Framework, protocolli e strumenti di intelligenza artificiale agentica su AWS

# AWS Guida prescrittiva

# AWS Guida prescrittiva: Framework, protocolli e strumenti di intelligenza artificiale agentica su AWS

# Table of Contents

# Framework, protocolli e strumenti di intelligenza artificiale agentica su AWS

Aaron Sempf e Joshua Samuel, Amazon Web Services ()AWS

Luglio 2025 (storia del documento)

L'intelligenza artificiale agentica è un potente paradigma all'intersezione tra intelligenza artificiale, sistemi distribuiti e ingegneria del software. È una classe di sistemi intelligenti composta da agenti software autonomi e asincroni. Gli agenti mostrano di agire, sono in grado di percepire il contesto, ragionare sugli obiettivi, prendere decisioni e intraprendere azioni mirate per conto degli utenti o dei sistemi. Questi agenti operano in modo indipendente, spesso collaborativo, all'interno di ambienti distribuiti e sono progettati per perseguire obiettivi delegati con intelligenza, memoria e intenti integrati.

Inoltre AWS, le organizzazioni possono sfruttare l'intelligenza artificiale agentica per automatizzare flussi di lavoro complessi, migliorare i processi decisionali e creare sistemi più reattivi. Questa guida fornisce informazioni sui componenti chiave necessari per creare soluzioni di intelligenza artificiale agentica efficaci:

- I framework di intelligenza artificiale agentica descrivono gli attuali framework di intelligenza artificiale agentica, comprese le revisioni dei relativi vantaggi e casi d'uso. Scopri come questi framework riducono il carico di lavoro indifferenziato tra modelli, protocolli e strumenti. Comprendi i criteri di selezione chiave per scegliere il framework giusto per le tue esigenze.
- I protocolli Agentic esplorano i protocolli di comunicazione standardizzati essenziali per le interazioni tra agenti. Agent-to-agentstanno emergendo protocolli, come l'open source Model Context Protocol (MCP) e Agent2Agent (A2A), oltre ad altre implementazioni proprietarie. Scopri come i protocolli comuni consentono a protocolli diversi di interagire senza problemi.
- Tools fornisce informazioni su strumenti basati su protocolli (come MCP), strumenti nativi del framework e meta-strumenti. Le organizzazioni possono creare un toolkit che si integra con i sistemi chiave dei loro flussi di lavoro, abilitando flussi di lavoro agentici sia per utenti finali che basati su server.

# Destinatari principali

Questa guida è rivolta ad architetti, sviluppatori e leader tecnologici che desiderano sfruttare la potenza degli agenti software basati sull'intelligenza artificiale all'interno delle moderne applicazioni native del cloud.

# Obiettivi

Questa guida ti consente di:

- Confronta diversi framework di intelligenza artificiale agentica per selezionare quello più appropriato per il tuo caso d'uso.
- Comprendi i vantaggi dei protocolli aperti per la creazione di architetture di intelligenza artificiale agentica sostenibili.
- Crea una strategia di integrazione degli strumenti appropriata durante la creazione di sistemi con agenti.

# Informazioni su questa serie di contenuti

Questa guida fa parte di una serie di pubblicazioni che forniscono modelli architettonici e linee guida tecniche per la creazione di agenti software basati sull'intelligenza artificiale. AWS La serie AWS Prescriptive Guidance include quanto segue:

- Rendere operativa l'intelligenza artificiale agentica su AWS
- Fondamenti dell'intelligenza artificiale agentica su AWS
- Modelli e flussi di lavoro di intelligenza artificiale agentica su AWS
- Framework, protocolli e strumenti di intelligenza artificiale agentica su (questa guida) AWS
- Creazione di architetture serverless per l'intelligenza artificiale agentica su AWS
- Creazione di architetture multi-tenant per l'intelligenza artificiale agentica su AWS

Per ulteriori informazioni su questa serie di contenuti, consulta Agentic AI.

# Framework di intelligenza artificiale agentica

Foundations of agentic AI on AWS esamina i modelli e i flussi di lavoro principali che consentono un comportamento autonomo e mirato. Alla base dell'implementazione di questi modelli c'è la scelta del framework. Un framework è la base software che fornisce la struttura, gli strumenti e le funzionalità di orchestrazione necessarie per creare agenti di intelligenza artificiale autonomi pronti per la produzione.

I framework di intelligenza artificiale agentica offrono diverse funzionalità essenziali che trasformano le interazioni grezze del modello di linguaggio di grandi dimensioni (LLM) in agenti robusti e autonomi in grado di operare in modo indipendente:

- L'orchestrazione degli agenti coordina il flusso di informazioni e il processo decisionale tra uno o più agenti per raggiungere obiettivi complessi senza l'intervento umano.
- L'integrazione degli strumenti consente agli agenti di interagire con sistemi esterni e fonti di dati per estendere le proprie funzionalità oltre l'elaborazione del linguaggio. APIs Per ulteriori informazioni, vedere Panoramica degli strumenti nella Strands Agents documentazione.
- La gestione della memoria fornisce uno stato persistente o basato sulla sessione per mantenere il contesto tra le interazioni, essenziale per attività autonome di lunga durata. Per ulteriori informazioni, consulta How to think about agent framework sul blog. LangChain
- La definizione del flusso di lavoro supporta modelli strutturati come catene, routing, parallelizzazione e cicli di riflessione che consentono un ragionamento autonomo sofisticato.
- L'implementazione e il monitoraggio facilitano la transizione dallo sviluppo alla produzione con l'osservabilità per i sistemi autonomi. Per ulteriori informazioni, consulta l'annuncio di LangGraphPlatform GA.

Queste funzionalità sono implementate con approcci ed enfasi diversi in tutto il panorama del framework, ognuna delle quali offre vantaggi distinti per diversi casi d'uso di agenti autonomi e contesti organizzativi.

Questa sezione descrive e confronta i principali framework per la creazione di soluzioni di intelligenza artificiale agentiche, con particolare attenzione ai loro punti di forza, limiti e casi d'uso ideali per il funzionamento autonomo:

- Agenti Strands
- LangChain e LangGraph

- Crew AI
- Agenti Amazon Bedrock
- AutoGen
- Confronto tra framework di intelligenza artificiale agentica

> **Note**
>
> Questa sezione tratta i framework che supportano specificamente l'agenzia dell'IA e non copre le interfacce frontend o l'IA generativa senza agenzia.

# Strands Agents

Strands Agentsè un SDK open source che è stato inizialmente rilasciato da AWS, come descritto nel blog Open Source.AWS Strands Agentsè progettato per creare agenti di intelligenza artificiale autonomi con un approccio basato sul modello e fornisce un framework flessibile ed estensibile progettato per funzionare senza interruzioni Servizi AWS pur rimanendo aperto all'integrazione con componenti di terze parti. Strands Agents è ideale per creare soluzioni completamente autonome.

## Caratteristiche principali di Strands Agents

Strands Agentsinclude le seguenti funzionalità chiave:

- Progettazione incentrata sul modello: basata sul concetto che il modello di base è il fulcro dell'intelligenza degli agenti e consente un ragionamento autonomo sofisticato. Per ulteriori informazioni, consulta Agent Loop nella documentazione. Strands Agents
- Integrazione MCP: supporto nativo per il Model Context Protocol (MCP), che consente la fornitura di un contesto standardizzato LLMs per un funzionamento autonomo e coerente.
- Servizio AWS integrazione: connessione perfetta ad Amazon Bedrock e altro Servizi AWS per AWS Step Functions flussi di lavoro autonomi completi. AWS Lambda Per ulteriori informazioni, consulta AWS Weekly Roundup (blog).AWS
- Selezione del modello Foundation: supporta vari modelli di base tra cui Anthropic Claude, Amazon Nova (Premier, Pro, Lite e Micro) su Amazon Bedrock e altri per ottimizzare diverse capacità di ragionamento autonomo. Per ulteriori informazioni, consulta Amazon Bedrock nella Strands Agents documentazione.

- Integrazione dell'API LLM: integrazione flessibile con diverse interfacce di servizio LLM tra cui Amazon Bedrock, OpenAI e altre per l'implementazione in produzione. Per ulteriori informazioni, consulta Amazon Bedrock Basic Usage nella Strands Agents documentazione.
- Funzionalità multimodali: supporto per più modalità tra cui l'elaborazione di testo, voce e immagini per interazioni complete con agenti autonomi. Per ulteriori informazioni, consulta Amazon Bedrock Multimodal Support nella Strands Agents documentazione.
- Ecosistema di strumenti: ricco set di strumenti per Servizio AWS l'interazione, con estensibilità per strumenti personalizzati che ampliano le capacità autonome. Per ulteriori informazioni, vedere Panoramica degli strumenti nella Strands Agents documentazione.

## Quando usare Strands Agents

Strands Agentsè particolarmente adatto per scenari con agenti autonomi, tra cui:

- Organizzazioni che si basano su AWS un'infrastruttura che desiderano un'integrazione nativa Servizi AWS per flussi di lavoro autonomi
- Team che richiedono funzionalità di sicurezza, scalabilità e conformità di livello aziendale per sistemi autonomi di produzione
- Progetti che richiedono flessibilità nella selezione dei modelli tra diversi fornitori per attività autonome specializzate
- Casi d'uso che richiedono una stretta integrazione con i AWS flussi di lavoro e le risorse esistenti per processi autonomi end-to-end

## Approccio di implementazione per Strands Agents

Strands Agentsfornisce un approccio di implementazione semplice per gli stakeholder aziendali, come illustrato nella sua Guida rapida. Il framework consente alle organizzazioni di:

- Seleziona modelli base come Amazon Nova (Premier, Pro, Lite o Micro) su Amazon Bedrock in base a requisiti aziendali specifici.
- Definisci strumenti personalizzati che si connettono ai sistemi e alle fonti di dati aziendali.
- Elabora più modalità tra cui testo, immagini e voce.
- Implementa agenti in grado di rispondere autonomamente alle domande aziendali ed eseguire attività.

Questo approccio di implementazione consente ai team aziendali di sviluppare e implementare rapidamente agenti autonomi senza una profonda esperienza tecnica nello sviluppo di modelli di intelligenza artificiale.

### Esempio reale di Strands Agents

AWS Transform for .NET lo utilizza per Strands Agents potenziare le proprie funzionalità di modernizzazione delle applicazioni, come descritto in AWS Transform per.NET, il primo servizio di intelligenza artificiale agentica per modernizzare le applicazioni.NET su larga scala (Blog).AWS Questo servizio di produzione impiega più agenti autonomi specializzati. Gli agenti collaborano per analizzare le applicazioni.NET legacy, pianificare strategie di modernizzazione ed eseguire trasformazioni del codice verso architetture native del cloud senza l'intervento umano. AWS Transform for .NET dimostra la disponibilità alla produzione dei sistemi autonomi aziendali. Strands Agents

## LangChain e LangGraph

LangChainè uno dei framework più affermati nell'ecosistema di intelligenza artificiale agentica. LangGraphestende le sue funzionalità per supportare flussi di lavoro complessi e basati sullo stato degli agenti, come descritto nel blog. LangChain Insieme, forniscono una soluzione completa per la creazione di sofisticati agenti di intelligenza artificiale autonomi con ricche capacità di orchestrazione per operazioni indipendenti.

