Agentische KI-Frameworks, -Plattformen, Protokolle und Tools auf AWS

AWS Präskriptive Leitlinien

AWS Präskriptive Leitlinien: Agentische KI-Frameworks, -Plattformen, Protokolle und Tools auf AWS

Table of Contents

Agentische KI-Frameworks, -Plattformen, Protokolle und Tools auf AWS

Aaron Sempf, Ansley Verzosa und Joshua Samuel, Amazon Web Services ()AWS

Januar 2026 (Geschichte der Dokumente)

Agentic AI ist ein leistungsstarkes Paradigma an der Schnittstelle von KI, verteilten Systemen und Softwareentwicklung. Es handelt sich um eine Klasse intelligenter Systeme, die aus autonomen, asynchronen Softwareagenten bestehen, die KI-Modelle verwenden und sich in Tools und Ressourcen integrieren lassen. Die Agenten zeigen Entscheidungsfreiheit, können den Kontext wahrnehmen, über Ziele nachdenken, Entscheidungen treffen und zielgerichtete Maßnahmen im Namen von Benutzern oder Systemen ergreifen. Diese Agenten arbeiten unabhängig, oft kollaborativ, in verteilten Umgebungen und sind darauf ausgelegt, delegierte Ziele mit integrierter Intelligenz, Gedächtnis und Absicht zu verfolgen.

Auf diese Weise können Unternehmen die KI von Agenten nutzen AWS, um komplexe Arbeitsabläufe zu automatisieren, Entscheidungsprozesse zu verbessern und reaktionsfähigere Systeme zu entwickeln. Dieser Leitfaden enthält Informationen zu den wichtigsten Komponenten, die für die Entwicklung effektiver KI-Lösungen für Agenturen erforderlich sind:

- Frameworks stellt aktuelle KI-Frameworks für Agenturen vor, einschließlich einer Übersicht über ihre Vorteile und Anwendungsfälle. Erfahren Sie, wie diese Frameworks undifferenzierte Schwerstarbeit bei Mustern, Protokollen und Tools reduzieren. Machen Sie sich mit den wichtigsten Auswahlkriterien vertraut, um das richtige Framework für Ihre Anforderungen auszuwählen.
- Platforms bietet einen Überblick über die wichtigsten KI-Plattformen (Managed Agent, Open-Source-Orchestrierung und Hybrid) sowie Überlegungen zur Auswahl oder zum Design.
- Protocols untersucht wichtige standardisierte Kommunikationsprotokolle für Agenteninteraktionen. Agent-to-agentProtokolle wie das Open-Source-Model Context Protocol (MCP) und Agent2Agent (A2A) sowie andere proprietäre Implementierungen sind im Entstehen begriffen. Erfahren Sie, wie gängige Protokolle die nahtlose Interaktion verschiedener Protokolle ermöglichen.
- Tools bietet Informationen über protokollbasierte Tools (wie das MCP), Framework-native Tools und Meta-Tools. Organizations können ein Toolkit erstellen, das sich in die wichtigsten Systeme ihrer Workflows integrieren lässt und sowohl Endbenutzer- als auch serverbasierte Agentenworkflows ermöglicht.

Zielgruppe

Dieser Leitfaden richtet sich an Architekten, Entwickler und Technologieführer, die das Potenzial KI-gestützter Softwareagenten in modernen Cloud-nativen Anwendungen nutzen möchten.

Ziele

Dieser Leitfaden hilft Ihnen bei folgenden Aufgaben:

- Vergleichen Sie verschiedene agentische KI-Frameworks, um das für Ihren Anwendungsfall am besten geeignete auszuwählen.
- Erfahren Sie mehr über agentische KI-Plattformen, die Funktionen bieten, um einzelne Agenten in koordinierte, anpassungsfähige Systeme umzuwandeln.
- Erfahren Sie mehr über die Vorteile offener Protokolle für den Aufbau nachhaltiger agentischer KI-Architekturen.
- Entwickeln Sie beim Aufbau von Agentensystemen eine geeignete Strategie zur Integration von Tools.

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

Frameworks

Foundations of Agentic AI on AWS untersucht die Kernmuster und Arbeitsabläufe, die autonomes, zielgerichtetes Verhalten ermöglichen. Im Mittelpunkt der Implementierung dieser Muster steht die Wahl des Frameworks. Ein Framework ist die Softwarebasis aus vordefiniertem Code, die eine strukturierte Umgebung und gemeinsame Funktionen für die Erstellung und Verwaltung von Tools und Orchestrierungsfunktionen bietet, die für die Entwicklung produktionsreifer autonomer KI-Agenten erforderlich sind.

Effektive agentische KI-Frameworks bieten mehrere grundlegende Funktionen, die unbearbeitete Interaktionen mit einem Large Language Model (LLM) in koordinierte, intelligente Systeme umwandeln, die in der Lage sind, zu denken, zusammenzuarbeiten und zu handeln:

- Die Agentenorchestrierung koordiniert den Informationsfluss und die Entscheidungsfindung zwischen einzelnen oder mehreren Agenten, um komplexe Ziele ohne menschliches Eingreifen zu erreichen.
- Die Toolintegration ermöglicht es Agenten, mit externen Systemen und Datenquellen zu interagieren APIs, um ihre Fähigkeiten über die Sprachverarbeitung hinaus zu erweitern. Weitere Informationen finden Sie in der Strands Agents Dokumentation unter Tool-Übersicht.
- Die Speicherverwaltung bietet einen dauerhaften oder sitzungsbasierten Status, um den Kontext zwischen Interaktionen aufrechtzuerhalten, was für lang andauernde oder adaptive Aufgaben unerlässlich ist. Fortgeschrittenere Frameworks verfügen über ein Langzeitgedächtnis zum Speichern von Zusammenfassungen und Benutzereinstellungen und ermöglichen so personalisierte und kontextsensitive Agentenerlebnisse. Weitere Informationen finden Sie im Blog unter How to think about Agent Frameworks. LangChain
- Die Workflow-Definition unterstützt strukturierte Muster wie Ketten, Routing, Parallelisierung und Reflexionsschleifen, die ein ausgeklügeltes autonomes Denken ermöglichen.
- Einsatz und Überwachung erleichtern den Übergang von der Entwicklung zur Produktion, wobei autonome Systeme beobachtbar sind. Weitere Informationen finden Sie in der Ankündigung zur AgentCore allgemeinen Verfügbarkeit von Amazon Bedrock.

Diese Funktionen werden in der gesamten Framework-Landschaft mit unterschiedlichen Ansätzen und Schwerpunkten implementiert und bieten jeweils unterschiedliche Vorteile für unterschiedliche Anwendungsfälle autonomer Agenten und Unternehmenskontexte.

In diesem Abschnitt werden die führenden Frameworks für die Entwicklung agentischer KI-Lösungen vorgestellt und verglichen, wobei der Schwerpunkt auf ihren Stärken, Grenzen und idealen Anwendungsfällen für den autonomen Betrieb liegt:

- Strands & Agenten
- LangChain und LangGraph
- Crew I
- AutoGen
- ???
- Vergleich agentischer KI-Frameworks

> **Note**
>
> In diesem Abschnitt werden die Frameworks behandelt, die speziell die Steuerung der KI unterstützen. Frontend-Schnittstellen oder generative KI ohne Agentur werden nicht behandelt.

Strands Agents

Strands Agentsist ein Open-Source-SDK, das ursprünglich von veröffentlicht wurde AWS, wie im AWS Open Source-Blog beschrieben. Strands Agentswurde für die Entwicklung autonomer KI-Agenten mit einem Model-First-Ansatz entwickelt. Es bietet ein flexibles, erweiterbares Framework, das so konzipiert ist, dass es nahtlos funktioniert und AWS-Services gleichzeitig offen für die Integration mit Komponenten von Drittanbietern bleibt. Strands Agents ist ideal für den Aufbau vollständig autonomer Lösungen.

Hauptmerkmale von Strands Agents

Strands Agentsumfasst die folgenden Hauptfunktionen:

- Modellorientiertes Design — Es basiert auf dem Konzept, dass das Basismodell den Kern der Agentenintelligenz bildet und ausgeklügeltes autonomes Denken ermöglicht. Weitere Informationen finden Sie in der Dokumentation unter Agent Loop. Strands Agents
- Muster für die Zusammenarbeit mehrerer Agenten — Integrierte Koordinationsmodelle wie Swarm-, Graph- und Workflow-Muster, die eine skalierbare Zusammenarbeit und Steuerung über verteilte

Agentennetzwerke hinweg ermöglichen. Weitere Informationen finden Sie unter Muster für mehrere Agenten in der Dokumentation zu Strands Agents.

- MCP-Integration — Systemeigene Unterstützung für das Model Context Protocol (MCP), wodurch eine standardisierte Bereitstellung von Kontext LLMs für einen konsistenten autonomen Betrieb ermöglicht wird.
- AWS-Service Integration — Nahtlose Verbindung zu Amazon Bedrock, AWS Lambda AWS Step Functions, und anderen AWS-Services für umfassende autonome Workflows. Weitere Informationen finden Sie unter AWS Weekly Roundup (Blog)AWS .
- Auswahl des Foundation-Modells — Unterstützt verschiedene Foundation-Modelle, darunter Anthropic Claude, Amazon Nova (Premier, Pro, Lite und Micro) auf Amazon Bedrock und andere, um für verschiedene Funktionen des autonomen Denkens zu optimieren. Weitere Informationen finden Sie in der Strands Agents Dokumentation unter Amazon Bedrock.
- LLM-API-Integration — Flexible Integration mit verschiedenen LLM-Serviceschnittstellen wie Amazon Bedrock, OpenAI und anderen für den Produktionseinsatz. Weitere Informationen finden Sie in der Strands Agents Dokumentation unter Amazon Bedrock Basic Usage.
- Multimodale Funktionen — Support mehrerer Modalitäten, einschließlich Text-, Sprach- und Bildverarbeitung für umfassende autonome Agenteninteraktionen. Weitere Informationen finden Sie in der Dokumentation unter Amazon Bedrock Multimodal Support. Strands Agents
- Tool-Ökosystem — Umfangreiches Angebot an Tools für die AWS-Service Interaktion mit Erweiterbarkeit für benutzerdefinierte Tools, die die autonomen Funktionen erweitern. Weitere Informationen finden Sie in der Strands Agents Dokumentation unter Tool-Übersicht.

Wann sollte es verwendet werden Strands Agents

Strands Agentseignet sich besonders gut für Szenarien mit autonomen Agenten, darunter:

- Organizations, die auf einer AWS Infrastruktur aufbauen und eine native Integration AWS-Services für autonome Workflows wünschen
- Teams, die Sicherheits-, Skalierbarkeits- und Compliance-Funktionen auf Unternehmensebene für autonome Produktionssysteme benötigen
- Projekte, die Flexibilität bei der Modellauswahl verschiedener Anbieter für spezielle autonome Aufgaben benötigen
- Anwendungsfälle, die eine enge Integration mit bestehenden AWS Workflows und Ressourcen für durchgängige autonome Prozesse erfordern

Implementierungsansatz für Strands Agents

Strands Agentsbietet einen unkomplizierten Implementierungsansatz für Unternehmensbeteiligte, wie im Schnellstartleitfaden beschrieben. Das Framework ermöglicht Organisationen:

- Wählen Sie Basismodelle wie Amazon Nova (Premier, Pro, Lite oder Micro) auf Amazon Bedrock auf der Grundlage spezifischer Geschäftsanforderungen aus.
- Definieren Sie benutzerdefinierte Tools, die eine Verbindung zu Unternehmenssystemen und Datenquellen herstellen.
- Verarbeiten Sie mehrere Modalitäten, darunter Text, Bilder und Sprache.
- Stellen Sie Agenten bereit, die selbstständig auf Geschäftsanfragen antworten und Aufgaben ausführen können.

Dieser Implementierungsansatz ermöglicht es Geschäftsteams, autonome Agenten schnell zu entwickeln und einzusetzen, ohne über fundiertes technisches Fachwissen in der Entwicklung von KI-Modellen zu verfügen.

Beispiel aus der Praxis für Strands Agents

AWS Transform for.NET nutzt Strands Agents zur Unterstützung seiner Funktionen zur Anwendungsmodernisierung, wie in AWS Transform for .NET beschrieben, dem ersten agentischen KI-Dienst für die Modernisierung von.NET-Anwendungen im großen Maßstab (Blog).AWS Dieser Produktionsservice setzt mehrere spezialisierte autonome Agenten ein. Die Agenten arbeiten zusammen, um ältere .NET-Anwendungen zu analysieren, Modernisierungsstrategien zu planen und Codetransformationen in cloudnative Architekturen ohne menschliches Eingreifen durchzuführen. AWS Transform for .NET demonstriert die Produktionsreife von autonomen Strands Agents Unternehmenssystemen.

LangChain und LangGraph

LangChainist eines der etabliertesten Frameworks im agentischen KI-Ökosystem. LangGrapherweitert seine Funktionen, um komplexe, zustandsbehaftete Agenten-Workflows zu unterstützen, wie im Blog beschrieben. LangChain Zusammen bieten sie eine umfassende Lösung für den Aufbau ausgeklügelter autonomer KI-Agenten mit umfangreichen Orchestrierungsfunktionen für einen unabhängigen Betrieb.

Hauptmerkmale von und LangChain LangGraph

LangChainund LangGraph beinhalten die folgenden Hauptmerkmale:

- Komponenten-Ökosystem — Umfangreiche Bibliothek mit vorgefertigten Komponenten für verschiedene Funktionen autonomer Agenten, die die schnelle Entwicklung spezialisierter Agenten ermöglicht. Weitere Informationen finden Sie in der Dokumentation unter Schnellstart. LangChain
- Auswahl von Foundation-Modellen — Support für verschiedene Foundation-Modelle, darunter Anthropic Claude, Amazon Nova-Modelle (Premier, Pro, Lite und Micro) auf Amazon Bedrock und andere für unterschiedliche Argumentationsfähigkeiten. Weitere Informationen finden Sie in der Dokumentation unter Eingaben und Ausgaben. LangChain
- LLM-API-Integration — Standardisierte Schnittstellen für mehrere Large Language Model (LLM) - Dienstleister, darunter Amazon Bedrock und andereOpenAI, für eine flexible Bereitstellung. Weitere Informationen finden Sie unter LLMs in der LangChain-Dokumentation.
- Multimodale Verarbeitung — Integrierte Unterstützung für Text-, Bild- und Audioverarbeitung, um umfangreiche multimodale autonome Interaktionen mit Agenten zu ermöglichen. Weitere Informationen finden Sie in der Dokumentation unter Multimodalität. LangChain
- Graphbasierte Workflows — LangGraph ermöglichen die Definition komplexer Verhaltensweisen autonomer Agenten als Zustandsmaschinen und unterstützen so eine ausgeklügelte Entscheidungslogik. Weitere Informationen finden Sie in der Ankündigung von LangGraphPlatform GA.
- Speicherabstraktionen — Mehrere Optionen für die Kurz- und Langzeitspeicherverwaltung, was für autonome Agenten, die den Kontext im Laufe der Zeit aufrechterhalten, unerlässlich ist. Weitere Informationen finden Sie in der Dokumentation unter So fügen Sie Chatbots Speicher hinzu. LangChain
- Tool-Integration — Umfangreiches Ökosystem von Tool-Integrationen für verschiedene Dienste und Erweiterung der APIs Funktionen autonomer Agenten. Weitere Informationen finden Sie in der LangChain Dokumentation unter Tools.
- LangGraph Plattform — Verwaltete Bereitstellungs- und Überwachungslösung für Produktionsumgebungen, die autonome Agenten mit langer Laufzeit unterstützt. Weitere Informationen finden Sie in der Ankündigung von LangGraphPlatform GA.

