menschliche Teams. Dies macht sie immer kostengünstiger, da der Einsatz skaliert und die Betriebserfahrung zunimmt.

Messung des Erfolgs und des ROI agentischer KI-Systeme

Die Erfolgsmessung bei der Implementierung von KI-Systemen durch Agenturen erfordert einen systematischen Ansatz. Dieser Abschnitt bietet eine klare Methodik für die Bewertung und kontinuierliche Optimierung, bei der Ihre bestehende Analyse verwendet wird, anstatt bei Null zu beginnen.

Schritt 1: Verwenden Sie Ihre bestehende Grundlage

Beginnen Sie mit einer umfassenden Kostenbewertung gemäß den Empfehlungen im Abschnitt [Bewertung Ihrer aktuellen Prozesskosten](#). Dies bietet eine betriebliche Grundlage für Ihre ROI-Berechnungen. Wie im Abschnitt [Risikofolgenabschätzung](#) beschrieben, wählen Sie zwischen den vier Autonomiestufen (vollständig autonom, menschliches Eingreifen, Co-Pilot, von Menschen geleitet mit Unterstützung von Agenten), um für jeden Prozess die geeigneten Messkriterien und Fehlertoleranzschwellen festzulegen.

Schritt 2: Setzen Sie sich klare Erfolgsziele

Legen Sie Architektur- und Erfolgsziele fest, bei denen lernfähige Systeme im Vordergrund stehen, wie im Abschnitt [Erfolgreiche Muster für die Implementierung agentischer KI-Systeme](#) beschrieben. Konzentrieren Sie sich eher auf kontinuierliche Verbesserungen als auf statische Leistung. Legen Sie Zeitpläne für den ROI fest, indem Sie die [Break-Even-Analysemethode verwenden, die in der Fallstudie: Vergleich der Kosten menschlicher und behördlicher KI für Personalbeschaffungsmaßnahmen demonstriert wurde](#). Geben Sie klare Entscheidungsgrundlagen für die Kündigung von Mitarbeitern an, die keine Leistungen erbringen.

Schritt 3: Verfolgen Sie wichtige Kennzahlen

Überwachen Sie Ihre finanzielle Leistung anhand Ihrer festgelegten Ausgangsbasis und verfolgen Sie Kosteneinsparungen und strategische Wertverbesserungen. Messen Sie betriebliche Kennzahlen, darunter Fehlerquoten innerhalb akzeptabler Schwellenwerte für die von Ihnen gewählte Autonomiestufe, Verbesserungen der Verarbeitungsgeschwindigkeit und Konsistenzgewinne. Konzentrieren Sie sich auf strategische Indikatoren, die die Lernfähigkeit und die Anpassung im Laufe der Zeit belegen.

Schritt 4: Verwenden AgentOps

Wenden Sie das Framework für kontinuierliches Lernen aus dem Abschnitt [Integration menschlichen Feedbacks in agentische KI-Systeme](#) an, um die Entscheidungsfindung durch systematische Integration von menschlichem Feedback zu optimieren. Entwickeln Sie Echtzeit-Lernsysteme, die menschliche Erkenntnisse zur Leistungssteigerung einbeziehen. Überwachen Sie den Wandel hin zu ergebnisorientierten Geschäftsmodellen, wie unter [Wirtschaftlicher Wandel hin zu ergebnisorientierter Preisgestaltung](#) für agentische KI-Systeme beschrieben. AWS

Bewertung Ihrer aktuellen Kosten für menschliche Prozesse

Ihre tatsächlichen Prozesskosten zu verstehen, ist von grundlegender Bedeutung, um fundierte Entscheidungen über Investitionen in Agentic KI-Systeme zu treffen. Zunächst müssen Sie eine genaue Bestandsaufnahme der aktuellen Prozesskosten erstellen, einschließlich aller versteckten Kosten, Ausfallraten und Opportunitätskosten. Dies hilft Ihnen, präzise ROI-Berechnungen zu entwickeln und strategische Entscheidungen zu treffen. Diese umfassende Kostenbewertung dient als wichtige Grundlage für die Bewertung, ob agentische KI-Systeme als produktive Begleiter einen echten Mehrwert bieten können.

Die Bewertung der Basiskosten ist aus den folgenden Hauptgründen unerlässlich:

- Genauigkeit der Investitionsrendite — Genaue Kostengrundlagen unterstützen realistische ROI-Prognosen, die das gesamte Spektrum der aktuellen Betriebskosten berücksichtigen.
- Implementierungsstrategie für Agenten — Ein umfassendes Verständnis der Kosten hilft Unternehmen dabei, die vielversprechendsten Prozesse für die erste Implementierung von KI-Systemen durch Agenturen zu identifizieren.
- Leistungsmessung — Etablierte Referenzwerte bilden den Rahmen für die Messung der tatsächlichen und prognostizierten Vorteile von KI-Implementierungen durch Agenturen.

Organizations müssen systematisch alle Kostenfaktoren identifizieren und bewerten, die die Prozessökonomie beeinflussen, bevor sie menschliche und behördliche Alternativen vergleichen. Diese Bewertung gewährleistet genaue Basisberechnungen, indem sowohl offensichtliche als auch versteckte Kostentreiber berücksichtigt werden. Dabei wird besonderer Wert auf Ausfallkosten, historische Ausfallraten und verpasste Geschäftschancen gelegt, die die tatsächlichen Gesamtkosten der aktuellen Prozesse darstellen.

In diesem Abschnitt wird beschrieben, wie Sie Daten in jeder Kostenkategorie sammeln, um genaue Basiswerte für Ihre aktuellen Prozesse zu ermitteln. Es werden Informationsquellen erörtert und Beispiele für die folgenden Kostenkategorien bereitgestellt:

- [Arbeitskosten](#)
- [Kosten für menschliche Leistung und Beständigkeit](#)
- [Technologie- und Infrastrukturkosten](#)
- [Kosten für entgangene Geschäftschancen](#)
- [Kosten für Risiken und Mängel](#)

Arbeitskosten

Extrahieren Sie Gehaltsabrechnungsdaten für 24 Monate, einschließlich Grundgehalt, Überstunden, Sozialleistungen und Schulungskosten. Verwenden Sie Ihr Personalinformationssystem (HRIS), um Personalbeschaffungskosten und Fluktuationsraten zu verfolgen. Zeiterfassungssysteme geben Aufschluss über die tatsächliche Produktivität im Vergleich zu den geplanten Arbeitsstunden. Plattformen für das Leistungsmanagement zeigen den Zusammenhang zwischen Qualifikationsniveau und Vergütungskosten. Berechnen Sie die Stundensätze bei voller Auslastung und die Zuweisung der Verwaltungskosten.