### Caratteristiche principali di e LangChainLangGraph

LangChaine LangGraph includono le seguenti funzionalità chiave:

- Ecosistema di componenti: ampia libreria di componenti predefiniti per varie funzionalità di agenti autonomi, che consente lo sviluppo rapido di agenti specializzati. Per ulteriori informazioni, consulta la documentazione relativa ad LangChain.
- Selezione del modello Foundation: supporto per diversi modelli di base, tra cui Anthropic Claude, Amazon Nova (Premier, Pro, Lite e Micro) su Amazon Bedrock e altri per diverse funzionalità di ragionamento. Per ulteriori informazioni, consulta Ingressi e uscite nella documentazione. LangChain
- Integrazione dell'API LLM: interfacce standardizzate per più fornitori di servizi LLM (Large Language Model) tra cui Amazon Bedrock e altri per una OpenAI distribuzione flessibile. Per ulteriori informazioni, consulta LLMs nella documentazione LangChain.

- Elaborazione multimodale: supporto integrato per l'elaborazione di testo, immagini e audio per consentire ricche interazioni multimodali tra agenti autonomi. Per ulteriori informazioni, vedete Multimodalità nella documentazione. LangChain
- Flussi di lavoro basati su grafici: LangGraph consentono di definire comportamenti complessi di agenti autonomi come macchine a stati, supportando una logica decisionale sofisticata. Per ulteriori informazioni, consulta l'annuncio di LangGraphPlatform GA.
- Astrazioni di memoria: opzioni multiple per la gestione della memoria a breve e lungo termine, essenziali per gli agenti autonomi che mantengono il contesto nel tempo. Per ulteriori informazioni, consulta Come aggiungere memoria ai chatbot nella documentazione. LangChain
- Integrazione con strumenti: ricco ecosistema di integrazioni di strumenti tra vari servizi e estensione delle funzionalità degli APIs agenti autonomi. Per ulteriori informazioni, consulta Tools nella LangChain documentazione.
- LangGraph piattaforma: soluzione gestita di implementazione e monitoraggio per ambienti di produzione, che supporta agenti autonomi a lunga durata. Per ulteriori informazioni, consulta l'annuncio di LangGraphPlatform GA.

## Quando usare LangChain e LangGraph

LangChaine LangGraph sono particolarmente adatti per scenari con agenti autonomi, tra cui:

- Flussi di lavoro complessi di ragionamento in più fasi che richiedono un'orchestrazione sofisticata per un processo decisionale autonomo
- Progetti che richiedono l'accesso a un ampio ecosistema di componenti e integrazioni predefiniti per diverse funzionalità autonome
- Team con infrastruttura ed esperienza di machine Python learning (ML) esistenti che desiderano creare sistemi autonomi
- Casi d'uso che richiedono una gestione dello stato complessa in sessioni di agenti autonomi di lunga durata

## Approccio di implementazione per LangChain e LangGraph

LangChaine LangGraph forniscono un approccio di implementazione strutturato per gli stakeholder aziendali, come dettagliato nella LangGraphdocumentazione. Il framework consente alle organizzazioni di:

- Definisci grafici sofisticati del flusso di lavoro che rappresentano i processi aziendali.
- Crea modelli di ragionamento in più fasi con punti decisionali e logica condizionale.
- Integra funzionalità di elaborazione multimodali per la gestione di diversi tipi di dati.
- Implementa il controllo di qualità attraverso meccanismi di revisione e convalida integrati.

Questo approccio basato su grafici consente ai team aziendali di modellare processi decisionali complessi come flussi di lavoro autonomi. I team hanno una chiara visibilità su ogni fase del processo di ragionamento e la capacità di verificare i percorsi decisionali.

### Esempio reale di e LangChainLangGraph

Vodafoneha implementato agenti autonomi utilizzando LangChain (eLangGraph) per migliorare i flussi di lavoro di ingegneria dei dati e operativi, come dettagliato nel case study LangChainEnterprise. Hanno creato assistenti di intelligenza artificiale interni che monitorano autonomamente le metriche delle prestazioni, recuperano informazioni dai sistemi di documentazione e presentano informazioni utili, il tutto attraverso interazioni in linguaggio naturale.

L'Vodafoneimplementazione utilizza caricatori di documenti LangChain modulari, integrazione vettoriale e supporto per più LLMs (OpenAI, 3 e) per prototipare e confrontare rapidamente queste pipeline. LLaMA Gemini Sono stati quindi utilizzati per LangGraph strutturare l'orchestrazione multiagente implementando agenti secondari modulari. Questi agenti eseguono attività di raccolta, elaborazione, riepilogo e ragionamento. LangGraphhanno integrato questi agenti APIs nei loro sistemi cloud.

## CrewAI

CrewAIè un framework open source incentrato specificamente sull'orchestrazione autonoma multiagente, disponibile su. GitHub Fornisce un approccio strutturato alla creazione di team di agenti autonomi specializzati che collaborano per risolvere compiti complessi senza l'intervento umano. CrewAIenfatizza il coordinamento basato sui ruoli e la delega dei compiti.

### Caratteristiche principali di CrewAI

CrewAIoffre le seguenti funzionalità chiave:

- Progettazione degli agenti basata sui ruoli: gli agenti autonomi sono definiti con ruoli, obiettivi e storie precedenti specifici per consentire competenze specializzate. Per ulteriori informazioni, consulta Crafting Effective Agents nella documentazione. CrewAI
- Delega delle attività: meccanismi integrati per l'assegnazione autonoma delle attività agli agenti appropriati in base alle loro capacità. Per ulteriori informazioni, consulta Tasks nella CrewAI documentazione.
- Collaborazione tra agenti: framework per la comunicazione autonoma tra agenti e la condivisione delle conoscenze senza la mediazione umana. Per ulteriori informazioni, consulta Collaborazione nella documentazione. CrewAI
- Gestione dei processi: flussi di lavoro strutturati per l'esecuzione di attività autonome sequenziali e parallele. Per ulteriori informazioni, consulta Processi nella CrewAI documentazione.
- Selezione del modello di base: supporto per vari modelli di base, tra cui Anthropic Claude, i modelli Amazon Nova (Premier, Pro, Lite e Micro) su Amazon Bedrock e altri per l'ottimizzazione per diverse attività di ragionamento autonomo. Per ulteriori informazioni, consulta LLMs nella documentazione CrewAI.
- Integrazione dell'API LLM: integrazione flessibile con più interfacce di servizio LLM, tra cui Amazon BedrockOpenAI, e implementazioni di modelli locali. Per ulteriori informazioni, consulta gli esempi di configurazione del provider nella documentazione. CrewAI
- Supporto multimodale: funzionalità emergenti per la gestione di testo, immagini e altre modalità per interazioni complete con agenti autonomi. Per ulteriori informazioni, consulta Using Multimodal Agents nella documentazione. CrewAI

## Quando usare CrewAI

CrewAIè particolarmente adatto per scenari con agenti autonomi, tra cui:

- Problemi complessi che traggono vantaggio da competenze specializzate e basate sui ruoli che operano in modo autonomo
- Progetti che richiedono una collaborazione esplicita tra più agenti autonomi
- Casi d'uso in cui la scomposizione dei problemi basata sul team migliora la risoluzione autonoma dei problemi
- Scenari che richiedono una chiara separazione delle preoccupazioni tra i diversi ruoli degli agenti autonomi

## Approccio di implementazione per CrewAI

CrewAIfornisce un'implementazione basata sui ruoli dell'approccio dei team di agenti di intelligenza artificiale agli stakeholder aziendali, come descritto in Getting Started nella CrewAI documentazione. Il framework consente alle organizzazioni di:

- Definisci agenti autonomi specializzati con ruoli, obiettivi e aree di competenza specifici.
- Assegna attività agli agenti in base alle loro capacità specializzate.
- Stabilisci dipendenze chiare tra le attività per creare flussi di lavoro strutturati.
- Orchestra la collaborazione tra più agenti per risolvere problemi complessi.

Questo approccio basato sui ruoli rispecchia le strutture dei team umani, il che lo rende intuitivo da comprendere e implementare per i leader aziendali. Le organizzazioni possono creare team autonomi con aree di competenza specializzate che collaborano per raggiungere gli obiettivi aziendali, in modo simile a come operano i team umani. Tuttavia, il team autonomo può lavorare ininterrottamente senza l'intervento umano.

## Esempio reale di CrewAI

AWS ha implementato sistemi multiagente autonomi utilizzando CrewAI integrato con Amazon Bedrock, come dettagliato nel CrewAI case study pubblicato. AWS e CrewAI ha sviluppato un framework sicuro e indipendente dal fornitore. L'architettura CrewAI open source «flows-and-crews» si integra perfettamente con i modelli di base, i sistemi di memoria e le barriere di conformità di Amazon Bedrock.

Gli elementi chiave dell'implementazione includono:

- Progetti e open source, oltre a progetti di riferimento CrewAI rilasciati che associano CrewAI gli agenti ai modelli AWS e agli strumenti di osservabilità di Amazon Bedrock. Hanno inoltre rilasciato sistemi esemplari come un team di controllo della AWS sicurezza composto da più agenti, flussi di modernizzazione del codice e automazione del back-office per i beni di consumo confezionati (CPG).
- Integrazione dello stack di osservabilità: la soluzione integra il monitoraggio con Amazon CloudWatch e consente la tracciabilità e LangFuse il debug dal proof of concept alla produzione. AgentOps

- Dimostrato ritorno sull'investimento (ROI): i primi progetti pilota mostrano importanti miglioramenti: un'esecuzione più rapida del 70% per un grande progetto di modernizzazione del codice e una riduzione di circa il 90% dei tempi di elaborazione per un flusso di back-office CPG.

# Agenti Amazon Bedrock

Amazon Bedrock Agents è un servizio completamente gestito che ti consente di creare e configurare agenti autonomi nelle tue applicazioni. Può orchestrare le interazioni tra modelli di base, fonti di dati, applicazioni software e conversazioni con gli utenti. Il suo approccio semplificato alla creazione di agenti non richiede la fornitura di capacità, la gestione dell'infrastruttura o la scrittura di codice personalizzato.