Wann sollte man verwenden LangChain und LangGraph

LangChainund LangGraph eignen sich besonders gut für Szenarien mit autonomen Agenten, darunter:

- Komplexe, mehrstufige Argumentationsabläufe, die eine ausgeklügelte Orchestrierung für autonome Entscheidungen erfordern
- Projekte, die Zugang zu einem großen Ökosystem vorgefertigter Komponenten und Integrationen für vielfältige autonome Funktionen benötigen
- Teams mit vorhandener Python Infrastruktur und Fachwissen für maschinelles Lernen (ML), die autonome Systeme aufbauen möchten
- Anwendungsfälle, die eine komplexe Statusverwaltung für lang andauernde autonome Agentensitzungen erfordern

Implementierungsansatz für und LangChain LangGraph

LangChainund LangGraph bieten einen strukturierten Implementierungsansatz für Unternehmensbeteiligte, wie in der LangGraphDokumentation detailliert beschrieben. Das Framework ermöglicht Organisationen:

- Definieren Sie ausgefeilte Workflow-Diagramme, die Geschäftsprozesse darstellen.
- Erstellen Sie mehrstufige Argumentationsmuster mit Entscheidungspunkten und bedingter Logik.
- Integrieren Sie multimodale Verarbeitungsfunktionen für den Umgang mit unterschiedlichen Datentypen.
- Implementieren Sie die Qualitätskontrolle mithilfe integrierter Überprüfungs- und Validierungsmechanismen.

Dieser grafisch gestützte Ansatz ermöglicht es Geschäftsteams, komplexe Entscheidungsprozesse als autonome Workflows zu modellieren. Die Teams haben einen klaren Überblick über jeden Schritt des Argumentationsprozesses und können Entscheidungswege überprüfen.

Ein Beispiel aus der Praxis für und LangChain LangGraph

Vodafonehat autonome Agenten implementiert, die LangChain (undLangGraph) verwenden, um seine Workflows für Datentechnik und Betrieb zu verbessern, wie in der Fallstudie LangChain Enterprise detailliert beschrieben. Sie haben interne KI-Assistenten entwickelt, die selbstständig

Leistungskennzahlen überwachen, Informationen aus Dokumentationssystemen abrufen und umsetzbare Erkenntnisse liefern — alles durch Interaktionen in natürlicher Sprache.

Die Vodafone Implementierung verwendet LangChain modulare Dokumentenlader, Vektorintegration und Unterstützung für mehrere LLMs (, LLaMA 3 undGemini), um diese Pipelines OpenAI schnell zu prototypisieren und zu vergleichen. Anschließend strukturierten sie die Orchestrierung LangGraph mit mehreren Agenten durch den Einsatz modularer Unteragenten. Diese Agenten führen Aufgaben zur Erfassung, Verarbeitung, Zusammenfassung und Argumentation durch. LangGraphhaben diese Agenten APIs in ihre Cloud-Systeme integriert.

CrewAI

CrewAIist ein Open-Source-Framework, das sich speziell auf die autonome Orchestrierung mehrerer Agenten konzentriert und unter verfügbar ist. GitHub Es bietet einen strukturierten Ansatz zur Bildung von Teams spezialisierter autonomer Agenten, die zusammenarbeiten, um komplexe Aufgaben ohne menschliches Eingreifen zu lösen. CrewAIbetont die rollenbasierte Koordination und Aufgabendelegierung.

Hauptmerkmale von CrewAI

CrewAIbietet die folgenden Hauptfunktionen:

- Rollenbasiertes Agentendesign — Autonome Agenten werden mit spezifischen Rollen, Zielen und Hintergrundinformationen definiert, um spezialisiertes Fachwissen zu ermöglichen. Weitere Informationen finden Sie in der Dokumentation unter Crefting Effective Agents. CrewAI
- Aufgabendelegierung — Integrierte Mechanismen für die autonome Zuweisung von Aufgaben an geeignete Agenten auf der Grundlage ihrer Fähigkeiten. Weitere Informationen finden Sie in der CrewAIDokumentation unter Aufgaben.
- Zusammenarbeit mit Agenten — Framework für autonome Kommunikation und Wissensaustausch zwischen Agenten ohne menschliche Vermittlung. Weitere Informationen finden Sie in der CrewAI Dokumentation unter Zusammenarbeit.
- Prozessmanagement — Strukturierte Workflows für sequentielle und parallel autonome Aufgabenausführung. Weitere Informationen finden Sie in der CrewAI Dokumentation unter Prozesse.
- Auswahl des Foundation-Modells — Support für verschiedene Foundation-Modelle, darunter Anthropic Claude, Amazon Nova-Modelle (Premier, Pro, Lite und Micro) auf Amazon Bedrock

und andere zur Optimierung für verschiedene Aufgaben des autonomen Denkens. Weitere Informationen finden Sie unter LLMs in der CrewAI-Dokumentation.

- LLM-API-Integration — Flexible Integration mit mehreren LLM-Serviceschnittstellen, einschließlich Amazon BedrockOpenAI, und lokalen Modellbereitstellungen. Weitere Informationen finden Sie in der Dokumentation unter Konfigurationsbeispielen für Anbieter. CrewAI
- Multimodale Unterstützung — Neue Funktionen für den Umgang mit Text, Bildern und anderen Modalitäten für umfassende Interaktionen mit autonomen Agenten. Weitere Informationen finden Sie in der Dokumentation unter Verwenden multimodaler Agenten. CrewAI

Wann sollte es verwendet werden CrewAI

CrewAIeignet sich besonders gut für Szenarien mit autonomen Agenten, darunter:

- Komplexe Probleme, die von spezialisiertem, rollenbasiertem Fachwissen profitieren, das autonom arbeitet
- Projekte, die eine ausdrückliche Zusammenarbeit zwischen mehreren autonomen Agenten erfordern
- Anwendungsfälle, in denen die teambasierte Problemzerlegung die autonome Problemlösung verbessert
- Szenarien, die eine klare Trennung der Anliegen zwischen den verschiedenen Rollen autonomer Agenten erfordern

Implementierungsansatz für CrewAI

CrewAIbietet eine rollenbasierte Implementierung eines Ansatzes für Teams von KI-Agenten für Geschäftsbeteiligte, wie in der CrewAI Dokumentation unter Erste Schritte beschrieben. Das Framework ermöglicht Organisationen:

- Definieren Sie spezialisierte autonome Agenten mit spezifischen Rollen, Zielen und Fachgebieten.
- Weisen Sie Agenten Aufgaben auf der Grundlage ihrer speziellen Fähigkeiten zu.
- Richten Sie klare Abhängigkeiten zwischen Aufgaben ein, um strukturierte Workflows zu erstellen.
- Orchestrieren Sie die Zusammenarbeit zwischen mehreren Agenten, um komplexe Probleme zu lösen.

Dieser rollenbasierte Ansatz spiegelt die menschlichen Teamstrukturen wider und macht es Unternehmensleitern leicht, ihn zu verstehen und umzusetzen. Organizations können autonome Teams mit speziellen Fachgebieten bilden, die zusammenarbeiten, um Geschäftsziele zu erreichen, ähnlich wie menschliche Teams arbeiten. Das autonome Team kann jedoch kontinuierlich ohne menschliches Eingreifen arbeiten.

Beispiel aus der Praxis für CrewAI

AWS hat autonome Multi-Agenten-Systeme mithilfe von CrewAI implementiert, die in Amazon Bedrock integriert sind, wie in der CrewAIveröffentlichten Fallstudie detailliert beschrieben. AWS und CrewAI entwickelte ein sicheres, herstellerneutrales Framework. Die CrewAI Open-Source-Architektur „Flows-and-Crews“ lässt sich nahtlos in die Grundmodelle, Speichersysteme und Compliance-Richtlinien von Amazon Bedrock integrieren.

Zu den wichtigsten Elementen der Implementierung gehören:

- Blueprints und Open Sourcing — AWS und CrewAI veröffentlichte Referenzdesigns, die CrewAI Agenten Amazon Bedrock-Modellen und Observability-Tools zuordnen. Außerdem veröffentlichten sie beispielhafte Systeme wie ein Team für AWS Sicherheitsüberprüfungen mit mehreren Mitarbeitern, Abläufe zur Code-Modernisierung und Backoffice-Automatisierung für Konsumgüter (CPG).
- Integration des Observability-Stacks — Die Lösung integriert die Überwachung in Amazon CloudWatch und ermöglicht so Rückverfolgbarkeit und LangFuse Debugging vom Machbarkeitsnachweis bis zur Produktion. AgentOps
- Nachgewiesene Investitionsrendite (ROI) — Erste Pilotprojekte zeigen wichtige Verbesserungen: 70 Prozent schnellere Ausführung bei einem großen Code-Modernisierungsprojekt und etwa 90 Prozent Verkürzung der Verarbeitungszeit für CPG-Backoffice-Abläufe.

AutoGen

AutoGenist ein Open-Source-Framework, das ursprünglich von veröffentlicht wurde. Microsoft AutoGenkonzentriert sich darauf, autonome KI-Agenten zur Konversation und Zusammenarbeit zu ermöglichen. Es bietet eine flexible Architektur für den Aufbau von Systemen mit mehreren Agenten, wobei der Schwerpunkt auf asynchronen, ereignisgesteuerten Interaktionen zwischen Agenten für komplexe autonome Workflows liegt.

Hauptmerkmale von AutoGen

AutoGenbietet die folgenden Hauptfunktionen:

- Konversationsagenten — Basiert auf Konversationen zwischen autonomen Agenten in natürlicher Sprache und ermöglicht so anspruchsvolles Denken im Dialog. Weitere Informationen finden Sie in der Dokumentation unter Multi-Agent Conversation Framework. AutoGen
- Asynchrone Architektur — Ereignisgesteuertes Design für blockierungsfreie autonome Agenteninteraktionen, das komplexe parallel Workflows unterstützt. Weitere Informationen finden Sie in der Dokumentation unter Mehrere Aufgaben in einer Folge von asynchronen Chats lösen. AutoGen
- H uman-in-the-loop — Starke Unterstützung für die optionale Beteiligung von Personen an ansonsten autonomen Agenten-Workflows, falls erforderlich. Weitere Informationen finden Sie in der AutoGen Dokumentation unter Zulassen von menschlichem Feedback bei Agenten.
- Codegenerierung und -ausführung — Spezialisierte Funktionen für codeorientierte autonome Agenten, die Code schreiben und ausführen können. Weitere Informationen finden Sie in der Dokumentation unter Codeausführung. AutoGen
- Individuell anpassbares Verhalten — Flexible autonome Agentenkonfiguration und Konversationssteuerung für verschiedene Anwendungsfälle. Weitere Informationen finden Sie in der Dokumentation unter agentchat.conversable_agent. AutoGen
- Auswahl des Foundation-Modells — Support für verschiedene Foundation-Modelle, darunter Anthropic Claude, Amazon Nova-Modelle (Premier, Pro, Lite und Micro) auf Amazon Bedrock und andere für verschiedene Funktionen für autonomes Denken. Weitere Informationen finden Sie in der Dokumentation unter LLM-Konfiguration. AutoGen
- LLM-API-Integration — Standardisierte Konfiguration für mehrere LLM-Serviceschnittstellen, darunter Amazon Bedrock, undOpenAI. Azure OpenAI Weitere Informationen finden Sie unter oai.openai_utils in der API-Referenz. AutoGen
- Multimodale Verarbeitung — Support für Text- und Bildverarbeitung, um umfangreiche multimodale Interaktionen mit autonomen Agenten zu ermöglichen. Weitere Informationen finden Sie in der Dokumentation unter Umgang mit multimodalen Modellen: GPT-4V. AutoGen AutoGen

Wann sollte es verwendet werden AutoGen

AutoGeneignet sich besonders gut für Szenarien mit autonomen Agenten, darunter:

- Anwendungen, die für komplexes Denken natürliche Konversationsabläufe zwischen autonomen Agenten erfordern
- Projekte, die sowohl einen vollständig autonomen Betrieb als auch optionale Funktionen zur menschlichen Überwachung benötigen
- Anwendungsfälle, die autonome Codegenerierung, Ausführung und Debugging ohne menschliches Eingreifen beinhalten
- Szenarien, die flexible, asynchrone Kommunikationsmuster für autonome Agenten erfordern

Implementierungsansatz für AutoGen

AutoGenbietet einen dialogorientierten Implementierungsansatz für Geschäftsbeteiligte, wie in der AutoGen Dokumentation unter Erste Schritte beschrieben. Das Framework ermöglicht Organisationen:

- Erstellen Sie autonome Agenten, die über Konversationen in natürlicher Sprache kommunizieren.
- Implementieren Sie asynchrone, ereignisgesteuerte Interaktionen zwischen mehreren Agenten.
- Kombinieren Sie bei Bedarf einen vollständig autonomen Betrieb mit optionaler menschlicher Aufsicht.
- Entwickeln Sie spezialisierte Agenten für verschiedene Geschäftsfunktionen, die im Dialog zusammenarbeiten.

Dieser dialogorientierte Ansatz macht die Argumentation des autonomen Systems transparent und für Geschäftsanwender zugänglich. Entscheidungsträger können den Dialog zwischen den Akteuren beobachten, um zu verstehen, wie Schlussfolgerungen gezogen werden, und optional an der Konversation teilnehmen, wenn menschliches Urteilsvermögen gefragt ist.

Ein Beispiel aus der Praxis für AutoGen

Magentic-Oneist ein generalistisches Open-Source-Multiagentensystem, das entwickelt wurde, um komplexe, mehrstufige Aufgaben in unterschiedlichen Umgebungen autonom zu lösen, wie im AI Frontiers-Blog beschrieben. Microsoft Sein Herzstück ist der Orchestrator-Agent, der übergeordnete Ziele zerlegt und Fortschritte mithilfe strukturierter Ledger verfolgt. Dieser Agent delegiert Unteraufgaben an spezialisierte Agenten (wieWebSurfer,, undComputerTerminal) und passt sich dynamisch an FileSurferCoder, indem er bei Bedarf neu plant.