Im Folgenden finden Sie ein Beispiel für Kostentreiber im Bereich Arbeit.

Kostentreiber	Auswirkungen auf das Geschäft
Basisvergütung	25—150 \$ pro Stunde bei voller Ladung
Leistungen und Lohnsteuern	25— 40% des Grundgehalts
Schulung und Entwicklung	5— 15% der jährlichen Arbeitskosten
Gemeinkosten für die Verwaltung	15-25% der direkten Arbeitskosten

Kosten für menschliche Leistung und Beständigkeit

Kombinieren Sie Daten aus Projektmanagementsystemen, die Abweichungen bei der Ausführung von Aufgaben aufzeigen, mit Anwesenheitssystemen. Dadurch können Fehlzeitenmuster und saisonale Veränderungen aufgedeckt werden. Kundenservice-

Plattformen zeigen anhand von Lösungskennzahlen individuelle Leistungsbereiche auf, und Daten zum Kundenbeziehungsmanagement (CRM) im Vertrieb können Effizienzunterschiede beim Geschäftsabschluss aufzeigen. Qualitätsmanagementsysteme liefern Team- und standortübergreifende Daten zu Fehlerquoten und Prozesskonformität. Workflow-Systeme erfassen Abschlusszeiten, Genehmigungsverzögerungen und die Häufigkeit der Ausnahmebehandlung. Kommunikationsanalysen zeigen anhand der Häufigkeit von Besprechungen und der Muster der Zusammenarbeit den Koordinationsaufwand auf.

Im Folgenden finden Sie eine Beispielliste von Kostentreibern für die Leistung und Konsistenz von Mitarbeitern.

Kostentreiber	Auswirkungen auf das Geschäft
Schwankungen der Produktivität	Leistungsbereich von 20 bis 50%
Fehlzeiten und Versicherungsschutz	15-25% zusätzliche Kapazität erforderlich
Erschöpfungs- und Motivationszyklen	Produktivitätsvarianz von 10— 30%
Inkonsistenzen bei den Verfahren	Effizienzverlust von 10— 40%
Variationen bei der Qualitätskontrolle	10— 30% der Gesamtkosten
Gemeinkosten für die Koordination	15-25% der Betriebskosten

Technologie- und Infrastrukturkosten

Lizenzverwaltungsplattformen zeigen Softwarekosten und Nutzungsraten. Die Überwachung der Infrastruktur liefert Daten zur Verfügbarkeit, Leistungskennzahlen und Wartungskosten. Helpdesk-Systeme verfolgen den Support-Aufwand und wiederkehrende technische Probleme. Anbieterverwaltungssysteme erfassen die gesamten Kosten für die Zusammenarbeit mit Technologien, einschließlich Integrationskosten und Leistung auf Serviceniveau.

Im Folgenden finden Sie eine Beispielliste von Kostentreibern für Technologie und Infrastruktur.

Kostentreiber	Auswirkungen auf das Geschäft
Technologiesysteme	50—500 \$ pro Benutzer und Monat

Arbeitsbereich und Ausrüstung

200—1.000 USD pro Mitarbeiter und Monat

Kosten für entgangene Geschäftschancen

CRM-Plattformen enthalten Lead-Antwortzeiten, Konversionsraten und Unterlagen zu verpassten Geschäftschancen. Die Marketingautomatisierung zeigt, dass sich Verzögerungen bei der Nachverfolgung auf Lead-Konversionen auswirken. Kundensupportsysteme zeigen, wie sich betriebliche Probleme auf Zufriedenheit und Kundenbindung auswirken. Die Wettbewerbsanalyse liefert Marktreaktionsanforderungen und Daten zu Gewinnen oder Verlusten, die die betriebliche Leistung mit den Umsatzergebnissen verknüpfen.

Im Folgenden finden Sie eine Beispielliste von Kostentreibern für entgangene Geschäftschancen.

Kostentreiber	Auswirkungen auf das Geschäft
Verzögerungen bei der Marktreaktion	Umsatz pro Tag der Verspätung
Kapazitätsengpässe	Verlorene Geschäftsmöglichkeiten
Allokation von Innovationsressourcen	Opportunitätskosten von Routinearbeit
Verzögerungen bei der Kundenakquise	Verlust von 50 bis 90% Lead durch langsame Reaktion

Kosten für Risiken und Fehler

In den Unterlagen zur Versicherungspolice sind die Kosten für allgemeine Haftung, Berufshaftpflicht, Arbeitnehmerentschädigung und Cyber-Haftpflichtversicherung aufgeführt. Interne Risikobewertungsberichte identifizieren betriebliche Schwachstellen und die damit verbundenen Minderungskosten. Systeme zur Fehlerverfolgung dokumentieren Produkt- oder Serviceausfälle, einschließlich Kosten für deren Erkennung, Ersatzkosten und Garantieansprüche. In Zeitplänen für den Austausch von Anlagen werden die Ausfallraten und die Wiederbeschaffungskosten von Geräten angegeben. In Berichten über Sicherheitsvorfälle werden Arbeitsunfälle und die damit verbundenen Schadensersatzansprüche der Arbeitnehmer erfasst. In Plänen zur Geschäftskontinuität werden die Kosten für das Backup-System und die Investitionen in die Notfallwiederherstellung detailliert beschrieben.

Im Folgenden finden Sie eine Beispielliste von Kostentreibern für Risiken und Fehler.