## Caratteristiche principali di Amazon Bedrock Agents

Amazon Bedrock Agents include le seguenti funzionalità chiave:

- Servizio completamente gestito: gestione completa dell'infrastruttura senza la necessità di fornire capacità o gestire i sistemi sottostanti. Per ulteriori informazioni, consulta Automatizza le attività nella tua applicazione utilizzando agenti AI nella documentazione di Amazon Bedrock.
- Sviluppo basato sulle API: definisci ed esegui gli agenti tramite semplici chiamate API specificando modelli, istruzioni, strumenti e parametri di configurazione. Per ulteriori informazioni, consulta Creare e configurare l'agente manualmente nella documentazione di Amazon Bedrock.
- Gruppi di azioni: definisci azioni specifiche che il tuo agente può eseguire creando gruppi di azioni con schemi API. Per ulteriori informazioni, consulta Utilizzare i gruppi di azioni per definire le azioni che il tuo agente deve eseguire nella documentazione di Amazon Bedrock.
- Integrazione con la Knowledge Base: connettiti senza problemi alle Knowledge Base di Amazon Bedrock per aumentare le risposte degli agenti con i dati della tua organizzazione. Per ulteriori informazioni, consulta Augment Response generation per il tuo agente con knowledge base nella documentazione di Amazon Bedrock.
- Modelli di prompt avanzati: personalizza il comportamento degli agenti tramite modelli di prompt per la preelaborazione, l'orchestrazione, la generazione di risposte della knowledge base e la post-elaborazione. Per ulteriori informazioni, consulta Migliora la precisione dell'agente utilizzando modelli di prompt avanzati in Amazon Bedrock nella documentazione di Amazon Bedrock.
- Tracciamento e osservabilità: monitora il processo di step-by-step ragionamento dell'agente utilizzando funzionalità di tracciamento integrate. Per ulteriori informazioni, consulta il processo di

step-by-step ragionamento dell'agente Track che utilizza trace nella documentazione di Amazon Bedrock.

- Controllo delle versioni e alias: crea più versioni del tuo agente e distribuiscile tramite alias per implementazioni controllate. Per ulteriori informazioni, consulta Implementare e utilizzare un agente Amazon Bedrock nella tua applicazione nella documentazione di Amazon Bedrock.

## Quando usare Amazon Bedrock Agents

Amazon Bedrock Agents è particolarmente adatto per scenari di agenti autonomi, tra cui:

- Organizzazioni che desiderano un'esperienza completamente gestita per la creazione e l'implementazione di agenti senza gestire l'infrastruttura
- Progetti che richiedono lo sviluppo e l'implementazione rapidi di agenti tramite la configurazione anziché il codice
- Casi d'uso che traggono vantaggio dalla stretta integrazione con altre funzionalità di Amazon Bedrock come Knowledge Bases e Guardrails
- I team non dispongono delle risorse interne necessarie per creare agenti partendo da zero, ma necessitano di funzionalità autonome pronte per la produzione

## Approccio di implementazione per Amazon Bedrock Agents

Amazon Bedrock Agents offre un approccio di implementazione basato sulla configurazione per gli stakeholder aziendali. Il servizio consente alle organizzazioni di:

- Definisci gli agenti tramite chiamate AWS Management Console o API senza scrivere codice complesso.
- Crea gruppi di azioni che specificano le operazioni APIs e che l'agente può eseguire.
- Connect le knowledge base per fornire all'agente informazioni specifiche del dominio.
- Verifica e ripeti il comportamento degli agenti tramite un'interfaccia visiva.

Questo approccio gestito consente ai team aziendali di sviluppare e implementare rapidamente agenti autonomi senza competenze tecniche approfondite nello sviluppo di modelli di intelligenza artificiale o nella gestione dell'infrastruttura.

## Esempio reale di Amazon Bedrock Agents

Una soluzione per le operazioni finanziarie (FinOps) descritta in questo post del AWS blog utilizza il framework multi-agente Amazon Bedrock per creare un assistente per la gestione dei costi del cloud basato sull'intelligenza artificiale. Il conveniente modello di base Amazon Nova alimenta la soluzione in cui un agente FinOps Supervisor centrale delega le attività ad agenti specializzati. Questi agenti recuperano e analizzano i dati di AWS spesa utilizzando AWS Cost Explorer e generano consigli per ridurre i costi utilizzando. AWS Trusted Advisor

Il sistema include l'accesso sicuro degli utenti tramite Amazon Cognito, un front-end ospitato su AWS Amplify, e gruppi di AWS Lambda azione per analisi e previsioni in tempo reale. I team finanziari possono porre domande in linguaggio naturale come «Quali erano i miei costi a febbraio 2025?» Il sistema risponde con interruzioni dettagliate, suggerimenti di ottimizzazione e previsioni, il tutto all'interno di un'architettura scalabile e senza server implementata utilizzando. AWS CloudFormation

# AutoGen

AutoGenè un framework open source rilasciato inizialmente da. Microsoft AutoGensi concentra sull'abilitazione di agenti di intelligenza artificiale autonomi conversazionali e collaborativi. Fornisce un'architettura flessibile per la creazione di sistemi multiagente con particolare attenzione alle interazioni asincrone e basate sugli eventi tra agenti per flussi di lavoro autonomi complessi.

## Caratteristiche principali di AutoGen

AutoGenoffre le seguenti funzionalità chiave:

- Agenti conversazionali: basati su conversazioni in linguaggio naturale tra agenti autonomi, consentono un ragionamento sofisticato attraverso il dialogo. Per ulteriori informazioni, consulta Multi-agent Conversation Framework nella documentazione. AutoGen
- Architettura asincrona: progettazione basata sugli eventi per interazioni non bloccanti tra agenti autonomi, che supporta flussi di lavoro paralleli complessi. Per ulteriori informazioni, consulta Risoluzione di più attività in una sequenza di chat asincrone nella documentazione. AutoGen
- H uman-in-the-loop — Forte supporto alla partecipazione umana opzionale a flussi di lavoro degli agenti altrimenti autonomi, quando necessario. Per ulteriori informazioni, consulta Consentire il feedback umano negli agenti nella AutoGen documentazione.

- Generazione ed esecuzione di codice: funzionalità specializzate per agenti autonomi incentrati sul codice in grado di scrivere ed eseguire codice. Per ulteriori informazioni, consulta Code Execution nella AutoGen documentazione.
- Comportamenti personalizzabili: configurazione flessibile e autonoma degli agenti e controllo delle conversazioni per diversi casi d'uso. Per ulteriori informazioni, consulta agentchat.conversable_agent nella documentazione. AutoGen
- Selezione del modello Foundation: supporto per vari modelli di base, tra cui Anthropic Claude, Amazon Nova (Premier, Pro, Lite e Micro) su Amazon Bedrock e altri per diverse funzionalità di ragionamento autonomo. Per ulteriori informazioni, consulta LLM Configuration nella documentazione. AutoGen
- Integrazione dell'API LLM: configurazione standardizzata per più interfacce di servizio LLM, tra cui Amazon OpenAI Bedrock e. Azure OpenAI Per ulteriori informazioni, consulta oai.openai_utils nel riferimento API. AutoGen
- Elaborazione multimodale: supporto per l'elaborazione di testo e immagini per consentire ricche interazioni multimodali con agenti autonomi. Per ulteriori informazioni, consulta Interagire con i modelli multimodali: GPT-4V nella documentazione. AutoGen AutoGen

## Quando usare AutoGen

AutoGenè particolarmente adatto per scenari con agenti autonomi, tra cui:

- Applicazioni che richiedono flussi conversazionali naturali tra agenti autonomi per ragionamenti complessi
- Progetti che richiedono sia un funzionamento completamente autonomo che capacità opzionali di supervisione umana
- Casi d'uso che prevedono la generazione, l'esecuzione e il debug di codice autonomi senza l'intervento umano
- Scenari che richiedono modelli di comunicazione tra agenti autonomi flessibili e asincroni

## Approccio di implementazione per AutoGen

AutoGenfornisce un approccio di implementazione conversazionale per gli stakeholder aziendali, come descritto in Getting Started nella AutoGen documentazione. Il framework consente alle organizzazioni di:

- Crea agenti autonomi che comunicano attraverso conversazioni in linguaggio naturale.
- Implementa interazioni asincrone basate sugli eventi tra più agenti.
- Combina un funzionamento completamente autonomo con la supervisione umana opzionale quando necessario.
- Sviluppa agenti specializzati per diverse funzioni aziendali che collaborino attraverso il dialogo.

Questo approccio conversazionale rende il ragionamento del sistema autonomo trasparente e accessibile agli utenti aziendali. I responsabili delle decisioni possono osservare il dialogo tra gli agenti per capire come vengono raggiunte le conclusioni e, facoltativamente, partecipare alla conversazione quando è richiesto il giudizio umano.

### Esempio reale di AutoGen

Magentic-Oneè un sistema multiagente generalista open source progettato per risolvere autonomamente attività complesse e in più fasi in diversi ambienti, come descritto nel blog AI Frontiers. Microsoft Alla base c'è l'agente Orchestrator, che scompone gli obiettivi di alto livello e monitora i progressi utilizzando registri strutturati. Questo agente delega le attività secondarie ad agenti specializzati (comeWebSurfer,, FileSurfer and) e si adatta dinamicamente ripianificando quando necessario. Coder ComputerTerminal

Il sistema è basato sul AutoGen framework ed è indipendente dal modello, l'impostazione predefinita è GPT-4o. Raggiunge prestazioni all'avanguardia su benchmark come, e, il tutto senza regolazioni specifiche per attività. GAIA AssistantBench WebArena Inoltre, supporta l'estensibilità modulare e una valutazione AutoGenBench rigorosa tramite suggerimenti.

## Confronto tra framework di intelligenza artificiale agentica

Quando scegli un framework di intelligenza artificiale agentica per lo sviluppo di agenti autonomi, considera in che modo ciascuna opzione si allinea ai tuoi requisiti specifici. Considerate non solo le sue capacità tecniche ma anche la sua idoneità organizzativa, tra cui l'esperienza del team, l'infrastruttura esistente e i requisiti di manutenzione a lungo termine. Molte organizzazioni potrebbero trarre vantaggio da un approccio ibrido, che sfrutta più framework per diversi componenti del loro ecosistema di intelligenza artificiale autonomo.