Das System basiert auf dem AutoGen Framework, ist modellunabhängig und verwendet standardmäßig GPT-4o. Es erreicht bei Benchmarks wie, und eine Leistung auf dem neuesten Stand der Technik — und das alles ohne aufgabenspezifische Feinabstimmung. GAIA AssistantBench WebArena Darüber hinaus unterstützt es modulare Erweiterbarkeit und eine strenge Evaluierung anhand von Vorschlägen. AutoGenBench

LlamaIndex

LlamaIndexist ein Datenframework, das speziell für die Verbindung umfangreicher Sprachmodelle (LLMs) mit externen Datenquellen entwickelt wurde, um anspruchsvolle Retrieval Augmented Generation (RAG) und agentische KI-Anwendungen zu ermöglichen. Das Framework bietet Abstraktionen und beschleunigte Entwicklungsworkflows für agentische Systeme, benutzerdefinierte Orchestrierungsmuster und Systemintegrationen, die für wissensgestützte KI-Lösungen weniger Aufwand bedeuten. time-to-production

Hauptmerkmale von LlamaIndex

LlamaIndexbietet einen umfassenden Funktionsumfang, der sich besonders für KI-Anwendungen in Unternehmen eignet:

- Datenzentrierte Architektur — Hervorragend geeignet für das Erfassen, Indexieren und Abrufen von Informationen aus über 100 Datenformaten, darunter Word-Dokumenten PDFs, Microsoft Tabellen und mehr. Das Framework wandelt Unternehmensdaten in abfragbare Wissensdatenbanken um, die für KI-Agenten optimiert sind. Weitere Informationen finden Sie in der LlamaIndex-Dokumentation.
- Produktionsbereite Bereitstellung — LlamaIndex bietet sowohl Open-Source-Frameworks als auch Managed Services und bietet Funktionen auf Unternehmensebene wie SicherheitskontrollenLlamaCloud, Skalierbarkeit, Integrationen zur Beobachtbarkeit und Flexibilität bei der Bereitstellung. Weitere Informationen finden Sie in der Framework-Dokumentation. LlamaIndex
- Erweiterte Dokumentenverarbeitung — LlamaCloud bietet Funktionen zum Analysieren, Extrahieren, Indexieren und Abrufen von Dokumenten, die komplexe Layouts, verschachtelte Tabellen, multimodale Inhalte und sogar handschriftliche Notizen verarbeiten. Dieses ausgeklügelte Parsing ermöglicht es Mitarbeitern, effektiv mit realen Unternehmensdokumenten zu arbeiten, die Diagramme, Diagramme und komplexe Formatierungen enthalten. Weitere Informationen finden Sie in der LlamaCloud-Dokumentation.

- Workflow-Orchestrierung — LlamaAgents bietet eine ereignisgesteuerte, asynchrone Orchestrierungs-Engine zum Aufbau mehrstufiger Agentensysteme. Workflows unterstützen komplexe Muster wie Schleifen, parallel Ausführung, bedingte Verzweigung und statusbehaftete Wiederaufnahme, wodurch sie sich ideal für anspruchsvolle Agenteninteraktionen eignen. Weitere Informationen finden Sie in der LlamaIndex Workflow-Dokumentation.
- Agentenabruffunktionen — Erweiterte Abrufmodi wie Hybridsuche, semantische Suche und automatisches Routing, die auf intelligente Weise die beste Abrufstrategie für jede Abfrage bestimmen. Das Framework unterstützt den kombinierten Abruf mehrerer Wissensdatenbanken mit Neueinstufungen zur Erhöhung der Genauigkeit. Weitere Informationen finden Sie in der LlamaIndex RAG-Dokumentation.
- Beobachtbarkeit und Bewertung — LlamaIndex lässt sich in eine Vielzahl von Beobachtungs- und Bewertungsinstrumenten integrieren. Diese Integrationsfunktion hilft Ihnen dabei, Ihre Anwendungen zu verfolgen und zu debuggen, ihre Leistung zu bewerten und die Kosten zu überwachen. Weitere Informationen finden Sie in der Dokumentation Tracing, Debugging and Evaluating. LlamaIndex

Wann sollte es verwendet werden LlamaIndex

LlamaIndexeignet sich besonders gut für agentische KI-Szenarien, in denen datenintensive Workflows und Wissensmanagement im Vordergrund stehen:

- Dokumentenintensive Anwendungen, bei denen Agenten große Mengen an Unternehmensdokumenten wie Verträgen, Berichten, Handbüchern und behördlichen Unterlagen verarbeiten, analysieren und Erkenntnisse daraus gewinnen müssen
- Schnelles Prototyping bis hin zu Produktionsszenarien, in denen Unternehmen schnell dokumentenorientierte Agenten ohne großen Aufwand für das Infrastrukturmanagement erstellen und einsetzen möchten
- RAG-First-Architekturen, bei denen Genauigkeit und Kontextrelevanz im Vordergrund stehen, insbesondere bei der Arbeit mit komplexen, multimodalen Dokumenten, die Tabellen, Bilder und strukturierte Daten enthalten
- Dokumenten-Workflows mit mehreren Agenten, die spezialisierte Agenten für verschiedene Aspekte der Dokumentenverarbeitung erfordern, z. B. Analyse, Zusammenfassung und Konformitätsprüfung

Implementierungsansatz für LlamaIndex

LlamaIndex bietet sowohl Bausteine auf niedriger Ebene als auch Abstraktionen auf hoher Ebene, die unterschiedlichen Implementierungsansätzen Rechnung tragen:

- Schnelle Entwicklung funktionaler RAG-Anwendungen in nur wenigen Codezeilen mithilfe LlamaIndex von High-Level. APIs Dieser Ansatz ist für Geschäftsteams und Entwickler LlamaIndex zugänglich, die noch keine Erfahrung mit agentischer KI haben.
- Unternehmensintegration durch LlamaHub für beliebte Unternehmenssysteme wie SharePoint Amazon Simple Storage Service (Amazon S3), Datenbanken und APIs. Dieser Ansatz ermöglicht eine nahtlose Integration in die bestehende Dateninfrastruktur.
- Flexible Bereitstellungsoptionen zwischen selbst gehosteten Open-Source-Bereitstellungen für maximale Kontrolle oder LlamaCloud verwalteten Diensten für weniger Betriebskosten und Unternehmensfunktionen.
- Anwendungen können mit einfachen Abfrage-Engines beginnen und nach und nach um Agentenfunktionen, Orchestrierung mehrerer Agenten und komplexe Workflows erweitern, wenn sich die Anforderungen ändern.

Ein Beispiel aus der Praxis für LlamaIndex

Dieses Beispiel konzentriert sich auf eine Tochtergesellschaft eines Luft- und Raumfahrtunternehmens, das sich auf Navigations- und Betriebslösungen für die Luftfahrt spezialisiert hat. Sie müssen sich einer wachsenden Herausforderung stellen, zu der die Erprobung unkoordinierter KI-Chatbot-Versuche gehört. Die Versuche führten zu wiederholter Arbeit, langen Entwicklungszyklen, Compliance-Hindernissen und isolierten Implementierungen im gesamten Unternehmen.

Sie entwickelten ein einheitliches Agenten-Framework, eine wiederverwendbare, auf Vorlagen basierende Lösung, die auf dem LlamaIndex Open-Source-Framework basiert und die Agentenerstellung erheblich effizienter macht. Sie verglichen mehrere konkurrierende Frameworks, sowohl kettenorientiert als auch grafisch. Letztlich entschieden sie sich LlamaIndex für drei entscheidende Vorteile: das flexible Design, die modularen Komponenten und die produktionsreife Orchestrierungssteuerung.

Die Plattform reduziert die Zeit für die Entwicklung und Bereitstellung von Agenten um 87% von 512 auf 64 Stunden. Diese Reduzierung wurde dadurch erreicht, dass Teams Agenten mit etwa

50 Codezeilen und einer JSON-Konfigurationsdatei erstellen konnten. Die Teams nutzten ein einheitliches Framework mit integrierter Sicherheit, Compliance und privilegiertem Systemzugriff. Weitere Informationen finden Sie in den Fallstudien von LlamaIndex Kunden.

Vergleich agentischer KI-Frameworks

Berücksichtigen Sie bei der Auswahl eines agentischen KI-Frameworks für die Entwicklung autonomer Agenten, wie jede Option Ihren spezifischen Anforderungen entspricht. Berücksichtigen Sie nicht nur die technischen Fähigkeiten, sondern auch die organisatorische Eignung, einschließlich der Expertise des Teams, der vorhandenen Infrastruktur und der langfristigen Wartungsanforderungen. Viele Unternehmen könnten von einem hybriden Ansatz profitieren, bei dem mehrere Frameworks für verschiedene Komponenten ihres autonomen KI-Ökosystems genutzt werden.

In der folgenden Tabelle werden die Reifegrade (am stärksten, am stärksten, angemessen oder schwach) der einzelnen Frameworks anhand der wichtigsten technischen Dimensionen verglichen. Für jedes Framework enthält die Tabelle auch Informationen zu den Optionen für den Einsatz in der Produktion und zur Komplexität der Lernkurve.

Framework	AWS Integration	Autonome Unterstützung für mehrere Agenten	Komplexität des autonomen Workflows	Multimodale Fähigkeiten	Auswahl des Fundamentmodells	LLM-API-Integration	Einsatz in der Produktion	Lernkurve
AutoGen	Schwach	Stark	Stark	Ausreichend	Ausreichend	Stark	Mach es selbst (DIY)	Steil
CrewAI	Schwach	Stark	Ausreichend	Schwach	Ausreichend	Ausreichend	DIY	Mittel
LangChain / LangGraph	Ausreichend	Stark	Am stärksten	Am stärksten	Am stärksten	Am stärksten	Plattform oder DIY	Steil

LlamaIndex	Ausreichend	Ausreichend	Stark	Ausreichend	Stark	Stark	Plattform oder DIY	Mittel
Strands Agents	Am stärksten	Stark	Am stärksten	Stark	Stark	Am stärksten	DIY	Mittel

Überlegungen bei der Auswahl eines agentischen KI-Frameworks

Berücksichtigen Sie bei der Entwicklung autonomer Agenten die folgenden Schlüsselfaktoren:

- AWS Infrastrukturintegration — Organizations, in die viel investiert AWS wird, werden am meisten von den nativen Integrationen von Strands Agents AWS-Services für autonome Workflows profitieren. Weitere Informationen finden Sie unter AWS Weekly Roundup (Blog)AWS .
- Auswahl des Foundation-Modells — Überlegen Sie anhand der Argumentationsanforderungen Ihres autonomen Agenten, welches Framework Ihre bevorzugten Foundation-Modelle am besten unterstützt (z. B. Amazon Nova-Modelle auf Amazon Bedrock oder Anthropic Claude). Weitere Informationen finden Sie auf der Website unter Building Effective Agents. Anthropic
- LLM-API-Integration — Evaluieren Sie Frameworks auf der Grundlage ihrer Integration mit Ihren bevorzugten Large Language Model (LLM) -Serviceschnittstellen (z. B. Amazon Bedrock oderOpenAI) für die Produktionsbereitstellung. Weitere Informationen finden Sie in der Dokumentation unter Model Interfaces. Strands Agents
- Multimodale Anforderungen — Bei autonomen Agenten, die Text, Bilder und Sprache verarbeiten müssen, sollten Sie die multimodalen Fähigkeiten der einzelnen Frameworks berücksichtigen. Weitere Informationen finden Sie in der Dokumentation unter Multimodalität. LangChain
- Komplexität autonomer Workflows — Komplexere autonome Workflows mit ausgefeilter Zustandsverwaltung könnten die fortschrittlichen Funktionen von State Machines begünstigen. von. LangGraph
- Autonome Teamzusammenarbeit — Projekte, die eine explizite, rollenbasierte, autonome Zusammenarbeit zwischen spezialisierten Mitarbeitern erfordern, können von der teamorientierten Architektur von profitieren. CrewAI
- Paradigma der autonomen Entwicklung — Teams, die asynchrone Konversationsmuster für autonome Agenten bevorzugen, bevorzugen möglicherweise die ereignisgesteuerte Architektur von. AutoGen

- Verwalteter oder codebasierter Ansatz — Organizations, die eine vollständig verwaltete Erfahrung mit minimalem Programmieraufwand wünschen, sollten Amazon Bedrock Agents in Betracht ziehen. Organizations, die eine tiefere Anpassung benötigen, bevorzugen Strands Agents möglicherweise andere Frameworks mit speziellen Funktionen, die besser auf die spezifischen Anforderungen autonomer Agenten zugeschnitten sind.
- Produktionsreife autonomer Systeme — Erwägen Sie Bereitstellungsoptionen, Überwachungsmöglichkeiten und Unternehmensfunktionen für autonome Produktionsagenten.

Plattformen

Agentic KI-Plattformen bieten die grundlegenden Laufzeit-, Orchestrierungs- und Integrationsebenen, die für die Bereitstellung, Skalierung und Verwaltung von agentischen Systemen in Produktionsqualität erforderlich sind. Frameworks definieren, wie Agenten aufgebaut sind, und Protokolle regeln, wie sie kommunizieren. Plattformen bieten die Umgebung, in der diese Agenten sicher und skalierbar arbeiten, zusammenarbeiten und sich weiterentwickeln können.

Agentenplattformen kombinieren Modellausführung, Kontextmanagement, Toolintegration, Beobachtbarkeit und Governance-Funktionen in einheitlichen Umgebungen. Diese Plattformen ermöglichen es Unternehmen, vom Experimentieren zur Bereitstellung auf Unternehmensebene überzugehen.

In diesem Abschnitt:

Warum Plattformen wichtig sind

Agentische KI-Plattformen sind für Unternehmen, die autonome Systeme in der Produktion operationalisieren möchten, von entscheidender Bedeutung. Sie bieten die folgenden Funktionen:

- Stellen Sie Runtime-Orchestrierung für das Hosten, Skalieren und Koordinieren von Agenten bereit.
- Verwalten Sie Status, Kontext und Speicher in Workflows mit mehreren Agenten.
- Bieten Sie Sicherheits-, Identitäts- und Governance-Kontrollen an, die den Unternehmensstandards entsprechen.
- Integrieren Sie mithilfe von Standards APIs oder Protokollen in Tooling-Ökosysteme und externe Systeme.
- Sorgen Sie für Beobachtbarkeit und Überprüfbarkeit aller Agenteninteraktionen und Ereignisabläufe.

- Support modellübergreifende Interoperabilität, sodass Agenten mehrere Basismodelle in einer einzigen Umgebung verwenden können.

Diese Funktionen machen aus einzelnen Agenten koordinierte, anpassungsfähige Systeme, die innerhalb der Unternehmens- und behördlichen Grenzen zuverlässig funktionieren können.

Arten von agentischen KI-Plattformen

Agentische KI-Plattformen lassen sich in der Regel in eine oder mehrere der folgenden Kategorien einteilen:

- Managed Agent — Vollständig verwaltete Plattformen bieten integrierte Infrastruktur-, Speicher- und Orchestrierungsfunktionen. Sie reduzieren den betrieblichen Aufwand und beschleunigen die Produktionszeit.
- Open-Source-Orchestrierung — Agenturbasierte Open-Source-Plattformen bieten Flexibilität und Transparenz für Unternehmen, die anpassbare Umgebungen oder die Bereitstellung vor Ort bevorzugen.
- Hybrides Unternehmen — Hybride Plattformen integrieren verwaltete und selbst gehostete Komponenten und kombinieren so die Skalierbarkeit von Cloud-verwalteten Diensten mit der Steuerung von Unternehmenssystemen.

Überlegungen zur Plattformauswahl

Bei der Auswahl oder Gestaltung einer agentischen KI-Plattform sollten Unternehmen Folgendes berücksichtigen:

- Integrationstiefe — Bewerten Sie, wie gut sich die Plattform in bestehende Datenquellen, Tools und Protokolle integrieren lässt.
- Skalierbarkeit — Stellen Sie sicher, dass die Plattform dynamisch skaliert werden kann, um autonome Workloads und die Zusammenarbeit mehrerer Agenten zu unterstützen.
- Sicherheit und Compliance — Beurteilen Sie die Datenschutz-, Verschlüsselungs- und Governance-Funktionen anhand organisatorischer und regionaler Anforderungen.
- Erweiterbarkeit — Wählen Sie Plattformen mit modularen Architekturen, die es ermöglichen, im Laufe der Zeit neue Tools, Modelle oder Agenten hinzuzufügen.