Kostentreiber	Auswirkungen auf das Geschäft
Versicherungskosten	1— 5% des operativen Budgets
Kosten von Fehlern	50—5.000 \$ pro Fehlerfall
Auswirkungen menschlicher Fehler	2— 15% der gesamten Betriebskosten
Fehlerquoten und Nachbesserungen	Das 1,5- bis 4-fache der ursprünglichen Kosten für Korrekturen

Erfolgreiche Muster für die Implementierung agentischer KI-Systeme auf AWS

[Der Stand der Einführung von KI in Unternehmen](#) (ISG-Bericht 2025) zeigt, dass das Haupthindernis für eine erfolgreiche KI-Implementierung nicht die technischen Fähigkeiten, sondern die Lernlücke sind. Dieser Begriff bezieht sich auf Systeme, die sich nicht anpassen, sich nicht an den Kontext erinnern oder sich im Laufe der Zeit verbessern können. Organizations, die statische KI-Tools implementieren, verzeichnen hohe Ausfallraten. Im Folgenden sind die gemeinsamen Merkmale agentischer KI-Systeme aufgeführt, die erfolgreich sind:

- Kontextuelles Gedächtnis — Systeme, die den Konversationsverlauf und die Benutzerpräferenzen speichern
- Integration von Feedback — Fähigkeit, aus Korrekturen zu lernen und die Leistung zu verbessern
- Workflow-Anpassung — Automatische Anpassung an sich ändernde Geschäftsanforderungen
- Kontinuierliche Verbesserung — Messbare Verbesserung durch Betriebserfahrung

Organizations, die erfolgreiche KI-Implementierungen durchführen, priorisieren häufig Folgendes:

- Nutzung umfassender Partnerökosysteme, anstatt KI-Fähigkeiten eigenständig aufzubauen und zu erforschen
- Lernfähige Systeme statt statischer Tools
- Konzentrieren Sie sich lieber auf Geschäftsergebnisse als auf den Vergleich technischer Merkmale
- Workflow-Integration statt eigenständiger Tools

- Kontinuierliche Anpassung statt einmaliger Implementierung

[Diese Muster stimmen mit vielen AWS-Service Funktionen überein, insbesondere mit dem Zugriff auf das Basismodell in Amazon Bedrock, der ereignisgesteuerten Architektur in und der umfassenden Überwachung AWS Lambda, die von Amazon angeboten wird. CloudWatch](#) Weitere Informationen zur Integration von menschlichem Feedback und lernfähigen Systemen finden Sie im Abschnitt Integration von [menschlichem Feedback in agentische KI-Systeme in diesem Leitfaden](#).

Integration von menschlichem Feedback in agentische KI-Systeme

Kein System ist zu 100% erfolgreich, und Fehler sind unvermeidlich. Mit jedem Ausfall sind Änderungskosten verbunden. Human in the Loop ist ein KI-Ansatz, bei dem KI eine Aufgabe ausführt, aber menschliches Eingreifen oder Genehmigung erforderlich ist. Dieser Ansatz muss verwendet werden, wenn die Kosten eines Fehlers höher sind als die Kosten einer human-in-the-loop Lösung.

Der Erfolg agentischer KI-Systeme hängt im Wesentlichen von der Fähigkeit der Agenten ab, durch menschliches Feedback zu lernen und sich zu verbessern. Die Kosten der menschlichen Arbeit müssen je nach Umfang des erforderlichen Aufwands berücksichtigt werden. Im Gegensatz zu statischen Automatisierungstools, die vorgegebene Regeln ausführen, verfügen human-in-the-loop Lösungen über lernfähige Agentensysteme, die eine dynamische Partnerschaft zwischen den autonomen Agenten und dem Menschen schaffen. Menschliches Fachwissen verbessert kontinuierlich die Leistung des Agenten, während Agenten die routinemäßige Verarbeitung in großem Umfang übernehmen. Dieser kollaborative Ansatz verwandelt die KI-Implementierung von einer einmaligen Bereitstellung in einen kontinuierlichen Optimierungsprozess. Das System passt sich organisatorischen Mustern an, verinnerlicht Qualitätsstandards und verfeinert seine Entscheidungsfähigkeit auf der Grundlage realer Betriebserfahrungen. Durch die systematische Erfassung menschlicher Korrekturen, Genehmigungen und Erkenntnisse können Unternehmen KI-Agenten entwickeln, die den Kontext verstehen, Muster erkennen und sich im Laufe der Zeit zunehmend an den Geschäftszielen orientieren.

Bei Lösungen, die kein menschliches Eingreifen oder Unterstützung erfordern, müssen die personenspezifischen Kosten nicht in die Wirtschaftlichkeit der Agenten einfließen.

Verhaltensorientiertes Lernen durch menschliche Bediener

Menschliche Bediener geben kritisches Feedback, das agentische KI-Systeme nutzen können, um zu lernen, sich anzupassen und ihre Reaktionen im Laufe der Zeit zu verbessern. Diese Feedback-

Schleife schafft eine kollaborative Umgebung, in der menschliches Fachwissen die Fähigkeiten der Agenten verbessert, während die Agenten die routinemäßige Bearbeitung übernehmen.

Durch die Erkennung menschlicher Verhaltensmuster lernen Agenten aus menschlichen Interaktionsmustern, um erfolgreiche Kommunikationsansätze widerzuspiegeln. Dies hilft ihnen, sich an organisatorische Entscheidungsmuster und Risikotoleranzen anzupassen. Systeme verinnerlichen Qualitätserwartungen durch menschliche Korrekturen und Genehmigungen. Sie können auch lernen, auf unterschiedliche Kundensegmente und Geschäftskontexte angemessen zu reagieren.

Effektive Mechanismen zur Erfassung von Feedback erfassen systematisch menschliche Änderungen und Änderungen an den Antworten der Mitarbeiter. Sie analysieren, was menschliche Prüfer an Empfehlungen von Mitarbeitern genehmigen, ablehnen oder ändern. Diese Systeme verstehen, warum in bestimmten Fällen menschliches Eingreifen erforderlich ist, und integrieren die menschliche Bewertung der Leistung der Mitarbeiter in verschiedenen Szenarien und Komplexitätsstufen. Dadurch werden ihre Fähigkeiten kontinuierlich weiterentwickelt, um sie besser an die Standards und Erwartungen der Organisation anzupassen.

Kontinuierliche Lernprozesse

Die Lernintegration in Echtzeit ermöglicht es den KI-Systemen von Agenten, menschliches Feedback zu berücksichtigen und die Reaktionen der Agenten durch dynamische Modellaktualisierung sofort zu verbessern. Diese Systeme nutzen menschliche Erkenntnisse, um neue Muster und Grenzfälle zu identifizieren. Dies verbessert ihre Fähigkeiten zur Mustererkennung und verbessert gleichzeitig das organisatorische Gedächtnis durch von Menschen geleitete Lernerfahrungen. Kontinuierliche Verbesserungen auf der Grundlage von Feedback und Geschäftsergebnissen durch menschliche Bediener sorgen für eine kontinuierliche Leistungsoptimierung.