La tabella seguente confronta i livelli di maturità (più forte, forte, adeguato o debole) di ciascun framework in base alle dimensioni tecniche chiave. Per ogni framework, la tabella include anche

informazioni sulle opzioni di implementazione in produzione e sulla complessità della curva di apprendimento.

| Framework | AWS integration | Supporto multiagente autonomo | Complessità del workflow autonomo | Funzionalità multimodali | Selezione del modello Foundation | Integrazione con l'API LLM | Distribuzione in produzione | Curva di apprendimento |
|---|---|---|---|---|---|---|---|---|
| Amazon Bedrock Agents | Il più forte | Sufficiente | Adegua | Forte | Forte | Forte | Completamente gestito | Bassa |
| AutoGen | Debole | Forte | Forte | Sufficiente | Adegua | Forte | Fai da te (fai da te) | Ripido |
| CrewAI | Debole | Forte | Sufficiente | Debole | Sufficiente | Adegua | FAI DA TE | Moderata |
| LangChain / LangGraph | adeguato | Forte | Il più forte | Il più forte | Il più forte | Il più forte | Piattaforma per il fai da te | Ripido |
| Strands Agents | Il più forte | Forte | Il più forte | Forte | Forte | Il più forte | FAI DA TE | Moderata |

## Considerazioni sulla scelta di un framework di intelligenza artificiale agentica

Nello sviluppo di agenti autonomi, considera i seguenti fattori chiave:

- AWS integrazione dell'infrastruttura: le organizzazioni in cui hanno investito molto AWS trarranno i maggiori vantaggi dalle integrazioni native di Strands Agents with Servizi AWS per flussi di lavoro autonomi. Per ulteriori informazioni, consulta AWS Weekly Roundup (blog).AWS

- Selezione del modello di base: valuta quale framework offre il supporto migliore per i tuoi modelli di base preferiti (ad esempio, i modelli Amazon Nova su Amazon Bedrock o Anthropic Claude), in base ai requisiti di ragionamento del tuo agente autonomo. Per ulteriori informazioni, consulta Building Effective Agents sul sito Web. Anthropic
- Integrazione dell'API LLM: valuta i framework in base alla loro integrazione con le tue interfacce di servizio LLM (Large Language Model) preferite (ad esempio Amazon Bedrock o) per l'implementazione in produzione. OpenAI Per ulteriori informazioni, consulta Model Interfaces nella documentazione. Strands Agents
- Requisiti multimodali: per gli agenti autonomi che devono elaborare testo, immagini e voce, considera le funzionalità multimodali di ciascun framework. Per ulteriori informazioni, vedete Multimodalità nella documentazione. LangChain
- Complessità del flusso di lavoro autonomo: flussi di lavoro autonomi più complessi con una sofisticata gestione degli stati potrebbero favorire le funzionalità avanzate delle macchine a stati di. LangGraph
- Collaborazione autonoma in team: i progetti che richiedono una collaborazione autonoma esplicita e basata sui ruoli tra agenti specializzati possono trarre vantaggio dall'architettura orientata al team di. CrewAI
- Paradigma di sviluppo autonomo: i team che preferiscono modelli conversazionali e asincroni per agenti autonomi potrebbero preferire l'architettura basata sugli eventi di. AutoGen
- Approccio gestito o basato sul codice: le organizzazioni che desiderano un'esperienza completamente gestita con una codifica minima dovrebbero prendere in considerazione Amazon Bedrock Agents. Le organizzazioni che richiedono una personalizzazione più profonda potrebbero preferire Strands Agents altri framework con funzionalità specializzate che si allineano meglio ai requisiti specifici degli agenti autonomi.
- Preparazione alla produzione per sistemi autonomi: prendete in considerazione le opzioni di implementazione, le funzionalità di monitoraggio e le funzionalità aziendali per gli agenti autonomi di produzione.

# Protocolli agentici

Gli agenti di intelligenza artificiale richiedono protocolli di comunicazione standardizzati per interagire con altri agenti e servizi. Le organizzazioni che implementano architetture di agenti devono affrontare sfide significative in termini di interoperabilità, indipendenza dai fornitori e preparazione dei propri investimenti al futuro.

Questa sezione aiuta a navigare nel panorama dei agent-to-agent protocolli con particolare attenzione agli standard aperti che massimizzano la flessibilità e l'interoperabilità. (Per informazioni sui agent-to-tool protocolli, consulta Strategia di integrazione degli strumenti più avanti in questa guida.)

Questa sezione evidenzia il Model Context Protocol (MCP), uno standard aperto originariamente sviluppato Anthropic nel 2024. Oggi supporta AWS attivamente MCP attraverso contributi allo sviluppo e all'implementazione del protocollo. AWS collabora con i principali framework di agenti open source, tra cui, e LangGraph CrewAILlamaIndex, per plasmare il futuro della comunicazione tra agenti sul protocollo. Per ulteriori informazioni, vedere Protocolli aperti per l'interoperabilità degli agenti, parte 1: Comunicazione tra agenti su MCP (Blog).AWS

In questa sezione



## Perché è importante la selezione del protocollo

La selezione dei protocolli determina fondamentalmente il modo in cui è possibile creare ed evolvere l'architettura degli agenti AI. Scegliendo protocolli che supportano la portabilità tra framework di agenti, ottieni la flessibilità necessaria per combinare diversi sistemi di agenti e flussi di lavoro per soddisfare le tue esigenze specifiche.

I protocolli aperti consentono di integrare gli agenti su più framework. Ad esempio, utilizzali LangChain per la prototipazione rapida e l'implementazione di sistemi di produzione comunicando

tramite un protocollo comuneStrands Agents, come MCP o il protocollo Agent2Agent (A2A). Questa flessibilità riduce la dipendenza da specifici provider di intelligenza artificiale, semplifica l'integrazione con i sistemi esistenti e consente di migliorare le capacità degli agenti nel tempo.

I protocolli ben progettati stabiliscono inoltre modelli di sicurezza coerenti per l'autenticazione e l'autorizzazione in tutto l'ecosistema degli agenti. Soprattutto, la portabilità del protocollo preserva la libertà di adottare nuovi framework e funzionalità degli agenti man mano che emergono. La scelta di protocolli aperti protegge l'investimento nello sviluppo di agenti mantenendo al contempo l'interoperabilità con i sistemi di terze parti.

### Vantaggi dei protocolli aperti

Quando si implementano le proprie estensioni o si creano sistemi di agenti personalizzati, i protocolli aperti offrono vantaggi convincenti:

- Documentazione e trasparenza: in genere forniscono una documentazione completa e implementazioni trasparenti
- Supporto comunitario: accesso a comunità di sviluppatori più ampie per la risoluzione dei problemi e le migliori pratiche
- Garanzie di interoperabilità: maggiore garanzia che le estensioni funzionino su diverse implementazioni
- Compatibilità futura: riduzione del rischio di interruzione delle modifiche o di obsolescenza
- Influenza sullo sviluppo: opportunità di contribuire all'evoluzione del protocollo

## Agent-to-agent protocolli

La tabella seguente fornisce una panoramica dei protocolli agentici che consentono a più agenti di collaborare, delegare attività e condividere informazioni.

| Protocollo | Ideale per | Considerazioni |
| --- | --- | --- |
| Comunicazione tra agenti MCP | Organizzazioni alla ricerca di modelli flessibili di collaborazione tra agenti | • Un'estensione del Model Context Protocol (MCP) proposta da That AWS si basa sulle sue basi esistenti |

| | | |
|---|---|---|
| | | per la comunicazione agent-to-agent<br>• Consente una collaborazione senza interruzioni tra gli agenti con una sicurezza basata sulla sicurezza OAuth |
| Protocollo A2A | Ecosistemi di agenti multipiattaforma | • Supportato da Google<br>• Standard più recente con un'adozione più limitata rispetto a MCP |
| AutoGenmultiagente | sistemi multiagente incentrati sulla ricerca | • Supportato da Microsoft<br>• Forte per interazioni complesse con agenti |
| CrewAI | Team di agenti basati sui ruoli | • Implementazione indipendente<br>• Ottimo per simulare strutture organizzative |

## Decidere tra le opzioni di protocollo

Quando implementate agent-to-agent la comunicazione, abbinate i vostri requisiti di comunicazione specifici alle funzionalità del protocollo appropriate. Modelli di interazione diversi richiedono funzionalità di protocollo diverse. La tabella seguente descrive i modelli di comunicazione più comuni e consiglia le scelte di protocollo più adatte per ogni scenario.

| Pattern | Descrizione | Scelta del protocollo ideale |
|---|---|---|
| Richiesta e risposta semplici | Interazioni una tantum tra agenti | MCP con flussi stateless |
| Dialoghi statici | Conversazioni continue con contesto | MCP con gestione delle sessioni |

| | | |
|---|---|---|
| Collaborazione tra più agenti | Interazioni complesse tra più agenti | Interagente MCP o AutoGen |
| Flussi di lavoro basati sul team | Team di agenti gerarchici con ruoli definiti | Interagente MCP, oppure CrewAI AutoGen |

Oltre ai modelli di comunicazione, diversi fattori tecnici e organizzativi possono influenzare la selezione del protocollo. La tabella seguente riporta le considerazioni chiave che possono aiutarvi a valutare quale protocollo si allinea maggiormente ai vostri requisiti di implementazione specifici.

| Considerazione | Descrizione | Esempio |
|---|---|---|
| Modello di sicurezza | Requisiti di autenticazione e autorizzazione | OAuth 2.0 in MCP |
| Ambiente di implementazione | Dove gli agenti funzioneranno e comunicheranno | Macchina distribuita o singola |
| Compatibilità con l'ecosistema | Integrazione con i framework di agenti esistenti | LangChain o Strands Agents |
| Esigenze di scalabilità | Crescita prevista delle interazioni tra agenti | Funzionalità di streaming di MCP |

# Selezione dei protocolli agentici

Per la maggior parte delle organizzazioni che creano sistemi di agenti di produzione, il Model Context Protocol (MCP) offre la base per la comunicazione più completa e ben supportata. agent-to-agent MCP beneficia dei contributi attivi allo sviluppo AWS e della comunità open source.

La selezione dei protocolli agentici giusti è importante per le organizzazioni che desiderano implementare l'intelligenza artificiale agentica in modo efficace. Le considerazioni variano in base al contesto organizzativo.

## Considerazioni sul protocollo aziendale

Le aziende dovrebbero prendere in considerazione le seguenti azioni:

- Dai priorità ai protocolli aperti come MCP per implementazioni strategiche a lungo termine con agenti.
- Implementa livelli di astrazione quando utilizzi protocolli proprietari per facilitare le migrazioni future.
- Partecipa allo sviluppo degli standard per influenzare l'evoluzione del protocollo.
- Prendi in considerazione approcci ibridi che utilizzano protocolli aperti per l'infrastruttura di base e protocolli proprietari per casi d'uso specifici.

## Considerazioni relative all'avvio e al protocollo SMB

Le startup e le aziende small-to-medium (PMI) dovrebbero prendere in considerazione le seguenti azioni:

- Bilancia velocità e flessibilità iniziando con protocolli proprietari ben supportati per uno sviluppo rapido.
- Pianifica i percorsi di migrazione per utilizzare standard più aperti man mano che le tue esigenze maturano.
- Valuta le tendenze di adozione dei protocolli per evitare di investire in standard in declino.
- Prendi in considerazione i servizi gestiti che astraggono la complessità del protocollo.