- Beobachtbarkeit — Bevorzugen Sie Plattformen, die detaillierte Telemetrie-, Rückverfolgbarkeits- und Auditprotokolle für Agenteninteraktionen bieten.
- Kosteneffizienz — Ziehen Sie serverlose oder nutzungsbasierte Modelle in Betracht, um die Kosten für variable Workloads zu optimieren.

Agenten für Amazon Bedrock

Amazon Bedrock Agents ist ein vollständig verwalteter Service, mit dem Sie autonome Agenten in Ihren Anwendungen erstellen und konfigurieren können. Er kann Interaktionen zwischen Basismodellen, Datenquellen, Softwareanwendungen und Benutzerkonversationen orchestrieren. Dank des optimierten Ansatzes zur Erstellung von Agenten müssen Sie keine Kapazität bereitstellen, die Infrastruktur verwalten oder benutzerdefinierten Code schreiben.

Hauptfunktionen von Amazon Bedrock Agents

Amazon Bedrock Agents umfasst die folgenden Hauptfunktionen:

- Vollständig verwalteter Service — Umfassendes Infrastrukturmanagement, ohne dass Kapazitäten bereitgestellt oder die zugrunde liegenden Systeme verwaltet werden müssen. Weitere Informationen finden Sie unter Automatisieren von Aufgaben in Ihrer Anwendung mithilfe von KI-Agenten in der Amazon Bedrock-Dokumentation.
- API-gesteuerte Entwicklung — Definieren Sie Agenten und führen Sie sie durch einfache API-Aufrufe aus, indem Sie Modelle, Anweisungen, Tools und Konfigurationsparameter angeben. Weitere Informationen finden Sie unter Agenten manuell erstellen und konfigurieren in der Amazon Bedrock-Dokumentation.
- Aktionsgruppen — Definieren Sie spezifische Aktionen, die Ihr Agent ausführen kann, indem Sie Aktionsgruppen mit API-Schemas erstellen. Weitere Informationen finden Sie in der Amazon Bedrock-Dokumentation unter Verwenden von Aktionsgruppen zur Definition von Aktionen, die Ihr Agent ausführen soll.
- Wissensdatenbank-Integration — Stellen Sie eine nahtlose Verbindung zu Amazon Bedrock Knowledge Bases her, um die Antworten der Agenten mit den Daten Ihres Unternehmens zu ergänzen. Weitere Informationen finden Sie unter Verbessern Sie die Antwortgenerierung für Ihren Agenten mit der Wissensdatenbank in der Amazon Bedrock-Dokumentation.
- Erweiterte Vorlagen für Eingabeaufforderungen — Passen Sie das Verhalten Ihrer Agenten mithilfe von Vorlagen für die Vorverarbeitung, Orchestrierung, Generierung und Nachbearbeitung von Antworten in der Wissensdatenbank individuell an. Weitere Informationen finden Sie in der Amazon

Bedrock-Dokumentation unter Verbessern Sie die Genauigkeit des Agenten mithilfe erweiterter Eingabeaufforderungsvorlagen in Amazon Bedrock.

- Rückverfolgung und Beobachtbarkeit — Verfolgen Sie den step-by-step Argumentationsprozess des Agenten mithilfe der integrierten Nachverfolgungsfunktionen. Weitere Informationen finden Sie unter Nachverfolgen des step-by-step Argumentationsprozesses des Agenten mithilfe von Trace in der Amazon Bedrock-Dokumentation.
- Versionierung und Aliase — Erstellen Sie mehrere Versionen Ihres Agenten und stellen Sie sie über Aliase für kontrollierte Rollouts bereit. Weitere Informationen finden Sie in der Amazon Bedrock-Dokumentation unter Bereitstellen und Verwenden eines Amazon Bedrock-Agenten in Ihrer Anwendung.

Wann sollten Sie Amazon Bedrock Agents verwenden

Amazon Bedrock Agents eignet sich besonders gut für autonome Agentenszenarien, darunter:

- Organizations, die ein vollständig verwaltetes Erlebnis für die Erstellung und Bereitstellung von Agenten wünschen, ohne die Infrastruktur verwalten zu müssen
- Projekte, die eine schnelle Entwicklung und Bereitstellung von Agenten durch Konfiguration und nicht durch Code erfordern
- Anwendungsfälle, die von einer engen Integration mit anderen Amazon Bedrock-Funktionen wie Knowledge Bases und Guardrails profitieren
- Teams, die nicht über die internen Ressourcen verfügen, um Agenten von Grund auf neu zu erstellen, benötigen aber produktionsreife autonome Funktionen

Implementierungsansatz für Amazon Bedrock Agents

Amazon Bedrock Agents bietet einen konfigurationsbasierten Implementierungsansatz für Geschäftsbeteiligte. Der Service ermöglicht Unternehmen:

- Definieren Sie Agenten über die AWS-Managementkonsole oder API-Aufrufe, ohne komplexen Code zu schreiben.
- Erstellen Sie Aktionsgruppen, die APIs angeben, welche Operationen der Agent ausführen kann.
- Connect Wissensdatenbanken, um dem Agenten domänenspezifische Informationen zur Verfügung zu stellen.
- Testen und wiederholen Sie das Verhalten von Agenten über eine visuelle Oberfläche.

Dieser verwaltete Ansatz ermöglicht es Geschäftsteams, autonome Agenten schnell zu entwickeln und einzusetzen, ohne über fundiertes technisches Fachwissen in der KI-Modellentwicklung oder im Infrastrukturmanagement verfügen zu müssen.

Ein Beispiel aus der Praxis für Amazon Bedrock Agents

Eine in diesem AWS Blogbeitrag beschriebene Financial Operations (FinOps) -Lösung verwendet das Amazon Bedrock Multi-Agent-Framework, um einen KI-gesteuerten Cloud-Kostenmanagement-Assistenten zu erstellen. Das kostengünstige Amazon Nova Foundation-Modell unterstützt die Lösung, bei der ein zentraler FinOps Supervisor-Agent Aufgaben an spezialisierte Agenten delegiert. Diese Agenten rufen Ausgabendaten mithilfe von Daten zu AWS Ausgaben ab AWS Cost Explorer und generieren daraus Empfehlungen zur Kosteneinsparung. AWS Trusted Advisor

Das System umfasst sicheren Benutzerzugriff über Amazon Cognito, ein auf dem gehostetes Frontend AWS Amplify, und AWS Lambda Aktionsgruppen für Analysen und Prognosen in Echtzeit. Finanzteams können Fragen in natürlicher Sprache stellen, z. B. „Wie hoch waren meine Kosten im Februar 2025?“ Das System reagiert mit detaillierten Aufschlüsselungen, Optimierungsvorschlägen und Prognosen — alles innerhalb einer skalierbaren, serverlosen Architektur, die mithilfe von AWS CloudFormation

Amazon Grundgestein AgentCore

Amazon Bedrock AgentCore ist eine Agentenplattform für die sichere und skalierbare Entwicklung, Bereitstellung und den Betrieb hochleistungsfähiger Agenten unter Verwendung eines beliebigen Frameworks, Modells oder Protokolls. Mit AgentCore ihr können Sie Folgendes tun, und das alles ohne Infrastrukturmanagement:

- Erstellen Sie Agenten schneller.
- Ermöglichen Sie es Agenten, tool- und datenübergreifend Maßnahmen zu ergreifen.
- Führen Sie Agenten sicher mit niedriger Latenz und längeren Laufzeiten aus.
- Überwachen Sie Agenten in der Produktion.

AgentCore macht den undifferenzierten Aufwand beim Aufbau einer speziellen Agenteninfrastruktur überflüssig und ermöglicht es Ihnen, Ihre Agenten schneller zur Produktion zu bringen. Die Dienste können zusammen oder unabhängig voneinander genutzt werden und sind mit allen Frameworks kompatibel, einschließlichCrewAI, LangGraphLlamaIndex, undStrands Agents. AgentCoreist auch mit

jedem Fundamentmodell kompatibel, das innerhalb oder außerhalb von Amazon Bedrock erhältlich ist, und bietet so ultimative Flexibilität.

AgentCore besteht aus mehreren wichtigen Diensten:

- Amazon Bedrock AgentCore Runtime — Bietet eine sichere, serverlose, skalierbare Umgebung zum Hosten und Ausführen Ihrer Agenten, ohne dass Sie die Infrastruktur verwalten müssen, die für die Bereitstellung und Ausführung von KI-Agenten oder -Tools erforderlich ist.
- Amazon Bedrock AgentCore Memory — Bietet ein verwaltetes Speichersystem, das es Agenten ermöglicht, den Kontext von Interaktionen für persönlichere und kohärentere Konversationen beizubehalten, indem sowohl sofortiges als auch langfristiges Wissen erhalten bleibt.
- Amazon Bedrock AgentCore Gateway — Vereinfacht den Prozess der Erstellung, Sicherung und Suche nach den richtigen Tools für Agenten. Mit AgentCore Gateway können Entwickler Lambda-Funktionen und bestehende Dienste in Model Context Protocol (MCP) -kompatible Tools konvertieren APIs und sie Agenten zur Verfügung stellen.
- Amazon Bedrock AgentCore Identity — Bietet einen sicheren, skalierbaren Identitäts- und Zugriffsverwaltungsservice für Agenten, der die Entwicklung von KI-Agenten beschleunigt. Mit AgentCore Identity können Sie Agenten eindeutige, überprüfbare Identitäten zuweisen und so eine detaillierte Zugriffskontrolle und sichere agentengestützte Interaktionen mit Unternehmenssystemen ermöglichen.
- AgentCore Integrierte Tools von Amazon Bedrock — Ermöglicht es Ihnen, integrierte Tools zu verwenden, um Ihren Entwicklungs- und Test-Workflow zu verbessern. Verwenden Sie diese Tools, um effektiv mit Ihrer Anwendung zu interagieren, sodass KI-Agenten Code in Sandbox-Umgebungen sicher schreiben und ausführen können. Verwenden Sie das Browser-Tool, damit KI-Agenten in großem Umfang mit Websites interagieren können.
- Amazon Bedrock AgentCore Observability — Bietet Protokollierungs- und Überwachungsfunktionen, mit denen Sie in Echtzeit Einblick in die Leistung und das Verhalten Ihres Agenten erhalten, um Debugging und Optimierung zu erleichtern.

Die wichtigsten Funktionen von AgentCore

AgentCore umfasst die folgenden Hauptfunktionen:

- Vollständig verwaltet und erweiterbar — AgentCore ist ein vollständig verwalteter Service, d. h. er AWS kümmert sich um die zugrunde liegende Infrastruktur und Wartung. Er ist außerdem erweiterbar, sodass Sie die Funktionalität Ihrer Agenten anpassen und erweitern können. Weitere

Informationen finden Sie in der AgentCore Dokumentation unter Erste Schritte mit AgentCore Runtime.

- Langzeit- und Kurzzeitgedächtnis — Sorgen Sie für persönlichere und relevantere Interaktionen, indem Sie Agenten mit einem Speichersystem ausstatten, mit dem sie den Kontext aktueller Konversationen und langfristiges Wissen abrufen können. Weitere Informationen finden Sie in der AgentCore Dokumentation unter Erste Schritte mit dem AgentCore Gedächtnis.
- Vereinfachte Entwicklung und Integration von Tools — Ermöglichen Sie es Ihren Agenten, Tools über einen einzigen, sicheren Endpunkt zu finden und zu verwenden. Verwandeln Sie Ihre vorhandenen Unternehmensressourcen mit nur wenigen Codezeilen schnell in agentenfähige Tools, sodass sich Entwickler auf die Entwicklung einzigartiger Funktionen konzentrieren können. Weitere Informationen finden Sie in der Dokumentation unter Erste Schritte mit AgentCore Gateway. AgentCore
- Sichere und skalierbare Infrastruktur — AgentCore bietet eine sichere und skalierbare Umgebung für die Bereitstellung und den Betrieb von Agenten. Sie umfasst Funktionen für Identitäts- und Zugriffsmanagement, Datenverschlüsselung und Netzwerksicherheit. Weitere Informationen finden Sie in der AgentCore Dokumentation unter Erste Schritte mit AgentCore Identity.
- Integration mit einer Vielzahl von Tools — Ermöglicht es Ihnen, Ihre Agenten mit einer Vielzahl von Tools zu integrieren, darunter einen Codeinterpreter und ein Browser-Tool, das Sie mithilfe der AgentCore integrierten Tools erstellen können. Weitere Informationen finden Sie in der Dokumentation unter Erste Schritte mit AgentCore Code Interpreter und Erste Schritte mit dem AgentCore AgentCore Browser.
- Umfassende Beobachtbarkeit und Überwachung — Verschaffen Sie sich einen umfassenden Einblick in Ihre Agenten mit umfassenden Tools, mit denen Sie deren Leistung in der Produktion verfolgen, debuggen und überwachen können. Visualisieren Sie den gesamten Ausführungspfad des Agenten, um seine Argumentation zu überprüfen und Fehler zu beheben. Verwenden Sie Echtzeit-Dashboards und standardisierte Telemetriedaten, um wichtige Betriebskennzahlen zu verfolgen. Weitere Informationen finden Sie in der Dokumentation unter Hinzufügen von Observability zu Ihren Amazon AgentCore Bedrock-Ressourcen. AgentCore

Wann sollte es verwendet werden AgentCore

AgentCore eignet sich besonders gut für Szenarien mit autonomen Agenten, darunter:

- Organizations, die mit einem vollständig verwalteten Service, der sich um Infrastruktur, Sicherheit, integrierte Tools, Beobachtbarkeit und Skalierung kümmert, die Entwicklung beschleunigen und die Betriebskosten senken möchten
- Projekte, die Flexibilität mit modularen Services benötigen, die zusammen oder unabhängig voneinander funktionieren und mit jedem Framework wie CrewAI Oder LangGraph und jedem Basismodell aus jeder Quelle kompatibel sind
- Anwendungsfälle, die zustandsorientierte, dialogorientierte Agenten erfordern, die den Kontext wahren und aus vergangenen Interaktionen lernen müssen, um personalisierte und relevante Antworten geben zu können
- Agenten sind in der Lage, komplexe Aufgaben durch einfache Integration mit verschiedenen Anwendungen, Datenquellen und APIs

Implementierungsansatz für AgentCore

AgentCore ist für Unternehmen konzipiert, die KI-Agenten von der Machbarkeitsstudie, die mithilfe von Open Source- oder kundenspezifischen Agenten-Frameworks erstellt wurde, in die Produktion überführen möchten. Mit AgentCore können Unternehmen Folgendes tun:

- Stellen Sie Agenten sicher auf einer serverlosen Infrastruktur bereit, die jedes Framework und jedes Modell unterstützt, mit Sitzungsisolierung und integriertem Identitäts- und Zugriffsmanagement für end-to-end Sicherheit und Compliance. Mithilfe des Starter-Toolkits können Sie im Handumdrehen AgentCore Runtime-Agenten für führende Agenten-Frameworks erstellen.
- Verbessern Sie die Agents durch die Integration von persistentem Speicher zur Kontexterhaltung und vereinfachen Sie so die Entwicklung und Integration von Tools über AgentCore Gateway. Nutzen Sie das integrierte Browser-Tool und den Codeinterpreter für erweiterte Workflows.
- Verfolgen, debuggen und überwachen Sie KI-Agenten in der Produktion mithilfe von Observability-Dashboards, die von Amazon CloudWatch Application Insights unterstützt werdenOpenTelemetry, und verfolgen Sie wichtige AgentCore Ressourcenmetriken (Laufzeit, Speicher, Gateway und Tools).
- Beschleunigen Sie die Bereitstellung und Innovation mit vollständig verwalteten, modularen Services, zusammensetzbaren Blöcken zusammen oder unabhängig voneinander, mit beliebigen Agenten-Frameworks und Modellanbietern. Diese Flexibilität hilft Unternehmen dabei, schneller vom Prototyp zur Produktion überzugehen.