Von Menschen geleitete Schulungen vermitteln Expertenwissen, um die Entscheidungsfähigkeit der Mitarbeiter zu verbessern. Es überträgt wichtiges Fachwissen von erfahrenen Bedienern auf das KI-System. Durch szenariobasiertes Lernen nutzen Systeme von Menschen geschaffene Beispiele, um ihren Umgang mit komplexen Situationen zu verbessern. Durch Qualitätskalibrierung stimmen sie auch die Leistungsstandards der Agenten mit den menschlichen Qualitätserwartungen ab. Dieser Ansatz beinhaltet menschliche Erkenntnisse über die Unternehmenskultur und die Kundenerwartungen. Diese kulturelle Anpassung hilft den Mitarbeitern, in verschiedenen Kontexten angemessen zu reagieren.

Operative Exzellenz durch Zusammenarbeit zwischen Mensch und KI

Die automatisierte risikobewusste Optimierung ermöglicht eine kontinuierliche Bewertung der Betriebsbedingungen und der Fehlerwahrscheinlichkeit unter menschlicher Aufsicht für Szenarien mit hohem Risiko. Dies hilft Systemen, aus menschlichen Risikobeurteilungen zu lernen und future Entscheidungen zu verbessern. [Amazon Bedrock](#) bietet Zugriff auf mehrere Fundamentmodelle mit unterschiedlichen Funktionen und Kostenprofilen. Dies ermöglicht intelligentes Routing, das sowohl Kosten- als auch Risikoprofile berücksichtigt und gleichzeitig menschliches Feedback berücksichtigt, um die Modellauswahl zu optimieren. Bei der Leistungsoptimierung wird ein Gleichgewicht zwischen Effizienz und Minimierung der Fehlerquote hergestellt, indem menschliches Feedback zu Qualitätsstandards und akzeptablen Leistungsabstrichen berücksichtigt wird. Automatisierte Entscheidungen berücksichtigen die risikobereinigten Gesamtbetriebskosten. Die Betreiber geben Hinweise zur organisatorischen Risikotoleranz und zur Gewichtung der Geschäftsprioritäten. Dies hilft Ihnen, die Kosten zu optimieren und gleichzeitig die Unternehmensziele in Einklang zu bringen.

Von Menschen unterstützte Lernsysteme priorisieren menschliche Eingaben anhand der Auswirkungen von Fehlern und der Auswirkungen auf das Unternehmen. Dadurch entstehen Lernsysteme, die durch risikogewichtetes Feedback sowohl die technische Genauigkeit als auch den geschäftlichen Kontext verstehen. Regelmäßige Leistungsanalysen umfassen Risikometriken und Fehlerkostenanalysen, wobei menschliche Erkenntnisse einen Kontext liefern, den automatisierte Systeme nicht erfassen können. Bei der Entwicklung bewährter Verfahren liegt der Schwerpunkt auf Risikomanagement und Fehlerprävention, indem automatisierte Mustererkennung mit menschlichem Fachwissen und Urteilsvermögen kombiniert wird. Durch den Aufbau organisatorischer Fähigkeiten durch Schulungsprogramme werden sowohl menschliche Fähigkeiten für die Verwaltung agentischer KI-Systeme als auch Fähigkeiten der Agenten zur Unterstützung menschlicher Entscheidungsfindung entwickelt. Dies gewährleistet einen umfassenden Ansatz für die Zusammenarbeit zwischen Mensch und KI, der beide Komponenten der Partnerschaft stärkt.

Wirtschaftliche Transformation hin zu ergebnisorientierter Preisgestaltung für agentische KI-Systeme am AWS

Die Umstellung von traditionellen Festkostenmodellen auf ergebnisorientierte Preisgestaltung stellt einen grundlegenden Wandel in der Art und Weise dar, wie Unternehmen ihre wirtschaftlichen Abläufe strukturieren und Risiken managen. Diese Transformation dient als Weg für die ständige Modernisierung vorhandener Prozesse und finanziert gleichzeitig die agentische KI-Transformation. Sie ermöglicht es Unternehmen, sich von statischen, ressourcenintensiven Abläufen hin zu dynamischen, ergebnisorientierten Geschäftsmodellen zu entwickeln.

Herkömmliches Modell, das direkt im Voraus erfolgt

Abteilungen agieren häufig als Kostenstellen mit direkten Personalkosten, die durch die Kostenzuweisung finanziert werden. Organizations möchten diese Kostenverteilung in der Regel reduzieren. Wenn der Prozess nicht modernisiert wird, muss die Abteilung dieselben Ergebnisse mit einer kleineren Belegschaft erzielen. Dies beeinträchtigt in der Regel die Qualität. Traditionelle Geschäftsmodelle bringen erhebliche Herausforderungen mit sich, darunter:

- Lineare Kostenskalierung bei steigendem Volumen — Dies erfordert, dass Unternehmen zusätzliches Personal einstellen, um das erhöhte Volumen bewältigen zu können.
- Verpflichtungen zu festen Kosten — Diese bleiben unabhängig von der Geschäftsleistung und Prozesseffizienz bestehen.
- Vorausschauende Planung — Eingeschränkte Flexibilität in Zeiten wirtschaftlicher Abschwünge und Kapazitätsengpässe erfordert eine vorausschauende Planung.
- Zyklus der Qualitätsverschlechterung — Kürzere Budgets führen zu einer verminderten Servicequalität, wenn die Kosten ohne Prozessverbesserungen gesenkt werden.