## Considerazioni sui protocolli governativi e di settore regolamentati

Le amministrazioni e le industrie regolamentate dovrebbero prendere in considerazione le seguenti azioni:

- Sottolineate l'importanza degli standard aperti per garantire l'accesso a lungo termine ed evitare il vincolo del fornitore.
- Assegna priorità ai protocolli con solidi modelli di sicurezza e meccanismi di autenticazione.
- Prendi in considerazione le implicazioni sulla sovranità dei dati dei modelli di implementazione remota e locale.
- Documenta le decisioni relative ai protocolli per i requisiti di conformità e governance.

## Strategia di implementazione per i protocolli agentici

Per implementare efficacemente i protocolli agentici in tutta l'organizzazione, prendete in considerazione i seguenti passaggi strategici:

1. Inizia con l'allineamento degli standard: adotta protocolli aperti consolidati, ove possibile.
2. Crea livelli di astrazione: implementa adattatori tra i tuoi sistemi e protocolli specifici.
3. Contribuisci agli standard aperti: partecipa alle comunità di sviluppo dei protocolli.
4. Monitora l'evoluzione del protocollo: rimani informato sugli standard e sugli aggiornamenti emergenti.
5. Verifica regolarmente l'interoperabilità: verifica che le implementazioni rimangano compatibili.

## Guida introduttiva a MCP

Per implementare il Model Context Protocol (MCP) nell'architettura del tuo agente, esegui le seguenti azioni:

1. Esplora le implementazioni MCP in framework come l'SDK. Strands Agents
2. Consulta la documentazione tecnica del Model Context Protocol.
3. Leggi Open Protocols for Agent Interoperability Part 1: Inter-Agent Communication on MCP (AWS Blog) per saperne di più sull'interoperabilità degli agenti.
4. Unisciti alla community MCP per influenzare l'evoluzione del protocollo.

MCP fornisce un livello di comunicazione che consente agli agenti di interagire con dati e servizi esterni e può essere utilizzato anche per consentire agli agenti di interagire con altri agenti. L'implementazione del trasporto HTTP Streamable del protocollo offre agli sviluppatori un set completo di modelli di interazione senza dover reinventare la ruota. Questi modelli supportano sia i request/response flussi stateless che la gestione delle sessioni stateful con persistent. IDs

Adottando protocolli aperti come MCP, consentite alla vostra organizzazione di creare sistemi di agenti che rimangano flessibili, interoperabili e adattabili di pari passo con l'evoluzione della tecnologia AI.

# Strumenti

Gli agenti di intelligenza artificiale offrono valore interagendo con strumenti esterni e fonti di dati per eseguire attività utili. APIs La giusta strategia di integrazione degli strumenti ha un impatto diretto sulle capacità, sul livello di sicurezza e sulla flessibilità a lungo termine degli agenti.

Questa sezione vi aiuta a navigare nel panorama dell'integrazione degli strumenti, con particolare attenzione agli standard aperti che massimizzano la libertà e la flessibilità. La sezione evidenzia il Model Context Protocol (MCP) per l'integrazione degli strumenti e esamina gli strumenti specifici del framework e i meta-strumenti specializzati che migliorano i flussi di lavoro degli agenti.

In questa sezione



## Categorie di strumenti

I sistemi Building Agent comprendono tre categorie principali di strumenti.

### Strumenti basati su protocolli

Gli strumenti basati su protocolli utilizzano protocolli di comunicazione standardizzati: agent-to-tool

- Strumenti MCP: strumenti standard aperti che funzionano su più framework con opzioni di esecuzione sia locali che remote.
- OpenAIfunction calling: strumenti proprietari specifici per i modelli. OpenAI
- Anthropictools — Strumenti proprietari specifici per i modelli di Anthropic Claude.

### Strumenti nativi del framework

Gli strumenti nativi del framework sono integrati direttamente in framework di agenti specifici:

- Strands Agents Pythonstrumenti: strumenti leggeri e specifici quick-to-implement per il framework. Strands Agents
- LangChainstrumenti: strumenti Python basati su strumenti strettamente integrati con l'LangChainecosistema.
- LlamaIndexstrumenti: strumenti ottimizzati per il recupero e l'elaborazione dei dati all'interno. LlamaIndex

### Meta-strumenti

I meta-strumenti migliorano i flussi di lavoro degli agenti senza intraprendere azioni esterne direttamente:

- Strumenti per il flusso di lavoro: gestisci il flusso di esecuzione degli agenti, la logica di ramificazione e la gestione dello stato.
- Strumenti grafici per agenti: coordina più agenti in flussi di lavoro complessi.
- Strumenti di memoria: forniscono l'archiviazione e il recupero persistenti delle informazioni tra le sessioni degli agenti.
- Strumenti di riflessione: consentono agli agenti di analizzare e migliorare le proprie prestazioni.

## Strumenti basati su protocolli

Quando si considerano gli strumenti basati sul protocollo, il Model Context Protocol (MCP) fornisce la base più completa e flessibile per l'integrazione degli strumenti. Come affermato nel post del blog AWS Open Source sull'interoperabilità degli agenti, AWS ha adottato MCP come protocollo strategico, contribuendo attivamente al suo sviluppo.

La tabella seguente descrive le opzioni per l'implementazione degli strumenti MCP.

| Modello di distribuzione | Descrizione | Ideale per | Implementazione |
| --- | --- | --- | --- |
| Basato su uno studio locale | Gli strumenti vengono eseguiti nello stesso processo dell'agente | Sviluppo, test e strumenti semplici | Rapido da implementare senza sovraccarico di rete |

| | | | |
|---|---|---|---|
| Basato su eventi inviati dal server locale (SSE) | Gli strumenti vengono eseguiti localment e ma comunicano tramite HTTP | Strumenti locali più complessi con separazione delle preoccupazioni | Migliore isolamento ma comunque bassa latenza |
| Basato su SSE remoto | Gli strumenti vengono eseguiti su server remoti | Ambienti di produzion e e strumenti condivisi | Scalabile e gestito centralmente |

Il Model Context Protocol SDKs ufficiale è disponibile per la creazione di strumenti MCP:

- PythonSDK: implementazione completa con supporto completo del protocollo
- TypeScriptSDK —JavaScript/TypeScriptimplementazione per applicazioni web
- JavaSDK: implementazione Java per applicazioni aziendali

Questi SDKs forniscono gli elementi costitutivi per la creazione di strumenti compatibili con MCP nel linguaggio preferito, con implementazioni coerenti delle specifiche del protocollo.

Inoltre, AWS ha implementato MCP nell'SDK. Strands Agents L'Strands AgentsSDK offre un modo semplice per creare e utilizzare strumenti compatibili con MCP. La documentazione completa è disponibile nel repository. Strands Agents GitHub Per casi d'uso più semplici o quando si lavora al di fuori del Strands Agents framework, gli MCP ufficiali SDKs offrono implementazioni dirette del protocollo in più lingue.

## Funzionalità di sicurezza degli strumenti MCP

Le funzionalità di sicurezza degli strumenti MCP includono quanto segue:

- OAuth Autenticazione 2.0/2.1: autenticazione standard del settore
- Ambito delle autorizzazioni: controllo granulare degli accessi per gli strumenti
- Scoperta delle funzionalità degli strumenti: individuazione dinamica degli strumenti disponibili
- Gestione strutturata degli errori: modelli di errore coerenti

## Guida introduttiva agli strumenti MCP

Per implementare MCP per l'integrazione degli strumenti, intraprendi le seguenti azioni:

1. Esplora l'Strands AgentsSDK per un'implementazione MCP pronta per la produzione.
2. Consulta la documentazione tecnica MCP per comprendere i concetti fondamentali.
3. Usa gli esempi pratici descritti in questo post del blog AWS Open Source.
4. Inizia con semplici strumenti locali prima di passare a strumenti remoti.
5. Unisciti alla community MCP per influenzare l'evoluzione del protocollo.

## Strumenti nativi del framework

Sebbene il Model Context Protocol (MCP) fornisca la base più flessibile, gli strumenti nativi del framework offrono vantaggi per casi d'uso specifici.

L'Strands AgentsSDK offre strumenti Python basati su un design leggero che richiede un sovraccarico minimo per operazioni semplici. Consentono un'implementazione rapida e consentono agli sviluppatori di creare strumenti con poche righe di codice. Inoltre, sono strettamente integrati per funzionare perfettamente all'interno del Strands Agents framework.

L'esempio seguente mostra come creare un semplice strumento meteorologico utilizzando. Strands Agents Gli sviluppatori possono trasformare rapidamente Python le funzioni in strumenti accessibili tramite agenti con un sovraccarico di codice minimo e generare automaticamente la documentazione appropriata dalla docstring della funzione.

```
#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."
```

Per la prototipazione rapida o per casi d'uso semplici, gli strumenti nativi del framework possono accelerare lo sviluppo. Tuttavia, per i sistemi di produzione, gli strumenti MCP offrono una migliore interoperabilità e flessibilità future rispetto agli strumenti nativi del framework.

La tabella seguente fornisce una panoramica di altri strumenti specifici del framework.

| Framework | Tipo di utensile | Vantaggi | Considerazioni |
| --- | --- | --- | --- |
| AutoGen | Definizioni delle funzioni | Forte supporto multiagente | Microsoftecosistema |
| LangChain | Pythonclassi | Ampio ecosistema di strumenti predefiniti | Framework lock-in |
| LlamaIndex | Funzioni Python | Ottimizzato per le operazioni relative ai dati | Limitato a LlamaIndex |

# Meta-strumenti

I meta-strumenti non interagiscono direttamente con i sistemi esterni. Al contrario, migliorano le capacità degli agenti implementando modelli agentici. Questa sezione illustra il flusso di lavoro, il grafico degli agenti e i meta-strumenti di memoria.

## Meta-strumenti per il flusso di lavoro

I meta-strumenti del flusso di lavoro gestiscono il flusso di esecuzione degli agenti:

- Gestione dello stato: mantenimento del contesto tra più interazioni tra agenti
- Logica di ramificazione: abilita percorsi di esecuzione condizionali
- Meccanismi di ripetizione dei tentativi: gestisci gli errori con sofisticate strategie di riprova

I framework di esempio con meta-strumenti per il flusso di lavoro includono funzionalità di workflow. LangGraphStrands Agents

## Meta-strumenti Agent Graph

I meta-strumenti Agent Graph coordinano più agenti che lavorano insieme:

- Delega delle attività: assegna attività secondarie ad agenti specializzati
- Aggregazione dei risultati: combina gli output di più agenti
- Risoluzione dei conflitti: risolvi i disaccordi tra agenti

I framework apprezzano AutoGene sono CrewAIspecializzati nella coordinazione grafica degli agenti.