Dieser verwaltete Ansatz ermöglicht es Unternehmen, schnell und sicher KI-Agenten und Multi-Agenten-Systeme auf Unternehmensebene in jeder Größenordnung zu erstellen, bereitzustellen und auszuführen.

Ein Beispiel aus der Praxis für AgentCore

AWS hat beobachtet, dass eine der größten Banken Lateinamerikas AI/ML seit Jahren ein hyperpersonalisiertes und sicheres digitales Bankerlebnis anbietet. Die Bank erweitert ihre KI-Dienste, AgentCore um ihren Kunden intuitive Interaktionen, mehr Sicherheit und mehr Automatisierung zu bieten. Nach Angaben des CTO AgentCore wird erwartet, dass es ihre Bemühungen unterstützt, die Verpflichtungen der Kunden in großem Umfang zu erfüllen. AgentCore bietet ihren Entwicklern die Tools und die Flexibilität, um Agenten aufzubauen und zu verwalten, und trägt gleichzeitig zur Einhaltung der Finanzvorschriften bei.

Protokolle

KI-Agenten benötigen standardisierte Kommunikationsprotokolle, um mit anderen Agenten und Diensten interagieren zu können. Organizations, die Agentenarchitekturen implementieren, stehen vor großen Herausforderungen in Bezug auf Interoperabilität, Herstellerunabhängigkeit und Zukunftssicherheit ihrer Investitionen.

Dieser Abschnitt hilft Ihnen, sich in der agent-to-agent Protokolllandschaft zurechtzufinden, wobei der Schwerpunkt auf offenen Standards liegt, die Flexibilität und Interoperabilität maximieren. (Informationen zu agent-to-tool Protokollen finden Sie weiter unten in diesem Handbuch unter Strategie zur Integration von Tools.)

In diesem Abschnitt wird das Model Context Protocol (MCP) vorgestellt, ein offener Standard, der ursprünglich 2024 Anthropic entwickelt wurde. Heute unterstützt er MCP AWS aktiv durch Beiträge zur Entwicklung und Implementierung des Protokolls. AWS arbeitet mit führenden Open-Source-Agenten-Frameworks zusammen, darunter, und LangGraph CrewAILlamaIndex, um die future der Agentenkommunikation auf dem Protokoll zu gestalten. Weitere Informationen finden Sie unter Offene Protokolle für Agenten-Interoperabilität, Teil 1: Agentenübergreifende Kommunikation auf MCP (Blog).AWS

In diesem Abschnitt:

- Warum die Protokollauswahl wichtig ist
- Agent-to-agent Protokolle
- Auswahl von Agentenprotokollen
- Implementierungsstrategie für behördliche Protokolle
- Erste Schritte mit MCP
- ???

Warum die Protokollauswahl wichtig ist

Die Protokollauswahl bestimmt grundlegend, wie Sie Ihre KI-Agentenarchitektur aufbauen und weiterentwickeln können. Durch die Auswahl von Protokollen, die die Portabilität zwischen Agenten-Frameworks unterstützen, erhalten Sie die Flexibilität, verschiedene Agentensysteme und Workflows zu kombinieren, um Ihren spezifischen Anforderungen gerecht zu werden.

Offene Protokolle ermöglichen es Ihnen, Agenten über mehrere Frameworks hinweg zu integrieren. Verwenden Sie es beispielsweise LangChain für die schnelle Prototypenentwicklung und Implementierung von ProduktionssystemenStrands Agents, die über ein gemeinsames Protokoll wie MCP oder das Agent2Agent (A2A) -Protokoll kommunizieren. Diese Flexibilität reduziert die Abhängigkeit von bestimmten KI-Anbietern, vereinfacht die Integration in bestehende Systeme und ermöglicht es Ihnen, die Fähigkeiten der Agenten im Laufe der Zeit zu verbessern.

Gut konzipierte Protokolle sorgen außerdem für einheitliche Sicherheitsmuster für die Authentifizierung und Autorisierung in Ihrem gesamten Agenten-Ökosystem. Am wichtigsten ist jedoch, dass Sie dank der Protokollportabilität die Freiheit haben, neue Agenten-Frameworks und -Funktionen zu übernehmen, sobald diese verfügbar sind. Die Wahl offener Protokolle schützt Ihre Investitionen in die Agentenentwicklung und gewährleistet gleichzeitig die Interoperabilität mit Systemen von Drittanbietern.

Vorteile offener Protokolle

Wenn Sie Ihre eigenen Erweiterungen implementieren oder benutzerdefinierte Agentensysteme erstellen, bieten offene Protokolle überzeugende Vorteile:

- Dokumentation und Transparenz — Sorgen in der Regel für eine umfassende Dokumentation und transparente Implementierungen
- Community-Support — Zugang zu breiteren Entwickler-Communitys für Problemlösungen und Best Practices
- Interoperabilitätsgarantien — Bessere Gewissheit, dass Ihre Erweiterungen in verschiedenen Implementierungen funktionieren
- Künftige Kompatibilität — Geringeres Risiko, dass Änderungen nicht mehr funktionieren oder nicht mehr unterstützt werden
- Einfluss auf die Entwicklung — Gelegenheit, zur Protokollentwicklung beizutragen

Agent-to-agent Protokolle

Die folgende Tabelle bietet einen Überblick über Agentenprotokolle, mit denen mehrere Agenten zusammenarbeiten, Aufgaben delegieren und Informationen austauschen können.

Protocol (Protokoll)	Ideal für	Überlegungen

MCP-Kommunikation zwischen Agenten	Organizations, die nach flexiblen Mustern für die Zusammenarbeit mit Agenten suchen	• Eine von AWS dieser Organisation vorgeschl agene Erweiterung des Model Context Protocol (MCP) baut auf der bestehenden Kommunika tionsgrundlage agent-to-agent auf • Ermöglicht eine nahtlose Zusammenarbeit der Agenten mit OAuth basierter Sicherheit
A2A-Protokoll	Plattformübergreifende Agenten-Ökosysteme	• Unterstützt von Google • Neuerer Standard mit geringerer Akzeptanz als MCP

Entscheidung zwischen Protokolloptionen

Passen Sie bei der Implementierung der agent-to-agent Kommunikation Ihre spezifischen Kommunikationsanforderungen an die entsprechenden Protokollfunktionen an. Verschiedene Interaktionsmuster erfordern unterschiedliche Protokollfunktionen. In der folgenden Tabelle werden gängige Kommunikationsmuster beschrieben und die für jedes Szenario am besten geeigneten Protokolloptionen empfohlen.

Muster	Beschreibung	Ideale Protokollwahl
Einfache Anfrage und Antwort	Einmalige Interaktionen zwischen Agenten	MCP mit zustandslosen Datenströmen
Zustandsreiche Dialoge	Laufende Gespräche mit Kontext	MCP mit Sitzungsverwaltung
Zusammenarbeit mit mehreren Agenten	Komplexe Interaktionen zwischen mehreren Agenten	MCP-Interagent oder AutoGen

Teambasierte Workflows	Hierarchische Agententeams mit definierten Rollen	MCP-Interagent, oder CrewAI AutoGen

Neben Kommunikationsmustern können verschiedene technische und organisatorische Faktoren Ihre Protokollauswahl beeinflussen. In der folgenden Tabelle sind die wichtigsten Überlegungen aufgeführt, anhand derer Sie beurteilen können, welches Protokoll Ihren spezifischen Implementierungsanforderungen am ehesten entspricht.

Überlegung	Beschreibung	Beispiel
Sicherheitsmodell	Anforderungen an Authentifizierung und Autorisierung	OAuth 2.0 im MCP
Bereitstellungsumgebung	Wo Agenten arbeiten und kommunizieren	Verteilter Computer oder einzelner Computer
Kompatibilität mit Ökosystemen	Integration mit bestehenden Agenten-Frameworks	LangChain oder Strands Agents
Anforderungen an Skalierbarkeit	Erwartetes Wachstum der Agenteninteraktionen	Streaming-Funktionen von MCP

Auswahl von Agentenprotokollen

Für die meisten Unternehmen, die Produktionsagentensysteme entwickeln, bietet das Model Context Protocol (MCP) die umfassendste und am besten unterstützte Kommunikationsgrundlage. agent-to-agent MCP profitiert von aktiven Entwicklungsbeiträgen der AWS Open-Source-Community.

Die Auswahl der richtigen behördlichen Protokolle ist wichtig für Unternehmen, die agentische KI effektiv implementieren möchten. Die Überlegungen unterscheiden sich je nach organisatorischem Kontext.

Überlegungen zur Auswahl von behördlichen Protokollen

Organizations sollten bei der Auswahl von Protokollen für agentische KI-Systeme die folgenden bewährten Methoden berücksichtigen:

- Priorisieren Sie offene Standards — Organizations sollten offene Protokolle wie das MCP einführen, um langfristige Interoperabilität und Erweiterbarkeit zu gewährleisten und das Risiko einer Lieferantenbindung zu verringern.
- Balance zwischen Geschwindigkeit und Flexibilität — Startups und Early Adopters können mit gut unterstützten proprietären Protokollen für eine schnelle Entwicklung beginnen, sollten jedoch einen Migrationspfad zu offenen Standards definieren, sobald die Systeme ausgereift sind.
- Implementieren Sie Abstraktionsebenen — Unternehmen sollten Protokollabstraktion implementieren, um die Migration zu vereinfachen, die Einführung von Hybridanwendungen zu ermöglichen und Integrationsstrategien zukunftssicher zu gestalten.
- Betonen Sie Sicherheit und Compliance — Organizations in regulierten Branchen sollten Protokolle mit robusten Authentifizierungs-, Verschlüsselungs- und Auditfunktionen wählen, um die Governance- und Compliance-Anforderungen zu erfüllen.
- Evaluieren Sie den Reifegrad des Ökosystems — Alle Organisationen sollten den Zustand, die Akzeptanz und die Unterstützung jedes einzelnen Protokolls bewerten, um die Nachhaltigkeit sicherzustellen und die technische Verschuldung zu minimieren.
- Sich an der Entwicklung von Standards beteiligen — Organizations sollten sich an Normungsgremien oder Open-Source-Communities beteiligen, um die Entwicklung von Protokollen mitzugestalten und bewährte Verfahren zu beeinflussen.
- Berücksichtigung der Datensouveränität — Regierung und regulierte Sektoren sollten sicherstellen, dass die Protokollauswahl den Anforderungen an die Datenresidenz und Souveränität in den Einsatzregionen entspricht.
- Nutzung verwalteter Dienste — Verwenden Sie nach Möglichkeit verwaltete oder serverlose Implementierungen behördlicher Protokolle, um die betriebliche Komplexität zu reduzieren und die Bereitstellung zu beschleunigen.

Implementierungsstrategie für behördliche Protokolle

Um behördliche Protokolle effektiv in Ihrem Unternehmen zu implementieren, sollten Sie die folgenden strategischen Schritte in Betracht ziehen:

1. Beginnen Sie mit der Angleichung der Standards — Verwenden Sie, wo immer möglich, etablierte offene Protokolle.
2. Erstellen Sie Abstraktionsebenen — Implementieren Sie Adapter zwischen Ihren Systemen und spezifischen Protokollen.

3. Tragen Sie zu offenen Standards bei — Nehmen Sie an Gemeinschaften zur Protokollentwicklung teil.
4. Überwachen Sie die Protokollentwicklung — Bleiben Sie über neue Standards und Updates auf dem Laufenden.
5. Testen Sie die Interoperabilität regelmäßig — Stellen Sie sicher, dass Ihre Implementierungen kompatibel bleiben.

Erste Schritte mit MCP

AWS unterstützt aktiv das Model Context Protocol (MCP) durch Beiträge zur Entwicklung und Implementierung des Protokolls. AWS arbeitet mit führenden Open-Source-Agenten-Frameworks zusammen, darunter, und LangGraph CrewAILlamaIndex, um die future der Agentenkommunikation auf dem Protokoll zu gestalten.

Gehen Sie wie folgt vor, um das MCP in Ihrer Agentenarchitektur zu implementieren:

1. Erkunden Sie MCP-Implementierungen in Frameworks wie dem SDK. Strands Agents
2. Lesen Sie die technische Dokumentation zum Model Context Protocol.
3. Lesen Sie Offene Protokolle für Agenten-Interoperabilität, Teil 1: Kommunikation zwischen Agenten auf MCP (AWS Blog), um mehr über die Interoperabilität von Agenten zu erfahren.
4. Treten Sie der MCP-Community bei, um die Entwicklung des Protokolls zu beeinflussen.

MCP bietet eine Kommunikationsebene, die es Agenten ermöglicht, mit externen Daten und Diensten zu interagieren, und kann auch verwendet werden, um Agenten die Interaktion mit anderen Agenten zu ermöglichen. Die Streamable HTTP-Transportimplementierung des Protokolls bietet Entwicklern umfassende Interaktionsmuster, ohne das Rad neu erfinden zu müssen. Diese Muster unterstützen sowohl statuslose request/response Datenflüsse als auch statusbehaftetes Sitzungsmanagement mit persistenter Funktion. IDs

Durch die Einführung offener Protokolle wie MCP können Sie Ihr Unternehmen in die Lage versetzen, Agentensysteme aufzubauen, die flexibel, interoperabel und anpassungsfähig bleiben, wenn sich die KI-Technologie weiterentwickelt. Informationen zur agent-to-tool Protokollimplementierung finden Sie weiter unten in diesem Handbuch unter Strategie zur Tool-Integration.

Erste Schritte mit A2A

Das Agent2Agent (A2A) -Protokoll ermöglicht die dezentrale Zusammenarbeit zwischen Agenten über eine gemeinsame semantische Ebene. Anstatt die gesamte Arbeit über einen zentralen Orchestrator weiterzuleiten, ermöglicht A2A den Agenten, sich gegenseitig zu entdecken, für ihre Fähigkeiten zu werben, Aufgaben auszuhandeln und den Kontext mithilfe eines einfachen JSON-basierten Protokolls gemeinsam zu nutzen. Jeder Agent veröffentlicht ein Funktionsmanifest.

Das folgende Beispiel zeigt ein vereinfachtes A2A-Fähigkeitsmanifest, das die unterstützten Aktionen, erforderlichen Eingaben und Betriebsmetadaten eines Agenten ankündigt, um die Erkennung und Aushandlung von Aufgaben zu ermöglichen:

```
{
  "can": ["summarize.text", "extract.keywords"],
  "needs": ["document.input"],
  "meta": { "version": "1.0.3", "latencyMs": 120 }
}
```

Dieses Modell ermöglicht einen dynamischen Kapazitätsabgleich, die Delegierung von Aufgaben während der Ausführung und die organisationsübergreifende Zusammenarbeit. Agenten können sich anhand von Aufgaben selbst organisieren, temporäre Arbeitsgruppen bilden und sich anpassen, wenn neue Funktionen in das System eingeführt oder verlassen werden.