Ergebnisorientiertes Modell

Moderne ergebnisorientierte Modelle verknüpfen Zahlungen direkt mit messbaren Geschäftsergebnissen, wie z. B. erfolgreichen Neueinstellungen, erreichten Qualitätskennzahlen, Verbesserungen der Prozesseffizienz oder realisierten Produktivitätssteigerungen. Dadurch wird das finanzielle Risiko grundlegend von den Geschäftsbereichen auf die Dienstleister verlagert und gleichzeitig eine natürliche Anreizausrichtung erreicht. Im Folgenden sind die wichtigsten Vorteile eines ergebnisorientierten Modells aufgeführt:

- Die Kosten skalieren direkt mit dem generierten Geschäftswert
- Natürliche Abstimmung zwischen Betriebskosten und Einnahmen
- Flexibilität bei der Anpassung der Kapazität an die Marktbedingungen
- Pay-per-success Modelle reduzieren das finanzielle Risiko, indem sie das finanzielle Risiko von den Anfangsinvestitionen auf die laufende betriebliche Leistung verlagern
- Konzentrieren Sie sich auf lernfähige Systeme, die sich im Laufe der Zeit verbessern, und nicht auf statische Alternativen

Diese Transformation geht weit über interne Kostenstellen hinaus und verändert die Art und Weise, wie Unternehmen mit externen Partnern und Dienstleistern umgehen, grundlegend. Durch die Anwendung ergebnisorientierter Preisgestaltung auf die Zusammenarbeit mit Partnern können Unternehmen langfristige Qualitätsverbesserungen vorantreiben und Kosten senken, während sie gleichzeitig indirekt den Schwerpunkt auf die Modernisierung der KI legen.

Organizations können schnell experimentieren, die Leistung klar messen und die Skalierung auf der Grundlage des tatsächlich generierten Geschäftswerts und nicht auf der Grundlage herkömmlicher fester Ressourcenverpflichtungen vornehmen. Dieser Ansatz ermöglicht Folgendes:

- Weiterentwicklung der Lieferantenbeziehungen — Partner investieren in den Kundenerfolg und nicht nur in die Erbringung von Dienstleistungen.
- Standardisierte Ergebniskennzahlen — Vereinfachen Sie die Beschaffungsprozesse bei mehreren Anbietern.
- Reaktionsfähigkeit auf den Markt — Passen Sie sich schnell an sich ändernde Marktbedingungen und Kundenbedürfnisse an.
- Wettbewerbsvorteil — Überlegene Ressourcennutzung und verbesserte Betriebsfähigkeiten.
- Qualitätsorientierte Partnerschaften — Die langfristige Zusammenarbeit konzentriert sich auf kontinuierliche Verbesserungen und messbare Ergebnisse.

Verwendung AWS Marketplace als Enabler pay-per-outcome

Die wichtigste Voraussetzung für diese Transformation ist [AWS Marketplace](#), dass sie als Transaktionsvehikel für Agenturarbeit und ergebnisorientierte Preisgestaltung dient. Es bietet Zugriff auf Hunderte von vorgefertigten KI-Agenten und Agentenlösungen mit transparenten, nutzungsbasierten Preismodellen. Es kann dazu beitragen, vorab anfallende Lizenzkosten zu vermeiden, die Komplexität der Implementierung zu reduzieren und es Unternehmen zu ermöglichen, sich auf lernfähige Systeme zu konzentrieren, die sich im Laufe der Zeit anpassen und verbessern, anstatt sich auf statische Alternativen zu konzentrieren

Die Verwendung AWS Marketplace kann die folgenden Vorteile bieten:

- Schnelles Experimentieren — Testen Sie mehrere Lösungen ohne nennenswerte Kapitalinvestitionen
- Transparente Preisgestaltung — nutzungsabhängige Kosten mit klarer Zuordnung zu den Geschäftsergebnissen
- Bewährte Lösungen — Zugang zu praxiserprobten Agenten von erfahrenen Anbietern

- Integrierte Integration — Nahtlose Konnektivität mit bestehenden AWS-Services
- Risikominderung — Möglichkeit, je nach Leistung den Anbieter zu wechseln
- Zugang zu Lernfähigkeiten — Verfügbarkeit adaptiver Systeme ohne interne Entwicklungskosten

Dieser Ansatz ermöglicht es Unternehmen, mehrere Optionen auf der Grundlage von Ergebnissen und Lernfähigkeiten zu vergleichen, anstatt anhand von Feature-Listen. Es kann Ihnen auch dabei helfen, klare Erfolgskriterien und Messmethoden festzulegen und ergebnisorientierte Preise auszuhandeln, die an Geschäftsergebnisse und Systemverbesserungen geknüpft sind. Durch die Finanzierung der behördlichen KI-Transformation durch ergebnisorientierte Modelle können Unternehmen ihre Prozesse kontinuierlich modernisieren und zahlen dabei nur für messbare Verbesserungen und erfolgreiche Ergebnisse.

Fallstudie: Vergleich der Kosten für menschliche und behördliche KI bei Personalbeschaffungsmaßnahmen

Personalbeschaffungsmaßnahmen bieten eine überzeugende Fallstudie zur Bewertung der wirtschaftlichen Kompromisse zwischen menschlichen und agentischen KI-Systemen. Die Berechnung des ROI hängt jedoch entscheidend von Ihrer aktuellen betrieblichen Ausgangsbasis ab. Organizations, die KI-Investitionen von Agenturen bewerten, stellen sich oft eine grundlegende Frage: „Was wäre, wenn wir stattdessen einfach unsere bestehenden menschlichen Prozesse optimieren würden?“ Um dieses Problem direkt anzugehen, werden in dieser Analyse zwei unterschiedliche Szenarien vorgestellt, die den Bereich der betrieblichen Effizienz der Mitarbeiter einbeziehen.

[Szenario A modelliert](#) die Dauer der Prüfung von Lebensläufen (CV) oder Lebensläufen von 45 Minuten. [Szenario B](#) zeigt optimierte menschliche Abläufe innerhalb von 15 Minuten pro Anwendung, was einer Effizienzsteigerung von 66% entspricht. Diese Verbesserung könnte beispielsweise durch optimierte Prozesse, erfahrene Personalvermittler oder spezielle Tools erreicht werden.