### Meta-strumenti di memoria

I meta-strumenti di memoria forniscono archiviazione e recupero persistenti:

- Cronologia delle conversazioni: mantieni il contesto tra le sessioni
- Basi di conoscenza: archivia e recupera informazioni specifiche del dominio
- Archivi vettoriali: abilitano le funzionalità di ricerca semantica

Il sistema di risorse di MCP offre un modo standardizzato per implementare meta-strumenti di memoria che funzionano su diversi framework di agenti.

## Strategia di integrazione degli strumenti

La strategia di integrazione degli strumenti scelta influisce direttamente su ciò che gli agenti possono realizzare e sulla facilità con cui il sistema può evolversi. Dai priorità ai protocolli aperti come il Model Context Protocol (MCP) utilizzando strategicamente strumenti e meta-strumenti nativi del framework. In questo modo, puoi creare un ecosistema di strumenti che rimanga flessibile e potente man mano che la tecnologia AI avanza.

Il seguente approccio strategico all'integrazione degli strumenti massimizza la flessibilità soddisfacendo al contempo le esigenze immediate dell'organizzazione:

1. Adottate MCP come base: MCP offre un modo standardizzato per collegare gli agenti a strumenti con potenti funzionalità di sicurezza. Inizia con MCP come protocollo di strumenti principale per:
   - Strumenti strategici che verranno utilizzati in più implementazioni di agenti.
   - Strumenti sensibili alla sicurezza che richiedono un'autenticazione e un'autorizzazione solide.
   - Strumenti che richiedono l'esecuzione remota in ambienti di produzione.
2. Usa strumenti nativi del framework quando appropriato: prendi in considerazione gli strumenti nativi del framework per:
   - Prototipazione rapida durante lo sviluppo iniziale.
   - Strumenti semplici non critici con requisiti di sicurezza minimi.
   - Funzionalità specifiche del framework che sfrutta funzionalità uniche.
3. Implementa meta-strumenti per flussi di lavoro complessi: aggiungi meta-strumenti per migliorare l'architettura degli agenti:

- Inizia in modo semplice con modelli di flusso di lavoro di base.
- Aggiungi complessità man mano che i tuoi casi d'uso maturano.
- Standardizza le interfacce tra agenti e meta-strumenti.

4. Pianifica l'evoluzione: costruisci pensando alla flessibilità futura:
   - Interfacce degli strumenti documentali indipendentemente dalle implementazioni.
   - Crea livelli di astrazione tra agenti e strumenti.
   - Stabilisci percorsi di migrazione dai protocolli proprietari a quelli aperti.

# Le migliori pratiche di sicurezza per l'integrazione degli strumenti

L'integrazione degli strumenti ha un impatto diretto sul tuo livello di sicurezza. Questa sezione descrive le best practice da prendere in considerazione per la propria organizzazione.

## Autenticazione e autorizzazione

Utilizzate i seguenti robusti controlli di accesso:

- Usa OAuth 2.0/2.1: implementa l'autenticazione standard del settore per gli strumenti remoti.
- Implementa il privilegio minimo: concedi agli strumenti solo le autorizzazioni di cui hanno bisogno.
- Ruota le credenziali: aggiorna regolarmente le chiavi API e i token di accesso.

## Protezione dei dati

Per contribuire alla protezione dei dati, adotta le seguenti misure:

- Convalida input e output: implementa la convalida dello schema per tutte le interazioni con gli strumenti.
- Crittografa i dati sensibili: utilizza TLS per tutte le comunicazioni con strumenti remoti.
- Implementa la riduzione al minimo dei dati: trasmetti solo le informazioni necessarie agli strumenti.

## Monitoraggio e controllo

Mantieni la visibilità e il controllo utilizzando questi meccanismi:

- Registra tutte le chiamate agli strumenti: mantieni percorsi di controllo completi.

- Monitoraggio delle anomalie: rileva modelli di utilizzo degli strumenti insoliti.
- Implementa la limitazione della velocità: previene gli abusi tramite un numero eccessivo di strumenti.

Il modello di sicurezza MCP risponde a queste preoccupazioni in modo completo. Per ulteriori informazioni, consulta Considerazioni sulla sicurezza nella documentazione MCP.

# Conclusioni

Il panorama dell'intelligenza artificiale agentica continua a evolversi rapidamente, offrendo alle organizzazioni nuovi modi potenti per costruire sistemi intelligenti e autonomi. Questa guida ha esplorato tre componenti essenziali per un'implementazione di successo: framework che forniscono le basi, protocolli che consentono la comunicazione e strumenti che estendono le funzionalità.

Man mano che i framework maturano, è possibile aspettarsi una maggiore interoperabilità, la standardizzazione su protocolli come il Model Context Protocol (MCP) e funzionalità di orchestrazione più sofisticate per agenti autonomi. Le organizzazioni che oggi acquisiscono competenze con questi framework saranno ben posizionate per creare agenti sempre più autonomi e intelligenti che offrano un valore aziendale significativo.

La scelta dei protocolli degli agenti rappresenta una decisione strategica che bilancia le esigenze di sviluppo immediate con la flessibilità e l'interoperabilità a lungo termine. Dando priorità ai protocolli aperti e creando livelli di astrazione appropriati, le organizzazioni possono creare sistemi di agenti che rimangono adattabili alle tecnologie in evoluzione e soddisfano al contempo i requisiti aziendali attuali.

Per la maggior parte delle organizzazioni, MCP rappresenta una solida base grazie al suo standard aperto, all'ecosistema in crescita, al supporto per i modelli di agent-to-agent comunicazione e alle capacità di integrazione degli strumenti. AWS ha adottato l'MCP come protocollo strategico, contribuendo attivamente al suo sviluppo e alla sua implementazione attraverso servizi come l'SDK. Strands Agents Utilizzando MCP insieme a strumenti e meta-strumenti nativi del framework appropriati, è possibile creare sistemi di agenti che offrono valore immediato pur rimanendo adattabili alle innovazioni future.

# Risorse

Utilizza le seguenti AWS e altre risorse relative allo sviluppo di agenti autonomi.

## AWS Blog

- Le migliori pratiche per creare solide applicazioni di intelligenza artificiale generativa con Amazon Bedrock Agents — Parte 1
- Best practice per creare solide applicazioni di intelligenza artificiale generativa con Amazon Bedrock Agents — Parte 2
- Presentazione Strands Agents di un AI Agents SDK open source
- Protocolli aperti per l'interoperabilità degli agenti, parte 1: Comunicazione tra agenti su MCP
- AWS Transform per.NET, il primo servizio di intelligenza artificiale agentica per modernizzare le applicazioni.NET su larga scala
- AWS Riepilogo settimanale: Strands Agents

## AWS Guida prescrittiva

- Rendere operativa l'intelligenza artificiale agentica su AWS
- Fondamenti dell'intelligenza artificiale agentica su AWS
- Modelli e flussi di lavoro di intelligenza artificiale agentica su AWS
- Creazione di architetture serverless per l'intelligenza artificiale agentica su AWS
- Creazione di architetture multi-tenant per l'intelligenza artificiale agentica su AWS
- Opzioni e architetture di Retrieval Augmented Generation su AWS

## AWS risorse

- Documentazione Amazon Bedrock
- Documentazione Amazon Nova
- AWS Server MCP () GitHub

# Altre risorse

- AutoGendocumentazione () Microsoft
- Creazione di agenti efficaci (Anthropic)
- CrewAI GitHubdeposito
- Documentazione di LangChain
- LangGraphpiattaforma
- documentazione del Model Context Protocol
- Documentazione di Strands Agents
- Strands AgentsPanoramica degli strumenti
- Strands AgentsGuida rapida

# Cronologia dei documenti

La tabella seguente descrive le modifiche significative apportate a questa guida. Per ricevere notifiche sugli aggiornamenti futuri, puoi abbonarti a un feed RSS.

| Modifica | Descrizione | Data |
| --- | --- | --- |
| Pubblicazione iniziale | — | 14 luglio 2025 |

# AWS Glossario delle linee guida prescrittive

I seguenti sono termini di uso comune nelle strategie, nelle guide e nei modelli forniti da AWS Prescriptive Guidance. Per suggerire voci, utilizza il link Fornisci feedback alla fine del glossario.

## Numeri

7 R

Sette strategie di migrazione comuni per trasferire le applicazioni sul cloud. Queste strategie si basano sulle 5 R identificate da Gartner nel 2011 e sono le seguenti:

- Rifattorizzare/riprogettare: trasferisci un'applicazione e modifica la sua architettura sfruttando appieno le funzionalità native del cloud per migliorare l'agilità, le prestazioni e la scalabilità. Ciò comporta in genere la portabilità del sistema operativo e del database. Esempio: migra il tuo database Oracle locale all'edizione compatibile con Amazon Aurora PostgreSQL.
- Ridefinire la piattaforma (lift and reshape): trasferisci un'applicazione nel cloud e introduci un certo livello di ottimizzazione per sfruttare le funzionalità del cloud. Esempio: migra il tuo database Oracle locale ad Amazon Relational Database Service (Amazon RDS) per Oracle in. Cloud AWS
- Riacquistare (drop and shop): passa a un prodotto diverso, in genere effettuando la transizione da una licenza tradizionale a un modello SaaS. Esempio: migra il tuo sistema di gestione delle relazioni con i clienti (CRM) su Salesforce.com.
- Eseguire il rehosting (lift and shift): trasferisci un'applicazione sul cloud senza apportare modifiche per sfruttare le funzionalità del cloud. Esempio: migra il database Oracle locale su Oracle su un'istanza in. EC2 Cloud AWS
- Trasferire (eseguire il rehosting a livello hypervisor): trasferisci l'infrastruttura sul cloud senza acquistare nuovo hardware, riscrivere le applicazioni o modificare le operazioni esistenti. Si esegue la migrazione dei server da una piattaforma locale a un servizio cloud per la stessa piattaforma. Esempio: migra un'applicazione suMicrosoft Hyper-V. AWS
- Riesaminare (mantenere): mantieni le applicazioni nell'ambiente di origine. Queste potrebbero includere applicazioni che richiedono una rifattorizzazione significativa che desideri rimandare a un momento successivo e applicazioni legacy che desideri mantenere, perché non vi è alcuna giustificazione aziendale per effettuarne la migrazione.
- Ritirare: disattiva o rimuovi le applicazioni che non sono più necessarie nell'ambiente di origine.

# A

ABAC

Vedi controllo degli accessi basato sugli attributi.

servizi astratti

Vedi servizi gestiti.