A2A unterstützt Interaktionen, die von einfachen Anfragen ohne Status bis hin zu mehrstufigen Verhandlungssitzungen reichen, darunter:

- peer-to-peerDirektnachrichten für Zusammenarbeit mit niedriger Latenz
- Semantische Aufgabenaushandlung, bei der Agenten den am besten geeigneten Peer auswählen
- Auf Fähigkeiten basierende Entdeckung, die eine sich abzeichnende Arbeitsteilung ermöglicht
- Sitzungsverankerung für zustandsbehaftete Interaktionen in mehreren Schritten

Durch die Einführung offener, agenteneigener Protokolle wie A2A schaffen Unternehmen KI-Systeme, die modular und interoperabel sind und eine grenzüberschreitende Zusammenarbeit ermöglichen. A2A stellt sicher, dass die Agenten-Ökosysteme flexibel bleiben und sich weiterentwickeln können, wenn neue Agenten, Teams oder externe Systeme eingeführt werden, ohne dass starre Orchestrierungsebenen oder vorherige Kopplungen erforderlich sind.

Gehen Sie wie folgt vor, um das A2A-Protokoll in Ihrer Agentenarchitektur zu implementieren:

1. Lesen Sie die A2A-Protokollspezifikation — Lesen Sie die neueste Version der Agent2Agent (A2A) -Protokollspezifikation, um zu erfahren, wie Funktionen funktionieren, wie die Verhandlungsabläufe ablaufen und der Agent-Handshake funktionieren.
2. Erkunden Sie A2A-kompatible Laufzeiten — Evaluieren Sie Frameworks wie das Strands Agents SDK oder benutzerdefinierte Runtime-Layer, die Funktionsmanifeste und -aushandlungen im A2A-Stil unterstützen. peer-to-peer
3. Implementieren Sie ein Funktionsmanifest für Ihre Agenten — Definieren Sie die `meta` Felder und Felder der einzelnen Agenten, um die Erkennung `canneeds`, das Matchmaking und die Zusammenarbeit auf Absichtsebene zu ermöglichen.
4. Experimentieren Sie mit A2A-Verhandlungsmustern — nutzen Sie die Anfrage-Angebot-Annahme-Schleife, strukturierte Funktionsabfragen oder eine auf Klatsch und Tratsch basierende Erkennung, um zu verstehen, wie Agenten entscheiden, wer eine Aufgabe bearbeiten soll.
5. Testen Sie A2A in einer gemischten Infrastrukturumgebung — Kombinieren Sie A2A-Peer-Negotiation mit Event-Routing, das standardmäßig AWS über Amazon erfolgt, um hybride Koordinationsmuster EventBridge zu bewerten.
6. Treten Sie der A2A-Community bei — Nehmen Sie an der offenen Arbeitsgruppe teil, um über Erweiterungen, Sicherheitsempfehlungen und herstellerübergreifende Interoperabilitätsverbesserungen auf dem Laufenden zu bleiben und zur Entwicklung des Protokolls beizutragen.

Tools

KI-Agenten bieten Mehrwert, indem sie mit externen Tools und Datenquellen interagieren, um nützliche Aufgaben auszuführen APIs. Die richtige Strategie zur Tool-Integration wirkt sich direkt auf die Fähigkeiten, die Sicherheitslage und die langfristige Flexibilität Ihres Agenten aus.

Dieser Abschnitt hilft Ihnen, sich in der Tool-Integrationslandschaft zurechtzufinden, wobei der Schwerpunkt auf offenen Standards liegt, die Ihre Freiheit und Flexibilität maximieren. In diesem Abschnitt wird das Model Context Protocol (MCP) für die Toolintegration vorgestellt und Framework-spezifische Tools und spezialisierte Meta-Tools zur Verbesserung der Arbeitsabläufe von Agenten besprochen.

In diesem Abschnitt:

Kategorien von Tools

Building Agent Systems umfasst drei Hauptkategorien von Tools.

Auf Protokollen basierende Tools

Protokollbasierte Tools verwenden standardisierte Protokolle für die Kommunikation: agent-to-tool

- MCP-Tools — Offene Standardtools, die in allen Frameworks funktionieren und sowohl lokale als auch Remote-Ausführungsoptionen bieten.
- OpenAIFunktionsaufruf — Proprietäre Tools, die spezifisch für OpenAI Modelle sind.
- AnthropicTools — Eine Variante des OpenAI Funktionsaufrufs für proprietäre Tools, die spezifisch für Anthropic Claude-Modelle sind.

Framework-native Tools

Framework-native Tools sind direkt in spezifische Agenten-Frameworks integriert:

- Strands Agents Tools — Leichte quick-to-implement Tools, die spezifisch für das Strands Agents Framework sind.
- LangChainPythonauf Tools basierende Tools, die eng in das LangChain Ökosystem integriert sind.
- LlamaIndexTools — Tools, die für den Datenabruf und die interne LlamaIndex Verarbeitung optimiert sind.

Meta-Tools

Meta-Tools verbessern die Arbeitsabläufe der Agenten, ohne direkt externe Maßnahmen ergreifen zu müssen:

- Workflow-Tools — Verwalten Sie den Ausführungsablauf der Agenten, die Verzweigungslogik und die Statusverwaltung.
- Tools für Agentendiagramme — Koordinieren Sie mehrere Agenten in komplexen Workflows.
- Speichertools — Sorgen für die persistente Speicherung und den dauerhaften Abruf von Informationen über Agentensitzungen hinweg.
- Reflexionstools — Ermöglichen es Agenten, ihre eigene Leistung zu analysieren und zu verbessern.

Auf Protokollen basierende Tools

Wenn es um protokollbasierte Tools geht, bietet das Model Context Protocol (MCP) die umfassendste und flexibelste Grundlage für die Toolintegration. Wie im AWS Open-Source-Blogbeitrag zur Interoperabilität von Agenten dargelegt, AWS hat sich das Unternehmen MCP als strategisches Protokoll zu eigen gemacht und aktiv zu seiner Entwicklung beigetragen.

In der folgenden Tabelle werden die Optionen für die Bereitstellung des MCP-Tools beschrieben.

Bereitstellungsmodell	Beschreibung	Ideal für	Implementierung

Lokales Studio	Tools werden im gleichen Prozess wie der Agent ausgeführt	Entwicklung, Testen und einfache Tools	Schnelle Implementierung ohne Netzwerk-Overhead
Auf lokalen serverges tützten Ereignissen (SSE)	Tools werden lokal ausgeführt, kommunizieren aber über HTTP	Komplexere lokale Tools mit getrennten Aufgabenbereichen	Bessere Isolierung, aber immer noch geringe Latenz
Remote-HTTP-Streamingfähig	Tools werden auf Remote-Servern ausgeführt	Produktionsumgebungen und gemeinsam genutzte Tools	Skalierbar und zentral verwaltet

Für die Erstellung von SDKs MCP-Tools stehen die offiziellen MCP zur Verfügung:

- PythonSDK — Umfassende Implementierung mit vollständiger Protokollunterstützung
- TypeScriptSDK —JavaScript//TypeScriptImplementierung für Webanwendungen
- JavaSDK — Java-Implementierung für Unternehmensanwendungen

Diese SDKs bieten die Bausteine für die Erstellung MCP-kompatibler Tools in Ihrer bevorzugten Sprache mit konsistenten Implementierungen der Protokollspezifikation.

Darüber hinaus AWS hat MCP im SDK implementiert. Strands Agents Das Strands Agents SDK bietet eine einfache Möglichkeit, MCP-kompatible Tools zu erstellen und zu verwenden. Eine umfassende Dokumentation ist im Repository verfügbar. Strands Agents GitHub Für einfachere Anwendungsfälle oder wenn Sie außerhalb des Strands Agents Frameworks arbeiten, SDKs bietet das offizielle MCP direkte Implementierungen des Protokolls in mehreren Sprachen an.

Sicherheitsfunktionen von MCP-Tools

Zu den Sicherheitsfunktionen der MCP-Tools gehören:

- OAuth 2.0/2.1-Authentifizierung — Authentifizierung nach Industriestandard
- Umfang der Berechtigungen — Präzise Zugriffskontrolle für Tools
- Erkennung der Funktionen von Tools — Dynamische Erkennung verfügbarer Tools
- Strukturierte Fehlerbehandlung — Konsistente Fehlermuster

Erste Schritte mit MCP-Tools

Gehen Sie wie folgt vor, um MCP für die Toolintegration zu implementieren:

1. Erkunden Sie das Strands AgentsSDK für eine produktionsreife MCP-Implementierung.
2. Lesen Sie die technische Dokumentation zu MCP, um die Kernkonzepte zu verstehen.
3. Verwenden Sie die in diesem AWS Open-Source-Blogbeitrag beschriebenen praktischen Beispiele.
4. Beginnen Sie mit einfachen lokalen Tools, bevor Sie zu Remote-Tools übergehen.
5. Treten Sie der MCP-Community bei, um die Entwicklung des Protokolls zu beeinflussen.

Erkunden Sie Gateway AgentCore

Amazon Bedrock AgentCore Gateway bietet Entwicklern eine einfache und sichere Möglichkeit, MCP-Tools und andere Zielendpunkte in großem Umfang zu entwickeln, bereitzustellen, zu entdecken und eine Verbindung zu ihnen herzustellen. Mit AgentCore Gateway können Entwickler AWS Lambda Funktionen und bestehende Services in MCP-kompatible Tools umwandeln APIs. Anschließend können sie diese Tools mit nur wenigen Codezeilen Agenten über AgentCore Gateway-Endpunkte zur Verfügung stellen. AgentCore Gateway unterstützt OpenAPISmithy, und Lambda als Eingabetypen und ist die einzige Lösung, die sowohl eine umfassende Eingangs- als auch Ausgangsauthentifizierung in einem vollständig verwalteten Service bietet.

Framework-native Tools

Obwohl das Model Context Protocol (MCP) die flexibelste Grundlage bietet, bieten Framework-native Tools Vorteile für bestimmte Anwendungsfälle.

Das Strands AgentsSDK bietet Python basierte Tools, die sich durch ihr leichtes Design auszeichnen und nur minimalen Aufwand für einfache Operationen erfordern. Sie ermöglichen eine schnelle Implementierung und ermöglichen es Entwicklern, Tools mit nur wenigen Codezeilen zu erstellen. Darüber hinaus sind sie eng integriert, sodass sie nahtlos in das Strands Agents Framework integriert werden können.

Das folgende Beispiel zeigt, wie Sie mit Hilfe von ein einfaches Wetter-Tool erstellenStrands Agents. Entwickler können Python Funktionen schnell und mit minimalem Codeaufwand in Tools umwandeln, auf die Agenten zugreifen können, und automatisch die entsprechende Dokumentation aus dem Docstring der Funktion generieren.

```
#Example of a simple Strands native tool

@tool

def weather(location: str) -> str:

"""Get the current weather for a location""" #

Implementation here

return f"The weather in {location} is sunny."
```

Für schnelles Prototyping oder einfache Anwendungsfälle können Framework-native Tools die Entwicklung beschleunigen. Für Produktionssysteme bieten MCP-Tools jedoch eine bessere Interoperabilität und future Flexibilität als Framework-native Tools.

Die folgende Tabelle bietet einen Überblick über andere Framework-spezifische Tools.

Framework	Art des Werkzeugs	Vorteile	Überlegungen
AutoGen	Funktionsdefinitionen	Starke Unterstützung für mehrere Agenten	MicrosoftÖkosystem
LangChain	PythonKlassen	Großes Ökosystem vorgefertigter Tools	Bindung an ein Framework
LlamaIndex	Python-Funktionen	Optimiert für Datenoperationen	Beschränkt auf LlamaIndex

Meta-Tools

Meta-Tools interagieren nicht direkt mit externen Systemen. Stattdessen verbessern sie die Fähigkeiten der Agenten, indem sie Agentenmuster implementieren. In diesem Abschnitt werden Workflows, Agentendiagramme und Speicher-Metatools behandelt.

Workflow-Metatools

Workflow-Metatools verwalten den Ablauf der Agentenausführung:

- Statusverwaltung — Behalten Sie den Kontext über mehrere Agenteninteraktionen hinweg bei

- Verzweigungslogik — Aktivieren Sie bedingte Ausführungspfade
- Wiederholungsmechanismen — Behandeln Sie Fehler mit ausgeklügelten Wiederholungsstrategien

Zu den Beispiel-Frameworks mit Workflow-Meta-Tools gehören auch LangGraphWorkflow-Funktionen. Strands Agents

Metatools für Agentengraphen

Die Metatools für Agentengraphen koordinieren die Zusammenarbeit mehrerer Agenten:

- Delegierung von Aufgaben — Weisen Sie spezialisierten Agenten Unteraufgaben zu
- Ergebnisaggregation — Kombinieren Sie die Ergebnisse mehrerer Agenten
- Konfliktlösung — Beilegung von Meinungsverschiedenheiten zwischen Agenten

Frameworks sind auf die Graphkoordination von Agenten CrewAIspezialisiert AutoGenund haben sich darauf spezialisiert.

Speicher-Metatools

Speicher-Metatools ermöglichen persistentes Speichern und Abrufen:

- Konversationsverlauf — Behalten Sie den Kontext zwischen den Sitzungen bei
- Wissensdatenbanken — Speichern und Abrufen domänenspezifischer Informationen
- Vektorspeicher — Ermöglichen semantische Suchfunktionen

Das Ressourcensystem von MCP bietet eine standardisierte Methode zur Implementierung von Speicher-Metatools, die über verschiedene Agenten-Frameworks hinweg funktionieren.

Strategie zur Integration von Tools

Ihre Wahl der Tool-Integrationsstrategie wirkt sich direkt darauf aus, was Ihre Agenten erreichen können und wie einfach sich Ihr System weiterentwickeln kann. Priorisieren Sie offene Protokolle wie das Model Context Protocol (MCP) und setzen Sie gleichzeitig Framework-native Tools und Meta-Tools strategisch ein. Auf diese Weise können Sie ein Tool-Ökosystem aufbauen, das auch im Zuge der Weiterentwicklung der KI-Technologie flexibel und leistungsstark bleibt.

Der folgende strategische Ansatz zur Toolintegration maximiert die Flexibilität und erfüllt gleichzeitig die unmittelbaren Bedürfnisse Ihres Unternehmens:

1. Nutzen Sie MCP als Grundlage — MCP bietet eine standardisierte Möglichkeit, Agenten mit Tools mit starken Sicherheitsfunktionen zu verbinden. Beginnen Sie mit MCP als primärem Tool-Protokoll für:
   - Strategische Tools, die in mehreren Agentenimplementierungen eingesetzt werden.
   - Sicherheitsempfindliche Tools, die eine zuverlässige Authentifizierung und Autorisierung erfordern.
   - Tools, die in Produktionsumgebungen remote ausgeführt werden müssen.
2. Verwenden Sie gegebenenfalls Framework-native Tools — Ziehen Sie Framework-native Tools in Betracht für:
   - Schnelles Prototyping während der ersten Entwicklung.
   - Einfache, unkritische Tools mit minimalen Sicherheitsanforderungen.
   - Framework-spezifische Funktionen, die einzigartige Funktionen nutzen.
3. Implementieren Sie Meta-Tools für komplexe Workflows — Fügen Sie Meta-Tools hinzu, um Ihre Agentenarchitektur zu verbessern:
   - Beginnen Sie einfach mit grundlegenden Workflow-Mustern.
   - Erhöhen Sie die Komplexität, wenn Ihre Anwendungsfälle immer ausgereifter werden.
   - Standardisieren Sie die Schnittstellen zwischen Agenten und Meta-Tools.
4. Für die Entwicklung planen — mit Blick auf future Flexibilität bauen:
   - Dokumentieren Sie die Benutzeroberflächen der Tools unabhängig von den Implementierungen.
   - Erstellen Sie Abstraktionsebenen zwischen Agenten und Tools.
   - Richten Sie Migrationspfade von proprietären zu offenen Protokollen ein.