Durch den Vergleich identischer Agentensystemfunktionen mit diesen unterschiedlichen Ausgangswerten für die menschliche Leistung zeigen wir auf, wie sich die bestehende Prozesseffizienz auf ROI-Berechnungen, Break-Even-Zeitpläne und strategische Implementierungsentscheidungen auswirkt. Dieser Ansatz mit zwei Szenarien dient mehreren Zwecken. Es verhindert, dass Unternehmen künstliche Intelligenz von Agenturen abtun, indem sie davon ausgehen, dass Prozessoptimierung allein ausreichend ist. Es hilft auch Unternehmen mit bereits effizienten Prozessen, ihre spezifischen wirtschaftlichen Aspekte zu verstehen. Darüber hinaus zeigen diese Szenarien, wann nichtfinanzielle Vorteile wie Verfügbarkeit und Skalierbarkeit rund um die Uhr zu den wichtigsten Entscheidungsfaktoren werden. Das Verständnis dieser wirtschaftlichen Dynamik über verschiedene Effizienzgrundlagen hinweg ermöglicht es Unternehmen, fundierte Entscheidungen darüber zu treffen, wo und wann sie agentische KI-Systeme einsetzen sollten, um maximale geschäftliche Auswirkungen zu erzielen.

Szenario A: 45 Minuten Screening-Zeit

Szenario A stellt Personalbeschaffungsmaßnahmen dar, bei denen Personalvermittler 45 Minuten damit verbringen, jeden Lebenslauf zu überprüfen. In diesem Szenario wird ein Personalvermittler auf mittlerer Ebene mit jährlichen Vollkosten von 112.250\$ modelliert. Dieser Personalvermittler bearbeitet Bewerbungen während der üblichen Geschäftszeiten mit typischen menschlichen Leistungsmerkmalen. Im Gegensatz dazu erfordert das KI-System von Agentice

eine Anfangsinvestition von 23.000 USD für Entwicklung, Anpassung und ATS-Integration und hat minimale monatliche Betriebskosten von 500 USD für die Cloud-Infrastruktur. Der Agent bearbeitet Anwendungen in nur 5 Minuten und ist rund um die Uhr verfügbar. Er erreicht eine Fehlerquote von 2% und eine monatliche Kapazität von mehr als 8.600 Anwendungen. Dies ist eine dramatische Effizienzlücke, da der Agent pro Anwendung neunmal schneller arbeitet und die monatliche Kapazität 39-mal höher ist. In diesem Abschnitt werden Kostenstrukturanalysen, betriebliche Kennzahlen, volumenbasierte Vergleiche und kumulative ROI-Berechnungen in den ersten sechs Betriebsmonaten untersucht.

Grundkostenstruktur

Die folgende Tabelle zeigt die anfänglichen Einrichtungskosten für Szenario A.

Komponente	Menschliche Operationen	Agentisches KI-System
Entwicklung und Anpassung von Agenten	–	15.000\$
Integration des Bewerbererfolgungssystems (ATS)	–	5.000\$
Schulung und Optimierung	–	3.000\$
Gesamtkosten für die Ersteinrichtung	0\$	23.000\$

Die folgende Tabelle zeigt die jährlichen Fixkosten für Szenario A.

Komponente	Menschliche Operationen	Agentisches KI-System
Grundgehalt	65.000\$	–
Leistungen (30%)	19.500\$	–
Arbeitsbereich und Ausrüstung	12.000\$	–
Aufsicht durch das Management (15%)	9.750\$	–

Schulung und Entwicklung	6.000\$	–
Jährliche Fixkosten insgesamt	112.250\$	N/A

Die folgende Tabelle zeigt die monatlichen Betriebskosten für Szenario A.

Komponente	Menschliche Operationen	Agentisches KI-System
Cloud-Computing	–	200 USD
Speicher	–	100 USD
Datenbankvorgänge	–	100 USD
Überwachen	–	100 USD
Monatliche Fixkosten insgesamt	9.354\$	500\$

Operationelle Metriken

Die folgende Tabelle zeigt Betriebskennzahlen für Szenario A.

Metrik	Menschliche Abläufe	Agentisches KI-System
Bearbeitungszeit pro Antrag	45 Minuten	5 Minuten
Kapazität pro Stunde	1.33 Anwendungen	12 Anwendungen
Tageskapazität (24 Stunden)	10-11 Anwendungen	288 Anwendungen
Monatliche Kapazität	220 Anwendungen	8.640 Bewerbungen
Kosten pro Anwendung	45\$	2,50\$
Kosten pro erfolgreicher Einstellung	2.200\$	\$125
Fehlerquote	5 %	2%

Kosten der Fehlerkorrektur 90\$ pro Fehler 45\$ pro Eskalation

Volumenbasierte Kostenanalyse

Die folgende Tabelle zeigt eine volumenbasierte Kostenanalyse für Szenario A. In diesem Beispiel beinhalten die Kosten des KI-Systems für Agenturen Fixkosten und amortisierte Einrichtungskosten in Höhe von 1.917\$ pro Monat über einen Zeitraum von 12 Monaten.

Monatliches Volumen	Menschliche Kosten	Kosten für Agentur-KI-Systeme	Monatliche Einsparungen
100 Anwendungen	4.500\$	750\$	3.750\$
500 Bewerbungen	22.500\$	2.667\$	19.833\$
1.000 Bewerbungen	45.000\$	4.917\$	40.083\$

ROI-Analyse

Die folgende Tabelle zeigt eine ROI-Analyse für Szenario A, die auf der Verarbeitung von 500 Anträgen pro Monat basiert.

Metrik	Wert
Monatliche Personalkosten	22.500\$
Monatliche Kosten für Makler	2.667\$
Monatliche Ersparnisse	19.833\$
Jährliche Einsparungen	237.996\$
Phase der Gewinnschwelle	1,16 Monate

Vergleich der kumulierten Kosten

Die folgende Tabelle zeigt einen Vergleich der kumulierten Kosten für Szenario A für die ersten sechs Monate, wobei von 500 Anträgen pro Monat ausgegangen wird.