ACIDO

Vedi atomicità, consistenza, isolamento, durata.

migrazione attiva-attiva

Un metodo di migrazione del database in cui i database di origine e di destinazione vengono mantenuti sincronizzati (utilizzando uno strumento di replica bidirezionale o operazioni di doppia scrittura) ed entrambi i database gestiscono le transazioni provenienti dalle applicazioni di connessione durante la migrazione. Questo metodo supporta la migrazione in piccoli batch controllati anziché richiedere una conversione una tantum. È più flessibile ma richiede più lavoro rispetto alla migrazione attiva-passiva.

migrazione attiva-passiva

Un metodo di migrazione del database in cui i database di origine e di destinazione vengono mantenuti sincronizzati, ma solo il database di origine gestisce le transazioni provenienti dalle applicazioni di connessione mentre i dati vengono replicati nel database di destinazione. Il database di destinazione non accetta alcuna transazione durante la migrazione.

funzione di aggregazione

Una funzione SQL che opera su un gruppo di righe e calcola un singolo valore restituito per il gruppo. Esempi di funzioni aggregate includono `SUM` e. `MAX`

Intelligenza artificiale

Vedi intelligenza artificiale.

AIOps

Guarda le operazioni di intelligenza artificiale.

anonimizzazione

Il processo di eliminazione permanente delle informazioni personali in un set di dati. L'anonimizzazione può aiutare a proteggere la privacy personale. I dati anonimi non sono più considerati dati personali.

anti-modello

Una soluzione utilizzata frequentemente per un problema ricorrente in cui la soluzione è controproducente, inefficace o meno efficace di un'alternativa.

controllo delle applicazioni

Un approccio alla sicurezza che consente l'uso solo di applicazioni approvate per proteggere un sistema dal malware.

portfolio di applicazioni

Una raccolta di informazioni dettagliate su ogni applicazione utilizzata da un'organizzazione, compresi i costi di creazione e manutenzione dell'applicazione e il relativo valore aziendale. Queste informazioni sono fondamentali per il processo di scoperta e analisi del portfolio e aiutano a identificare e ad assegnare la priorità alle applicazioni da migrare, modernizzare e ottimizzare.

intelligenza artificiale (IA)

Il campo dell'informatica dedicato all'uso delle tecnologie informatiche per svolgere funzioni cognitive tipicamente associate agli esseri umani, come l'apprendimento, la risoluzione di problemi e il riconoscimento di schemi. Per ulteriori informazioni, consulta la sezione Che cos'è l'intelligenza artificiale?

operazioni di intelligenza artificiale (AIOps)

Il processo di utilizzo delle tecniche di machine learning per risolvere problemi operativi, ridurre gli incidenti operativi e l'intervento umano e aumentare la qualità del servizio. Per ulteriori informazioni su come AIOps viene utilizzata nella strategia di AWS migrazione, consulta la guida all'integrazione delle operazioni.

crittografia asimmetrica

Un algoritmo di crittografia che utilizza una coppia di chiavi, una chiave pubblica per la crittografia e una chiave privata per la decrittografia. Puoi condividere la chiave pubblica perché non viene utilizzata per la decrittografia, ma l'accesso alla chiave privata deve essere altamente limitato.

atomicità, consistenza, isolamento, durabilità (ACID)

Un insieme di proprietà del software che garantiscono la validità dei dati e l'affidabilità operativa di un database, anche in caso di errori, interruzioni di corrente o altri problemi.

Controllo degli accessi basato su attributi (ABAC)

La pratica di creare autorizzazioni dettagliate basate su attributi utente, come reparto, ruolo professionale e nome del team. Per ulteriori informazioni, consulta ABAC AWS nella documentazione AWS Identity and Access Management (IAM).

fonte di dati autorevole

Una posizione in cui è archiviata la versione principale dei dati, considerata la fonte di informazioni più affidabile. È possibile copiare i dati dalla fonte di dati autorevole in altre posizioni allo scopo di elaborarli o modificarli, ad esempio anonimizzandoli, oscurandoli o pseudonimizzandoli.

Zona di disponibilità

Una posizione distinta all'interno di un edificio Regione AWS che è isolata dai guasti in altre zone di disponibilità e offre una connettività di rete economica e a bassa latenza verso altre zone di disponibilità nella stessa regione.

AWS Cloud Adoption Framework (CAF)AWS

Un framework di linee guida e best practice AWS per aiutare le organizzazioni a sviluppare un piano efficiente ed efficace per passare con successo al cloud. AWS CAF organizza le linee guida in sei aree di interesse chiamate prospettive: business, persone, governance, piattaforma, sicurezza e operazioni. Le prospettive relative ad azienda, persone e governance si concentrano sulle competenze e sui processi aziendali; le prospettive relative alla piattaforma, alla sicurezza e alle operazioni si concentrano sulle competenze e sui processi tecnici. Ad esempio, la prospettiva relativa alle persone si rivolge alle parti interessate che gestiscono le risorse umane (HR), le funzioni del personale e la gestione del personale. In questa prospettiva, AWS CAF fornisce linee guida per lo sviluppo delle persone, la formazione e le comunicazioni per aiutare a preparare l'organizzazione all'adozione del cloud di successo. Per ulteriori informazioni, consulta il sito web di AWS CAF e il white paper AWS CAF.

AWS Workload Qualification Framework (WQF)AWS

Uno strumento che valuta i carichi di lavoro di migrazione dei database, consiglia strategie di migrazione e fornisce stime del lavoro. AWS WQF è incluso in (). AWS Schema Conversion Tool AWS SCT Analizza gli schemi di database e gli oggetti di codice, il codice dell'applicazione, le dipendenze e le caratteristiche delle prestazioni e fornisce report di valutazione.

# B

bot difettoso

Un bot che ha lo scopo di interrompere o causare danni a individui o organizzazioni.

BCP

Vedi la pianificazione della continuità operativa.

grafico comportamentale

Una vista unificata, interattiva dei comportamenti delle risorse e delle interazioni nel tempo. Puoi utilizzare un grafico comportamentale con Amazon Detective per esaminare tentativi di accesso non riusciti, chiamate API sospette e azioni simili. Per ulteriori informazioni, consulta Dati in un grafico comportamentale nella documentazione di Detective.

sistema big-endian

Un sistema che memorizza per primo il byte più importante. Vedi anche endianness.

Classificazione binaria

Un processo che prevede un risultato binario (una delle due classi possibili). Ad esempio, il modello di machine learning potrebbe dover prevedere problemi come "Questa e-mail è spam o non è spam?" o "Questo prodotto è un libro o un'auto?"

filtro Bloom

Una struttura di dati probabilistica ed efficiente in termini di memoria che viene utilizzata per verificare se un elemento fa parte di un set.

distribuzioni blu/verdi

Una strategia di implementazione in cui si creano due ambienti separati ma identici. La versione corrente dell'applicazione viene eseguita in un ambiente (blu) e la nuova versione dell'applicazione nell'altro ambiente (verde). Questa strategia consente di ripristinare rapidamente il sistema con un impatto minimo.

bot

Un'applicazione software che esegue attività automatizzate su Internet e simula l'attività o l'interazione umana. Alcuni bot sono utili o utili, come i web crawler che indicizzano le informazioni su Internet. Alcuni altri bot, noti come bot dannosi, hanno lo scopo di disturbare o causare danni a individui o organizzazioni.

botnet

Reti di bot infettate da malware e controllate da un'unica parte, nota come bot herder o bot operator. Le botnet sono il meccanismo più noto per scalare i bot e il loro impatto.

ramo

Un'area contenuta di un repository di codice. Il primo ramo creato in un repository è il ramo principale. È possibile creare un nuovo ramo a partire da un ramo esistente e quindi sviluppare funzionalità o correggere bug al suo interno. Un ramo creato per sviluppare una funzionalità viene comunemente detto ramo di funzionalità. Quando la funzionalità è pronta per il rilascio, il ramo di funzionalità viene ricongiunto al ramo principale. Per ulteriori informazioni, consulta Informazioni sulle filiali (documentazione). GitHub

accesso break-glass

In circostanze eccezionali e tramite una procedura approvata, un mezzo rapido per consentire a un utente di accedere a un sito a Account AWS cui in genere non dispone delle autorizzazioni necessarie. Per ulteriori informazioni, vedere l'indicatore Implementate break-glass procedures nella guida Well-Architected AWS .

strategia brownfield

L'infrastruttura esistente nell'ambiente. Quando si adotta una strategia brownfield per un'architettura di sistema, si progetta l'architettura in base ai vincoli dei sistemi e dell'infrastruttura attuali. Per l'espansione dell'infrastruttura esistente, è possibile combinare strategie brownfield e greenfield.

cache del buffer

L'area di memoria in cui sono archiviati i dati a cui si accede con maggiore frequenza.

capacità di business

Azioni intraprese da un'azienda per generare valore (ad esempio vendite, assistenza clienti o marketing). Le architetture dei microservizi e le decisioni di sviluppo possono essere guidate dalle capacità aziendali. Per ulteriori informazioni, consulta la sezione Organizzazione in base alle funzionalità aziendali del whitepaper Esecuzione di microservizi containerizzati su AWS.

pianificazione della continuità operativa (BCP)

Un piano che affronta il potenziale impatto di un evento che comporta l'interruzione dell'attività, come una migrazione su larga scala, sulle operazioni e consente a un'azienda di riprendere rapidamente le operazioni.

# C

CAF

Vedi AWS Cloud Adoption Framework.

implementazione canaria

Il rilascio lento e incrementale di una versione agli utenti finali. Quando sei sicuro, distribuisci la nuova versione e sostituisci la versione corrente nella sua interezza.

CCoE

Vedi Cloud Center of Excellence.

CDC

Vedi Change Data Capture.

Change Data Capture (CDC)

Il processo di tracciamento delle modifiche a un'origine dati, ad esempio una tabella di database, e di registrazione dei metadati relativi alla modifica. È possibile utilizzare CDC per vari scopi, ad esempio il controllo o la replica delle modifiche in un sistema di destinazione per mantenere la sincronizzazione.

ingegneria del caos

Introduzione intenzionale di guasti o eventi dirompenti per testare la resilienza di un sistema. Puoi usare AWS Fault Injection Service (AWS FIS) per eseguire esperimenti che stressano i tuoi AWS carichi di lavoro e valutarne la risposta.

CI/CD

Vedi integrazione continua e distribuzione continua.

classificazione

Un processo di categorizzazione che aiuta a generare previsioni. I modelli di ML per problemi di classificazione prevedono un valore discreto. I valori discreti sono sempre distinti l'uno dall'altro. Ad esempio, un modello potrebbe dover valutare se in un'immagine è presente o meno un'auto.

crittografia lato client

Crittografia dei dati a livello locale, prima che il destinatario li Servizio AWS riceva.