Bewährte Sicherheitsverfahren für die Integration von Tools

Die Tool-Integration wirkt sich direkt auf Ihre Sicherheitslage aus. In diesem Abschnitt werden bewährte Methoden beschrieben, die Sie für Ihr Unternehmen berücksichtigen sollten.

Authentifizierung und Autorisierung

Nutzen Sie die folgenden robusten Zugriffskontrollen:

- Verwenden Sie OAuth 2.0/2.1 — Implementieren Sie die branchenübliche Authentifizierung für Remote-Tools.
- Implementierung der geringsten Rechte — Gewähren Sie Tools nur die Berechtigungen, die sie benötigen.
- Zugangsdaten wechseln — Aktualisieren Sie regelmäßig API-Schlüssel und Zugriffstoken.

Datenschutz

Ergreifen Sie zum Schutz Ihrer Daten die folgenden Maßnahmen:

- Eingaben und Ausgaben validieren — Implementieren Sie die Schemavalidierung für alle Werkzeuginteraktionen.
- Vertrauliche Daten verschlüsseln — Verwenden Sie TLS für die gesamte Remote-Tool-Kommunikation.
- Implementieren Sie Datenminimierung — Geben Sie nur die erforderlichen Informationen an die Tools weiter.

Überwachung und Prüfung

Sorgen Sie mit den folgenden Mechanismen für Transparenz und Kontrolle:

- Protokollieren Sie alle Tool-Aufrufe — Pflegen Sie umfassende Prüfprotokolle.
- Auf Anomalien achten — Erkennen Sie ungewöhnliche Nutzungsmuster von Tools.
- Implementieren Sie eine Ratenbegrenzung — Beugen Sie Missbrauch durch übermäßige Tool-Aufrufe vor.

Das Sicherheitsmodell Model Context Protocol (MCP) geht umfassend auf diese Bedenken ein. Weitere Informationen finden Sie in der MCP-Dokumentation unter Überlegungen zur Sicherheit.

Schlussfolgerung

Die Landschaft der agentischen KI entwickelt sich weiterhin rasant und bietet Unternehmen leistungsstarke neue Möglichkeiten zum Aufbau intelligenter, autonomer Systeme. In diesem Leitfaden wurden drei wesentliche Komponenten für eine erfolgreiche Implementierung untersucht: Frameworks, die die Grundlage bilden, Plattformen, die die Umgebung bereitstellen, Protokolle, die die Kommunikation ermöglichen, und Tools, die die Funktionen erweitern.

Mit zunehmender Reife der Frameworks können Sie mit einer erhöhten Interoperabilität, einer Standardisierung von Protokollen wie dem Model Context Protocol (MCP) und ausgefeilteren Orchestrierungsfunktionen für autonome Agenten rechnen. Organizations, die sich heute mit diesen Frameworks auskennen, sind gut positioniert, um zunehmend autonome, intelligente Agenten zu entwickeln, die einen erheblichen Geschäftswert bieten.

Plattformen bieten die Ausführungs-, Verwaltungs- und Lebenszyklusumgebung, in der agentische Systeme betrieben werden. Sie kümmern sich um Themen wie Identität, Sicherheitsgrenzen, Beobachtbarkeit, Speicherverwaltung, Sitzungsbindung und sichere Interaktion mit Tools und Daten. In AWS Umgebungen ermöglichen Plattformen wie Managed Agent Runtimes und Orchestration Services Unternehmen, autonome Agenten und Agentensysteme in großem Umfang einzusetzen, zu überwachen, weiterzuentwickeln und zu steuern. Plattformen verbinden grundlegende Frameworks mit realen betrieblichen Anforderungen.

Die Wahl der Agentenprotokolle stellt eine strategische Entscheidung dar, bei der unmittelbare Entwicklungsanforderungen mit langfristiger Flexibilität und Interoperabilität in Einklang gebracht werden. Durch die Priorisierung offener Protokolle und die Schaffung geeigneter Abstraktionsebenen können Unternehmen Agentensysteme aufbauen, die an sich entwickelnde Technologien anpassbar bleiben und gleichzeitig den aktuellen Geschäftsanforderungen entsprechen.

Für die meisten Unternehmen stellt MCP aufgrund seines offenen Standards, seines wachsenden Ökosystems, der Unterstützung von agent-to-agent Kommunikationsmustern und der Fähigkeit zur Toolintegration eine solide Grundlage dar. AWS hat MCP und Agent2Agent (A2A) als strategische Protokolle eingeführt und aktiv zu deren Entwicklung und Implementierung in Diensten wie dem SDK beigetragen. Strands Agents Durch die Verwendung von MCP oder A2A zusammen mit geeigneten Framework-nativen Tools und Meta-Tools können Sie Agentensysteme erstellen, die sofort einen Mehrwert bieten und gleichzeitig an future Innovationen anpassbar sind.

Ressourcen

Nutzen Sie die folgenden AWS und andere Ressourcen im Zusammenhang mit der Entwicklung autonomer Agenten.

AWS Blogs

- Amazon Bedrock AgentCore Memory: Aufbau kontextsensitiver Agenten
- Bewährte Methoden für die Erstellung robuster generativer KI-Anwendungen mit Amazon Bedrock Agents — Teil 1
- Bewährte Methoden für die Erstellung robuster generativer KI-Anwendungen mit Amazon Bedrock Agents — Teil 2
- Erstellen Sie leistungsstarke RAG-Pipelines mit Amazon LlamaIndex Bedrock
- Bauen Sie mit Amazon Bedrock AgentCore Observability vertrauenswürdige KI-Agenten auf
- Evaluieren Sie die Antworten von RAG mit Amazon Bedrock LlamaIndex und RAGAS
- Wir stellen vor: den Amazon Bedrock AgentCore Code Interpreter
- Wir stellen vor: Amazon Bedrock AgentCore Gateway: Transformation der Entwicklung von KI-Agententools für Unternehmen
- Wir stellen vor: Amazon Bedrock AgentCore Identity: Sicherung agentischer KI im großen Maßstab
- Wir stellen vor: Strands Agents ein Open-Source-SDK für KI-Agenten
- Offene Protokolle für Agenten-Interoperabilität Teil 1: Agentenübergreifende Kommunikation auf MCP
- Starten und skalieren Sie Ihre Agenten und Tools sicher auf Amazon Bedrock Runtime AgentCore
- AWS Transform für.NET, den ersten agentischen KI-Service für die Modernisierung von.NET-Anwendungen im großen Maßstab
- AWS Wöchentliche Zusammenfassung: Strands Agents

AWS Präskriptive Leitlinien

- Operationalisierung der Agenten-KI am AWS
- Grundlagen der Agenten-KI auf AWS
- Agentengestützte KI-Muster und Workflows auf AWS

- Aufbau serverloser Architekturen für agentische KI auf AWS
- Aufbau von Mehrmandantenarchitekturen für agentische KI auf AWS
- Sicherheit für Agenten-KI auf AWS
- Optionen und Architekturen für Augmented Generation abrufen auf AWS

AWS Ressourcen

- Dokumentation zu Amazon Bedrock
- Dokumentation zu Amazon Bedrock AgentCore
- Amazon Bedrock AgentCore Starter Toolkit (Repository) GitHub
- Amazon Nova-Dokumentation
- AWS MCP-Server (GitHubRepository)

Sonstige Ressourcen

- AutoGenDokumentation () Microsoft
- Wirksame Agenten aufbauen (Anthropic)
- CrewAI GitHubEndlager
- LangChain-Dokumentation
- LangGraphPlattform
- LlamaIndex-Dokumentation
- Dokumentation zum Model Context Protocol
- Strands Agents-Dokumentation
- Strands AgentsÜberblick über die Tools
- Strands AgentsSchnellstart-Anleitung

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
Neuer Abschnitt	Abschnitt „Plattformen“ hinzugefügt	16. Januar 2026
Erste Veröffentlichung	—	14. Juli 2025

AWS Glossar zu präskriptiven Leitlinien

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

Zahlen

7 Rs

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

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

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

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

A

ABAC

Siehe attributbasierte Zugriffskontrolle.

abstrahierte Dienste

Siehe Managed Services.

ACID

Siehe Atomarität, Konsistenz, Isolierung und Haltbarkeit.

Aktiv-Aktiv-Migration

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

Aktiv-Passiv-Migration

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

Aggregatfunktion

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

AI

Siehe künstliche Intelligenz.

AIOps

Siehe Operationen im Bereich künstliche Intelligenz.

Anonymisierung

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

Anti-Muster

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

Anwendungssteuerung

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

Anwendungsportfolio

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

künstliche Intelligenz (KI)

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

Operationen mit künstlicher Intelligenz (AIOps)

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

Asymmetrische Verschlüsselung

Ein Verschlüsselungsalgorithmus, der ein Schlüsselpaar, einen öffentlichen Schlüssel für die Verschlüsselung und einen privaten Schlüssel für die Entschlüsselung verwendet. Sie können den öffentlichen Schlüssel teilen, da er nicht für die Entschlüsselung verwendet wird. Der Zugriff auf den privaten Schlüssel sollte jedoch stark eingeschränkt sein.

Atomizität, Konsistenz, Isolierung, Haltbarkeit (ACID)

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

Attributbasierte Zugriffskontrolle (ABAC)

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

autoritative Datenquelle

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

Availability Zone

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

AWS Framework für die Einführung der Cloud (AWS CAF)

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

Cloud-Einführung vorzubereiten. Weitere Informationen finden Sie auf der AWS -CAF-Webseite und dem AWS -CAF-Whitepaper.

AWS Workload-Qualifizierungsrahmen (AWS WQF)

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

B

schlechter Bot

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

BCP

Siehe Planung der Geschäftskontinuität.

Verhaltensdiagramm

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

Big-Endian-System

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

Binäre Klassifikation

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

Bloom-Filter

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

Blau/Grün-Bereitstellung

Eine Bereitstellungsstrategie, bei der Sie zwei separate, aber identische Umgebungen erstellen. Sie führen die aktuelle Anwendungsversion in einer Umgebung (blau) und die neue Anwendungsversion in der anderen Umgebung (grün) aus. Mit dieser Strategie können Sie schnell und mit minimalen Auswirkungen ein Rollback durchführen.

Bot

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

Botnetz

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

branch

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

Zugang durch Glasbruch

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

Brownfield-Strategie

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

Puffer-Cache

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

Geschäftsfähigkeit

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

Planung der Geschäftskontinuität (BCP)

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

C

CAF

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

Bereitstellung auf Kanaren

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

CCoE

Weitere Informationen finden Sie im Cloud Center of Excellence.

CDC

Siehe Erfassung von Änderungsdaten.

Erfassung von Datenänderungen (CDC)

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

Chaos-Technik

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

CI/CD

Siehe Continuous Integration und Continuous Delivery.

Klassifizierung

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

clientseitige Verschlüsselung

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

Cloud-Exzellenzzentrum (CCoE)

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

Cloud Computing

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

Cloud-Betriebsmodell

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

Phasen der Einführung der Cloud

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

- Projekt – Durchführung einiger Cloud-bezogener Projekte zu Machbarkeitsnachweisen und zu Lernzwecken
- Fundament — Tätigen Sie grundlegende Investitionen, um Ihre Cloud-Einführung zu skalieren (z. B. Einrichtung einer landing zone, Definition eines CCo E, Einrichtung eines Betriebsmodells)
- Migration – Migrieren einzelner Anwendungen
- Neuentwicklung – Optimierung von Produkten und Services und Innovation in der Cloud

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

CMDB

Siehe Datenbank für das Konfigurationsmanagement.

Code-Repository

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

Kalter Cache

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

Kalte Daten

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

Computer Vision (CV)

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

Drift in der Konfiguration

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

Verwaltung der Datenbankkonfiguration (CMDB)

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

Konformitätspaket

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

Kontinuierliche Bereitstellung und kontinuierliche Integration (CI/CD)

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

CV

Siehe Computer Vision.

D

Daten im Ruhezustand

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

Datenklassifizierung

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

Datendrift

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

Daten während der Übertragung

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

Datennetz

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

Datenminimierung

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

Datenperimeter

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

Vorverarbeitung der Daten

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

Herkunft der Daten

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

betroffene Person

Eine Person, deren Daten gesammelt und verarbeitet werden.

Data Warehouse

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

Datenbankdefinitionssprache (DDL)

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

Datenbankmanipulationssprache (DML)

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

DDL

Siehe Datenbankdefinitionssprache.

Deep-Ensemble

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

Deep Learning

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

defense-in-depth

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

delegierter Administrator

In AWS Organizations kann ein kompatibler Dienst ein AWS Mitgliedskonto registrieren, um die Konten der Organisation und die Berechtigungen für diesen Dienst zu verwalten. Dieses Konto wird als delegierter Administrator für diesen Service bezeichnet. Weitere Informationen und eine Liste kompatibler Services finden Sie unter Services, die mit AWS Organizations funktionieren in der AWS Organizations -Dokumentation.

Einsatz

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

Entwicklungsumgebung

Siehe Umgebung.

Detektivische Kontrolle

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

Abbildung des Wertstroms in der Entwicklung (DVSM)

Ein Prozess zur Identifizierung und Priorisierung von Einschränkungen, die sich negativ auf Geschwindigkeit und Qualität im Lebenszyklus der Softwareentwicklung auswirken. DVSM erweitert den Prozess der Wertstromanalyse, der ursprünglich für Lean-Manufacturing-Praktiken

konzipiert wurde. Es konzentriert sich auf die Schritte und Teams, die erforderlich sind, um durch den Softwareentwicklungsprozess Mehrwert zu schaffen und zu steigern.

digitaler Zwilling

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

Maßtabelle

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

Katastrophe

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

Notfallwiederherstellung (DR)

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

DML

Siehe Sprache zur Datenbankmanipulation.

Domainorientiertes Design

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

DR

Siehe Disaster Recovery.

Erkennung von Driften

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

DVSM

Siehe Abbildung des Wertstroms in der Entwicklung.

E

EDA

Siehe explorative Datenanalyse.

EDI

Siehe elektronischer Datenaustausch.

Edge-Computing

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

elektronischer Datenaustausch (EDI)

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

Verschlüsselung

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

Verschlüsselungsschlüssel

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

Endianismus

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

Endpunkt

Siehe Service-Endpunkt.

Endpunkt-Services

Ein Service, den Sie in einer Virtual Private Cloud (VPC) hosten können, um ihn mit anderen Benutzern zu teilen. Sie können einen Endpunktdienst mit anderen AWS-Konten oder AWS Identity and Access Management (IAM AWS PrivateLink -) Prinzipalen erstellen und diesen Berechtigungen gewähren. Diese Konten oder Prinzipale können sich privat mit Ihrem Endpunktservice verbinden, indem sie Schnittstellen-VPC-Endpunkte erstellen. Weitere Informationen finden Sie unter Einen Endpunkt-Service erstellen in der Amazon Virtual Private Cloud (Amazon VPC)-Dokumentation.