Monat	Menschliche Kosten	Kosten für Agentur-KI-Systeme	Kumulierte Einsparungen
1	22.500\$	25.667\$	-3.167 \$
2	45.000\$	28.334\$	16.666\$
3	67.500\$	31.001\$	36.499\$
4	90.000\$	33.668\$	56.332\$
5	112.500\$	36.335\$	76.165\$
6	135.000\$	39.002\$	95.998\$

Zusätzliche Vorteile des agentischen KI-Systems

Im Folgenden sind die zusätzlichen Vorteile aufgeführt, die das agentische KI-System in Szenario A bietet:

- Skalierbarkeit — Kann Volumenspitzen ohne zusätzliche Kosten bewältigen
- Verfügbarkeit — Betrieb rund um die Uhr mit sofortiger Reaktion
- Konsistenz — Anwendung einheitlicher Screening-Kriterien
- Zeiteffizienz — Deutlich reduziert time-to-hire
- Benutzererfahrung — Sofortiges Feedback an die Kandidaten

Szenario B: 15-minütige Screening-Zeit

Szenario B modelliert optimierte Rekrutierungsabläufe, bei denen Personalvermittler ihren Auswahlprozess auf 15 Minuten pro Bewerbung reduziert haben. Dies entspricht einer Effizienzsteigerung von 66% gegenüber Szenario A. Bei diesem Szenario werden dieselben

jährlichen Gesamtkosten von 112.250\$ für Personalvermittler auf mittlerer Ebene beibehalten. Es zeigt jedoch eine deutlich höhere Produktivität der Mitarbeiter: Die tägliche Kapazität steigt auf 32 Anwendungen während einer 8-Stunden-Schicht und der monatliche Durchsatz erreicht 660 Bewerbungen. Durch die verbesserte Effizienz der Mitarbeiter werden die Kosten pro Anwendung von 45 USD auf 15 USD gesenkt, wodurch die wirtschaftliche Kluft zum agentischen KI-System verringert wird. Der Agent behält jedoch seine strukturellen Vorteile bei: eine Bearbeitungszeit von 5 Minuten, eine Verfügbarkeit rund um die Uhr, die 288 tägliche Anwendungen ermöglicht, eine geringere Fehlerquote von 2% im Vergleich zu 5% und eine monatliche Kapazität von mehr als 8.600 Anwendungen. Diese Effizienzsteigerung verlängert zwar die Gewinnschwelle von 1,16 Monaten auf 4,76 Monate und reduziert die monatlichen Einsparungen von 19.833\$ auf 4.833\$. Die Analyse zeigt jedoch, dass Agentensysteme auch dann wirtschaftlich rentabel bleiben, wenn sie mit hochoptimierten menschlichen Abläufen konkurrieren — eine wichtige Erkenntnis für Unternehmen, die beurteilen, ob ihre aktuelle Prozesseffizienz agentische KI-Investitionen rechtfertigt.

Grundlegende Kostenstruktur

Die folgende Tabelle zeigt die jährlichen Fixkosten für Szenario B.

Komponente	Menschliche Operationen	Agentisches KI-System
Grundgehalt	65.000\$	–
Leistungen (30%)	19.500\$	–
Arbeitsraum und Ausstattung	12.000\$	–
Aufsicht durch das Management (15%)	9.750\$	–
Schulung und Entwicklung	6.000\$	–
Jährliche Fixkosten insgesamt	112.250\$	N/A

Die folgende Tabelle zeigt die Implementierungskosten für Szenario B.

Komponente	Menschliche Operationen	Agentisches KI-System
Erstes Einrichten von	–	23.000\$

Monatliche Fixkosten	9.354\$	500\$
----------------------	---------	-------

Operationelle Metriken

Die folgende Tabelle zeigt Betriebskennzahlen für Szenario B.

Metrik	Menschliche Abläufe	Agentisches KI-System
Bearbeitungszeit pro Antrag	15 Minuten	5 Minuten
Kapazität pro Stunde	4 Anwendungen	12 Anwendungen
Tageskapazität (8-Stunden-Schicht)	32 Anwendungen	288 Anwendungen
Monatliche Kapazität	660 Anwendungen	8.640 Bewerbungen
Kosten pro Anwendung	15\$	2,50\$
Kosten pro erfolgreicher Einstellung	2.200\$	\$125
Fehlerquote	5 %	2%
Kosten der Fehlerkorrektur	30\$ pro Fehler	45\$ pro Eskalation

Volumenbasierte Kostenanalyse

Die folgende Tabelle zeigt eine volumenbasierte Kostenanalyse für Szenario B. In diesem Beispiel beinhalten die Kosten des KI-Systems für Agenturen Fixkosten und amortisierte Einrichtungskosten in Höhe von 1.917\$ pro Monat über einen Zeitraum von 12 Monaten.

Monatliches Volumen	Menschliche Kosten	Kosten für Agentur-KI-Systeme	Monatliche Einsparungen
100 Anwendungen	1.500\$	750\$	750\$
500 Bewerbungen	7.500\$	2.667\$	4.833\$

1.000 Bewerbungen	15.000\$	4.917\$	10.083\$
-------------------	----------	---------	----------

ROI-Analyse

Die folgende Tabelle zeigt eine ROI-Analyse für Szenario B, die auf der Verarbeitung von 500 Anträgen pro Monat basiert.

Metrik	Wert
Monatliche Personalkosten	7.500\$
Monatliche Kosten für das KI-System von Agenturen	2.667\$
Monatliche Ersparnisse	4.833\$
Jährliche Einsparungen	57.996\$
Phase der Gewinnschwelle	4,76 Monate

Vergleich der kumulierten Kosten

Die folgende Tabelle zeigt einen Vergleich der kumulierten Kosten für Szenario B für die ersten sechs Monate, wobei von 500 Anträgen pro Monat ausgegangen wird.

Monat	Menschliche Kosten	Kosten für Agentur-KI-Systeme	Kumulierte Einsparungen
1	7.500\$	25.667\$	-18.167 \$
2	15.000\$	28.334\$	-13.334 \$
3	22.500\$	31.001\$	-8.501 \$
4	30.000\$	33.668\$	-3.668 \$
5	37.500\$	36.335\$	1.165\$

6	45.000\$	39.002\$	5.998\$
---	----------	----------	---------

Vergleich von Kosten und Nutzen für jedes Szenario

Metrik	Szenario A	Szenario B	Auswirkung
Zeit der Vorführung	45 Minuten	15 Minuten	Verbesserung um 66%
Tägliche Kapazität	10—11 Anwendungen	32 Anwendungen	Steigerung um 200%
Kosten pro Anwendung	45\$	15\$	Ermäßigung um 66%
Monatliche Einsparungen (500 Anwendungen)	19.833\$	4.833\$	Rückgang um 76%
Phase der Gewinnschwelle	1,16 Monate	4,76 Monate	310% länger

Szenario B zeigt deutliche Effizienzsteigerungen bei menschlichen Abläufen mit Verbesserungen der Verarbeitungszeit, wodurch die Kapazität ohne zusätzlichen Personalbestand erhöht und die Kosten pro Anwendung erheblich gesenkt werden. Die finanziellen Auswirkungen zeigen jedoch ein differenzierteres Bild: Während der ROI weiterhin positiv ist, müssen Unternehmen im Vergleich zu Szenario A mit einer längeren Gewinnschwelle und geringeren monatlichen Einsparungen rechnen. Diese Ergebnisse heben wichtige Entscheidungsfaktoren für die Implementierung hervor — das Agentensystem bleibt auch bei optimierten Personalabläufen finanziell tragfähig, aber Unternehmen müssen bei der Bewertung von Bereitstellungszeitplänen und erwarteten Renditen eine längerfristige Investitionsperspektive verfolgen und Volumenschwankungen und Skalierbarkeitserfordernisse sorgfältig berücksichtigen.