Unternehmensressourcenplanung (ERP)

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

Envelope-Verschlüsselung

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

Umgebung

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

- Entwicklungsumgebung – Eine Instance einer laufenden Anwendung, die nur dem Kernteam zur Verfügung steht, das für die Wartung der Anwendung verantwortlich ist. Entwicklungsumgebungen werden verwendet, um Änderungen zu testen, bevor sie in höhere Umgebungen übertragen werden. Diese Art von Umgebung wird manchmal als Testumgebung bezeichnet.
- Niedrigere Umgebungen – Alle Entwicklungsumgebungen für eine Anwendung, z. B. solche, die für erste Builds und Tests verwendet wurden.

- Produktionsumgebung – Eine Instance einer laufenden Anwendung, auf die Endbenutzer zugreifen können. In einer CI/CD Pipeline ist die Produktionsumgebung die letzte Bereitstellungsumgebung.
- Höhere Umgebungen – Alle Umgebungen, auf die auch andere Benutzer als das Kernentwicklungsteam zugreifen können. Dies kann eine Produktionsumgebung, Vorproduktionsumgebungen und Umgebungen für Benutzerakzeptanztests umfassen.

Epics

In der agilen Methodik sind dies funktionale Kategorien, die Ihnen helfen, Ihre Arbeit zu organisieren und zu priorisieren. Epics bieten eine allgemeine Beschreibung der Anforderungen und Implementierungsaufgaben. Zu den Sicherheitsepen AWS von CAF gehören beispielsweise Identitäts- und Zugriffsmanagement, Detektivkontrollen, Infrastruktursicherheit, Datenschutz und Reaktion auf Vorfälle. Weitere Informationen zu Epics in der AWS -Migrationsstrategie finden Sie im Leitfaden zur Programm-Implementierung.

ERP

Siehe Enterprise Resource Planning.

Explorative Datenanalyse (EDA)

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

F

Faktentabelle

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

schnell scheitern

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

Grenze zur Fehlerisolierung

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

Feature-Zweig

Siehe Zweig.

Features

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

Bedeutung der Feature

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

Featuretransformation

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

Eingabeaufforderung mit wenigen Klicks

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

FGAC

Siehe detaillierte Zugriffskontrolle.

Feinkörnige Zugriffskontrolle (FGAC)

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

Flash-Cut-Migration

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

FM

Siehe Fundamentmodell.

Fundamentmodell (FM)

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

G

Generative KI

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

Geoblocking

Siehe geografische Einschränkungen.

Geografische Einschränkungen (Geoblocking)

Bei Amazon eine Option CloudFront, um zu verhindern, dass Benutzer in bestimmten Ländern auf Inhaltsverteilungen zugreifen. Sie können eine Zulassungsliste oder eine Sperrliste verwenden,

um zugelassene und gesperrte Länder anzugeben. Weitere Informationen finden Sie in der Dokumentation unter Beschränkung der geografischen Verteilung Ihrer Inhalte. CloudFront

Gitflow-Workflow

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

goldenes Bild

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

Greenfield-Strategie

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

Integritätsschutz

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

H

HEKTAR

Siehe Hochverfügbarkeit.

Heterogene Datenbankmigration

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

hohe Verfügbarkeit (HA)

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

historische Modernisierung

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

Daten zurückhalten

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

Homogene Datenbankmigration

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

heiße Daten

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

Hotfix

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

Hypercare-Phase

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

I

IaC

Sehen Sie Infrastruktur als Code.

Identitätsbasierte Richtlinie

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

Leerlaufanwendung

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

IIoT

Siehe Industrielles Internet der Dinge.

unveränderliche Infrastruktur

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

Eingehende (ingress) VPC

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

Inkrementelle Migration

Eine Cutover-Strategie, bei der Sie Ihre Anwendung in kleinen Teilen migrieren, anstatt eine einziges vollständiges Cutover durchzuführen. Beispielsweise könnten Sie zunächst nur einige Microservices oder Benutzer auf das neue System umstellen. Nachdem Sie sich vergewissert haben, dass alles ordnungsgemäß funktioniert, können Sie weitere Microservices oder Benutzer schrittweise verschieben, bis Sie Ihr Legacy-System außer Betrieb nehmen können. Diese Strategie reduziert die mit großen Migrationen verbundenen Risiken.

Industrie 4.0

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

Infrastruktur

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

Infrastructure as Code (IaC)

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

industrielles Internet der Dinge (T) IIo

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

Inspektions-VPC

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

Internet of Things (IoT)

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

Interpretierbarkeit

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

IoT

Siehe Internet der Dinge.

IT information library (ITIL, IT-Informationsbibliothek)

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

T service management (ITSM, IT-Servicemanagement)

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

BIS

Siehe IT-Informationsbibliothek.

ITSM

Siehe IT-Servicemanagement.

L

Labelbasierte Zugangskontrolle (LBAC)

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

Landing Zone

Eine landing zone ist eine gut strukturierte AWS Umgebung mit mehreren Konten, die skalierbar und sicher ist. Dies ist ein Ausgangspunkt, von dem aus Ihre Organisationen Workloads und Anwendungen schnell und mit Vertrauen in ihre Sicherheits- und Infrastrukturumgebung starten und bereitstellen können. Weitere Informationen zu Landing Zones finden Sie unter Einrichtung einer sicheren und skalierbaren AWS -Umgebung mit mehreren Konten..

großes Sprachmodell (LLM)

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

Große Migration

Eine Migration von 300 oder mehr Servern.

SCHWARZ

Siehe Labelbasierte Zugriffskontrolle.

Geringste Berechtigung

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

Lift and Shift

Siehe 7 Rs.

Little-Endian-System

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

LLM

Siehe großes Sprachmodell.

Niedrigere Umgebungen

Siehe Umgebung.

M

Machine Learning (ML)

Eine Art künstlicher Intelligenz, die Algorithmen und Techniken zur Mustererkennung und zum Lernen verwendet. ML analysiert aufgezeichnete Daten, wie z. B. Daten aus dem Internet der Dinge (IoT), und lernt daraus, um ein statistisches Modell auf der Grundlage von Mustern zu erstellen. Weitere Informationen finden Sie unter Machine Learning.

Hauptzweig

Siehe Filiale.

Malware

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

verwaltete Dienste

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

Manufacturing Execution System (MES)

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

MAP

Siehe Migration Acceleration Program.

Mechanismus

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

Mitgliedskonto

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

MES

Siehe Manufacturing Execution System.

Message Queuing-Telemetrietransport (MQTT)

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

Microservice

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

Microservices-Architekturen

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

Migration Acceleration Program (MAP)

Ein AWS Programm, das Beratung, Unterstützung, Schulungen und Services bietet, um Unternehmen dabei zu unterstützen, eine solide betriebliche Grundlage für die Umstellung auf

die Cloud zu schaffen und die anfänglichen Kosten von Migrationen auszugleichen. MAP umfasst eine Migrationsmethode für die methodische Durchführung von Legacy-Migrationen sowie eine Reihe von Tools zur Automatisierung und Beschleunigung gängiger Migrationsszenarien.

Migration in großem Maßstab

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

Migrationsfabrik

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

Migrationsmetadaten

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

Migrationsmuster

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

Migration Portfolio Assessment (MPA)

Ein Online-Tool, das Informationen zur Validierung des Geschäftsszenarios für die Migration auf das bereitstellt. AWS Cloud MPA bietet eine detaillierte Portfoliobewertung (richtige Servergröße, Preisgestaltung, Gesamtbetriebskostenanalyse, Migrationskostenanalyse) sowie Migrationsplanung (Anwendungsdatenanalyse und Datenerfassung, Anwendungsgruppierung,

Migrationspriorisierung und Wellenplanung). Das MPA-Tool (Anmeldung erforderlich) steht allen AWS Beratern und APN-Partnerberatern kostenlos zur Verfügung.

Migration Readiness Assessment (MRA)

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

Migrationsstrategie

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

ML

Siehe maschinelles Lernen.

Modernisierung

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

Bewertung der Modernisierungsfähigkeit

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

Monolithische Anwendungen (Monolithen)

Anwendungen, die als ein einziger Service mit eng gekoppelten Prozessen ausgeführt werden. Monolithische Anwendungen haben verschiedene Nachteile. Wenn ein Anwendungs-Feature stark nachgefragt wird, muss die gesamte Architektur skaliert werden. Das Hinzufügen oder

Verbessern der Feature einer monolithischen Anwendung wird ebenfalls komplexer, wenn die Codebasis wächst. Um diese Probleme zu beheben, können Sie eine Microservices-Architektur verwenden. Weitere Informationen finden Sie unter Zerlegen von Monolithen in Microservices.

MPA

Siehe Bewertung des Migrationsportfolios.

MQTT

Siehe Message Queuing-Telemetrietransport.

Mehrklassen-Klassifizierung

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

veränderbare Infrastruktur

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

O

OAC

Siehe Origin Access Control.

EICHE

Siehe Zugriffsidentität von Origin.

COM

Siehe organisatorisches Change-Management.

Offline-Migration

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

OI

Siehe Betriebsintegration.

OLA

Siehe Vereinbarung auf operativer Ebene.

Online-Migration

Eine Migrationsmethode, bei der der Quell-Workload auf das Zielsystem kopiert wird, ohne offline genommen zu werden. Anwendungen, die mit dem Workload verbunden sind, können während der Migration weiterhin funktionieren. Diese Methode beinhaltet keine bis minimale Ausfallzeit und wird in der Regel für kritische Produktionsworkloads verwendet.

OPC-UA

Siehe Open Process Communications — Unified Architecture.

Offene Prozesskommunikation — Einheitliche Architektur (OPC-UA)

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

Vereinbarung auf Betriebsebene (OLA)

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

Überprüfung der Betriebsbereitschaft (ORR)

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

Betriebstechnologie (OT)

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

Betriebsintegration (OI)

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

Organisationspfad

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

Organisatorisches Veränderungsmanagement (OCM)

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

Ursprungszugriffskontrolle (OAC)

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

Ursprungszugriffsidentität (OAI)

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

ORR

Weitere Informationen finden Sie unter Überprüfung der Betriebsbereitschaft.

NICHT

Siehe Betriebstechnologie.

Ausgehende (egress) VPC

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

P

Berechtigungsgrenze

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

persönlich identifizierbare Informationen (PII)

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

Personenbezogene Daten

Siehe persönlich identifizierbare Informationen.

Playbook

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

PLC

Siehe programmierbare Logiksteuerung.

PLM

Siehe Produktlebenszyklusmanagement.

policy

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

Polyglotte Beharrlichkeit

Unabhängige Auswahl der Datenspeichertechnologie eines Microservices auf der Grundlage von Datenzugriffsmustern und anderen Anforderungen. Wenn Ihre Microservices über dieselbe Datenspeichertechnologie verfügen, kann dies zu Implementierungsproblemen oder zu Leistungseinbußen führen. Microservices lassen sich leichter implementieren und erzielen eine bessere Leistung und Skalierbarkeit, wenn sie den Datenspeicher verwenden, der ihren Anforderungen am besten entspricht.

Portfoliobewertung

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

predicate

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

Prädikat Pushdown

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

Präventive Kontrolle

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

Prinzipal

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

Datenschutz von Natur aus

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

Privat gehostete Zonen

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

proaktive Steuerung

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

Produktlebenszyklusmanagement (PLM)

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

Produktionsumgebung

Siehe Umgebung.

Speicherprogrammierbare Steuerung (SPS)

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

schnelle Verkettung

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

Pseudonymisierung

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

publish/subscribe (pub/sub)

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

Q

Abfrageplan

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

Abfrageplanregression

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

R

RACI-Matrix

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

RAG

Siehe Erweiterte Generierung beim Abrufen.

Ransomware

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

RASCI-Matrix

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

RCAC

Siehe Zugriffskontrolle für Zeilen und Spalten.

Read Replica

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

neu strukturieren

Siehe 7 Rs.

Recovery Point Objective (RPO)

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

Wiederherstellungszeitziel (RTO)

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

Refaktorierung

Siehe 7 Rs.

Region

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

Regression

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

rehosten

Siehe 7 Rs.

Veröffentlichung

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

umziehen

Siehe 7 Rs.

neue Plattform

Siehe 7 Rs.

Rückkauf

Siehe 7 Rs.

Ausfallsicherheit

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

Ressourcenbasierte Richtlinie

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

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

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

Reaktive Kontrolle

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

Beibehaltung

Siehe 7 Rs.

zurückziehen

Siehe 7 Rs.

Retrieval Augmented Generation (RAG)

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

Drehung

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

Zugriffskontrolle für Zeilen und Spalten (RCAC)

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

RPO

Siehe Recovery Point Objective.

RTO

Siehe Ziel für die Erholungszeit.

Runbook

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

S

SAML 2.0

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

SCADA

Siehe Aufsichtskontrolle und Datenerfassung.

SCP

Siehe Richtlinie zur Dienstkontrolle.

Secret

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

Sicherheit durch Design

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

Sicherheitskontrolle

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

Härtung der Sicherheit

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

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

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

Automatisierung von Sicherheitsreaktionen

Eine vordefinierte und programmierte Aktion, die darauf ausgelegt ist, automatisch auf ein Sicherheitsereignis zu reagieren oder es zu beheben. Diese Automatisierungen dienen als detektive oder reaktionsschnelle Sicherheitskontrollen, die Sie bei der Implementierung bewährter AWS Sicherheitsmethoden unterstützen. Beispiele für automatisierte Antwortaktionen sind das Ändern einer VPC-Sicherheitsgruppe, das Patchen einer Amazon EC2 EC2-Instance oder das Rotieren von Anmeldeinformationen.

Serverseitige Verschlüsselung

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

Service-Kontrollrichtlinie (SCP)

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

Service-Endpunkt

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

Service Level Agreement (SLA)

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

Service-Level-Indikator (SLI)

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

Service-Level-Ziel (SLO)

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

Modell der geteilten Verantwortung

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

SIEM

Siehe Sicherheitsinformations- und Event-Management-System.

Single Point of Failure (SPOF)

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

SLA

Siehe Service Level Agreement.

SLI

Siehe Service-Level-Indikator.

ALSO

Siehe Service-Level-Ziel.

split-and-seed Modell

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

SPOTTEN

Siehe Single Point of Failure.

Sternschema

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

Strangler-Fig-Muster

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

Subnetz

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

Aufsichtskontrolle und Datenerfassung (SCADA)

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

Symmetrische Verschlüsselung

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

synthetisches Testen

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

Systemaufforderung

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

T

tags

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

Zielvariable

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

Aufgabenliste

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

Testumgebungen

Siehe Umgebung.

Training

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

Transit-Gateway

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

Stammbasierter Workflow

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

Vertrauenswürdiger Zugriff

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

Optimieren

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

Zwei-Pizzen-Team

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

U

Unsicherheit

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

undifferenzierte Aufgaben

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

höhere Umgebungen

Siehe Umgebung.

V

Vacuuming

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

Versionskontrolle

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

VPC-Peering

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

Schwachstelle

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

W

Warmer Cache

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

warme Daten

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

Fensterfunktion

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

Workload

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

Workstream

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

WURM

Sehen Sie einmal schreiben, viele lesen.

WQF

Siehe AWS Workload-Qualifizierungsrahmen.

einmal schreiben, viele lesen (WORM)

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

Z

Zero-Day-Exploit

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

Zero-Day-Sicherheitslücke

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

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.