Das agentische KI-System bietet jedoch immer noch wichtige betriebliche Vorteile, die über reine Kosteneinsparungen hinausgehen. Es bietet eine Verfügbarkeit rund um die Uhr für eine sofortige Kandidatenbindung, unabhängig von Zeitzonen oder Geschäftszeiten. Es bietet eine gleichbleibende Screening-Qualität, indem es für jede Bewerbung einheitliche Kriterien anwendet und skaliert, um

Volumenspitzen zu bewältigen, ohne dass zusätzliche Kosten anfallen. Es bietet eine sofortige Antwort auf die Bewerber, was die Arbeitgebermarke und das Bewerbererlebnis verbessert, und es arbeitet ohne Ermüdungsfaktor, sodass bei der ersten Bewerbung die gleiche hohe Qualität gewährleistet ist wie bei der tausendsten Bewerbung.

Menschliche Fehler sind in der Regel auf Müdigkeit, Ablenkung oder Wissenslücken zurückzuführen und führen häufig zu Fehlkommunikation oder falschen Informationen. Agentenfehler in KI-Systemen sind in der Regel auf Randfälle, mehrdeutige Eingaben oder Einschränkungen bei den Trainingsdaten zurückzuführen. Diese Fehler sind tendenziell konsistenter Natur.

Qualitäts- und Erfahrungskennzahlen lassen klare Kompromisse zwischen den Fähigkeiten von Mitarbeitern und Mitarbeitern erkennen:

- Kundenzufriedenheit — Menschen zeichnen sich durch Empathie und komplexe Problemlösung aus, und Mitarbeiter stellen konsistente, genaue Informationen für Routineanfragen bereit.
- Reaktionszeit — Die Reaktionszeit begünstigt Agenten mit sofortiger Verfügbarkeit rund um die Uhr. Mitarbeiter bieten Support während der Geschäftszeiten, was zu Verzögerungen in der Warteschlange führen kann.
- Konsistenz — Agenten geben identische Antworten auf ähnliche Anfragen. Menschen können in ihrer Herangehensweise und in der Anwendung von Wissen unterschiedlich sein.
- Umgang mit Eskalationen — Komplexe Probleme, die Urteilsvermögen, Kreativität oder emotionale Intelligenz erfordern, bleiben menschliche Stärken.

Fazit und Ressourcen

Die Wirtschaftlichkeit menschlicher Systeme im Vergleich zu agentischen KI-Systemen ist mehr als eine technologische Entscheidung. Sie spiegelt einen grundlegenden Wandel in der Art und Weise wider, wie Unternehmen Wert schaffen, Risiken managen und Wettbewerbsvorteile erzielen. Erfolg erfordert eine systematische Bewertung der Berufsmerkmale, eine umfassende Messung der Ergebnisse (einschließlich Risikofaktoren) und eine strategische Skalierung auf der Grundlage nachgewiesener Ergebnisse.

[Der Stand der Einführung von KI in Unternehmen](#) (ISG-Bericht 2025) zeigt, dass die meisten KI-Implementierungen aufgrund von Lernlücken scheitern — Systeme, die sich nicht anpassen, sich nicht an den Kontext erinnern oder sich im Laufe der Zeit nicht verbessern können. Erfolgreiche Unternehmen konzentrieren sich auf lernfähige Systeme, die sich tief in Arbeitsabläufe integrieren lassen und durch menschliches Feedback und Betriebserfahrung kontinuierliche Verbesserungen nachweisen.

Organizations, die diese Prinzipien verstehen — angefangen mit geeigneten Jobs, Zerlegung von Jobs in Aufgaben, Messung aller Aspekte, einschließlich der Risikoauswirkungen, und Skalierung dessen, was funktioniert — werden durch optimale Ressourcennutzung und ergebnisorientierte Automatisierung, die mit dem Geschäftserfolg wächst, einen nachhaltigen Wettbewerbsvorteil erzielen.

Die future gehört Organisationen, die menschliches Fachwissen auf intelligente Weise mit agentischen KI-Fähigkeiten kombinieren können. Dadurch entstehen Hybridmodelle, die hervorragende Ergebnisse liefern und gleichzeitig die Flexibilität, Lernfähigkeit und die Vorteile der Zusammenarbeit beibehalten, die für dynamische Marktbedingungen erforderlich sind.

Ressourcen

Die folgenden Ressourcen können Ihnen bei der Planung, Gestaltung und Implementierung agentischer KI-Systeme helfen für: AWS

- [Aufbau serverloser Architekturen für agentic AI on \(Prescriptive Guidance\) AWS](#)
- [Operationalisierung der behördlichen KI auf \(Prescriptive Guidance\) AWS](#)
- [Behördliche KI-Muster und Arbeitsabläufe auf \(Prescriptive Guidance\) AWS](#)
- [Agentische KI \(Prescriptive Guidance\) AWS](#)

- [AWS Cost Optimization Hub](#) (AWS-Service)
- [Amazon Bedrock-Dokumentation](#) (AWS-Service)
- [Säule Kostenoptimierung](#) (AWS Well-Architected Framework)
- [KI-Agenten und -Lösungen](#) (AWS Marketplace)

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	—	28. Januar 2026

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 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 stressen, 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 oder Bitbucket 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 unter [Kontinuierliche 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 Kerngeschäftsziele 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 Transaktionsdaten. 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.

|

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) Ilo

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 Infrastrukturmgebung 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-Entitäts 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 `false` zurückgibt, 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.

Refaktorisierung

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 unbekanntere 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.