Centro di eccellenza cloud (CCoE)

Un team multidisciplinare che guida le iniziative di adozione del cloud in tutta l'organizzazione, tra cui lo sviluppo di best practice per il cloud, la mobilitazione delle risorse, la definizione delle tempistiche di migrazione e la guida dell'organizzazione attraverso trasformazioni su larga scala. Per ulteriori informazioni, consulta gli CCoE post sull' Cloud AWS Enterprise Strategy Blog.

cloud computing

La tecnologia cloud generalmente utilizzata per l'archiviazione remota di dati e la gestione dei dispositivi IoT. Il cloud computing è generalmente collegato alla tecnologia di edge computing.

modello operativo cloud

In un'organizzazione IT, il modello operativo utilizzato per creare, maturare e ottimizzare uno o più ambienti cloud. Per ulteriori informazioni, consulta Building your Cloud Operating Model.

fasi di adozione del cloud

Le quattro fasi che le organizzazioni in genere attraversano quando migrano verso Cloud AWS:

- Progetto: esecuzione di alcuni progetti relativi al cloud per scopi di dimostrazione e apprendimento
- Fondamento: effettuare investimenti fondamentali per scalare l'adozione del cloud (ad esempio, creazione di una landing zone, definizione di una CCo E, definizione di un modello operativo)
- Migrazione: migrazione di singole applicazioni
- Reinvenzione: ottimizzazione di prodotti e servizi e innovazione nel cloud

Queste fasi sono state definite da Stephen Orban nel post sul blog The Journey Toward Cloud-First & the Stages of Adoption on the Enterprise Strategy. Cloud AWS Per informazioni su come si relazionano alla strategia di AWS migrazione, consulta la guida alla preparazione alla migrazione.

CMDB

Vedi database di gestione della configurazione.

repository di codice

Una posizione in cui il codice di origine e altri asset, come documentazione, esempi e script, vengono archiviati e aggiornati attraverso processi di controllo delle versioni. Gli archivi cloud più comuni includono GitHub oBitbucket Cloud. Ogni versione del codice è denominata ramo. In una

struttura a microservizi, ogni repository è dedicato a una singola funzionalità. Una singola pipeline CI/CD può utilizzare più repository.

cache fredda

Una cache del buffer vuota, non ben popolata o contenente dati obsoleti o irrilevanti. Ciò influisce sulle prestazioni perché l'istanza di database deve leggere dalla memoria o dal disco principale, il che richiede più tempo rispetto alla lettura dalla cache del buffer.

dati freddi

Dati a cui si accede raramente e che in genere sono storici. Quando si eseguono interrogazioni di questo tipo di dati, le interrogazioni lente sono in genere accettabili. Lo spostamento di questi dati su livelli o classi di storage meno costosi e con prestazioni inferiori può ridurre i costi.

visione artificiale (CV)

Un campo dell'intelligenza artificiale che utilizza l'apprendimento automatico per analizzare ed estrarre informazioni da formati visivi come immagini e video digitali. Ad esempio, Amazon SageMaker AI fornisce algoritmi di elaborazione delle immagini per CV.

deriva della configurazione

Per un carico di lavoro, una modifica della configurazione rispetto allo stato previsto. Potrebbe causare la non conformità del carico di lavoro e in genere è graduale e involontaria.

database di gestione della configurazione (CMDB)

Un repository che archivia e gestisce le informazioni su un database e il relativo ambiente IT, inclusi i componenti hardware e software e le relative configurazioni. In genere si utilizzano i dati di un CMDB nella fase di individuazione e analisi del portafoglio della migrazione.

Pacchetto di conformità

Una raccolta di AWS Config regole e azioni correttive che puoi assemblare per personalizzare i controlli di conformità e sicurezza. È possibile distribuire un pacchetto di conformità come singola entità in una regione Account AWS and o all'interno di un'organizzazione utilizzando un modello YAML. Per ulteriori informazioni, consulta i Conformance pack nella documentazione. AWS Config

integrazione e distribuzione continua (continuous integration and continuous delivery, CI/CD)

Il processo di automazione delle fasi di origine, compilazione, test, gestione temporanea e produzione del processo di rilascio del software. CI/CD viene comunemente descritto come una pipeline. CI/CD può aiutarvi ad automatizzare i processi, migliorare la produttività, migliorare

la qualità del codice e velocizzare le consegne. Per ulteriori informazioni, consulta Vantaggi della distribuzione continua. CD può anche significare continuous deployment (implementazione continua). Per ulteriori informazioni, consulta Distribuzione continua e implementazione continua a confronto.

CV

Vedi visione artificiale.

# D

dati a riposo

Dati stazionari nella rete, ad esempio i dati archiviati.

classificazione dei dati

Un processo per identificare e classificare i dati nella rete in base alla loro criticità e sensibilità. È un componente fondamentale di qualsiasi strategia di gestione dei rischi di sicurezza informatica perché consente di determinare i controlli di protezione e conservazione appropriati per i dati. La classificazione dei dati è un componente del pilastro della sicurezza nel AWS Well-Architected Framework. Per ulteriori informazioni, consulta Classificazione dei dati.

deriva dei dati

Una variazione significativa tra i dati di produzione e i dati utilizzati per addestrare un modello di machine learning o una modifica significativa dei dati di input nel tempo. La deriva dei dati può ridurre la qualità, l'accuratezza e l'equità complessive nelle previsioni dei modelli ML.

dati in transito

Dati che si spostano attivamente attraverso la rete, ad esempio tra le risorse di rete.

rete di dati

Un framework architettonico che fornisce la proprietà distribuita e decentralizzata dei dati con gestione e governance centralizzate.

riduzione al minimo dei dati

Il principio della raccolta e del trattamento dei soli dati strettamente necessari. Praticare la riduzione al minimo dei dati in the Cloud AWS può ridurre i rischi per la privacy, i costi e l'impronta di carbonio delle analisi.

perimetro dei dati

Una serie di barriere preventive nell' AWS ambiente che aiutano a garantire che solo le identità attendibili accedano alle risorse attendibili delle reti previste. Per ulteriori informazioni, consulta Building a data perimeter on. AWS

pre-elaborazione dei dati

Trasformare i dati grezzi in un formato che possa essere facilmente analizzato dal modello di ML. La pre-elaborazione dei dati può comportare la rimozione di determinate colonne o righe e l'eliminazione di valori mancanti, incoerenti o duplicati.

provenienza dei dati

Il processo di tracciamento dell'origine e della cronologia dei dati durante il loro ciclo di vita, ad esempio il modo in cui i dati sono stati generati, trasmessi e archiviati.

soggetto dei dati

Un individuo i cui dati vengono raccolti ed elaborati.

data warehouse

Un sistema di gestione dei dati che supporta la business intelligence, come l'analisi. I data warehouse contengono in genere grandi quantità di dati storici e vengono generalmente utilizzati per interrogazioni e analisi.

linguaggio di definizione del database (DDL)

Istruzioni o comandi per creare o modificare la struttura di tabelle e oggetti in un database.

linguaggio di manipolazione del database (DML)

Istruzioni o comandi per modificare (inserire, aggiornare ed eliminare) informazioni in un database.

DDL

Vedi linguaggio di definizione del database.

deep ensemble

Combinare più modelli di deep learning per la previsione. È possibile utilizzare i deep ensemble per ottenere una previsione più accurata o per stimare l'incertezza nelle previsioni.

deep learning

Un sottocampo del ML che utilizza più livelli di reti neurali artificiali per identificare la mappatura tra i dati di input e le variabili target di interesse.

defense-in-depth

Un approccio alla sicurezza delle informazioni in cui una serie di meccanismi e controlli di sicurezza sono accuratamente stratificati su una rete di computer per proteggere la riservatezza, l'integrità e la disponibilità della rete e dei dati al suo interno. Quando si adotta questa strategia AWS, si aggiungono più controlli a diversi livelli della AWS Organizations struttura per proteggere le risorse. Ad esempio, un defense-in-depth approccio potrebbe combinare l'autenticazione a più fattori, la segmentazione della rete e la crittografia.

amministratore delegato

In AWS Organizations, un servizio compatibile può registrare un account AWS membro per amministrare gli account dell'organizzazione e gestire le autorizzazioni per quel servizio. Questo account è denominato amministratore delegato per quel servizio specifico. Per ulteriori informazioni e un elenco di servizi compatibili, consulta Servizi che funzionano con AWS Organizations nella documentazione di AWS Organizations .

implementazione

Il processo di creazione di un'applicazione, di nuove funzionalità o di correzioni di codice disponibili nell'ambiente di destinazione. L'implementazione prevede l'applicazione di modifiche in una base di codice, seguita dalla creazione e dall'esecuzione di tale base di codice negli ambienti applicativi.

Ambiente di sviluppo

Vedi ambiente.

controllo di rilevamento

Un controllo di sicurezza progettato per rilevare, registrare e avvisare dopo che si è verificato un evento. Questi controlli rappresentano una seconda linea di difesa e avvisano l'utente in caso di eventi di sicurezza che aggirano i controlli preventivi in vigore. Per ulteriori informazioni, consulta Controlli di rilevamento in Implementazione dei controlli di sicurezza in AWS.

mappatura del flusso di valore dello sviluppo (DVSM)

Un processo utilizzato per identificare e dare priorità ai vincoli che influiscono negativamente sulla velocità e sulla qualità nel ciclo di vita dello sviluppo del software. DVSM estende il processo di

mappatura del flusso di valore originariamente progettato per pratiche di produzione snella. Si concentra sulle fasi e sui team necessari per creare e trasferire valore attraverso il processo di sviluppo del software.

gemello digitale

Una rappresentazione virtuale di un sistema reale, ad esempio un edificio, una fabbrica, un'attrezzatura industriale o una linea di produzione. I gemelli digitali supportano la manutenzione predittiva, il monitoraggio remoto e l'ottimizzazione della produzione.

tabella delle dimensioni

In uno schema a stella, una tabella più piccola che contiene gli attributi dei dati quantitativi in una tabella dei fatti. Gli attributi della tabella delle dimensioni sono in genere campi di testo o numeri discreti che si comportano come testo. Questi attributi vengono comunemente utilizzati per il vincolo delle query, il filtraggio e l'etichettatura dei set di risultati.

disastro

Un evento che impedisce a un carico di lavoro o a un sistema di raggiungere gli obiettivi aziendali nella sua sede principale di implementazione. Questi eventi possono essere disastri naturali, guasti tecnici o il risultato di azioni umane, come errori di configurazione involontari o attacchi di malware.

disaster recovery (DR)

La strategia e il processo utilizzati per ridurre al minimo i tempi di inattività e la perdita di dati causati da un disastro. Per ulteriori informazioni, consulta Disaster Recovery of Workloads su AWS: Recovery in the Cloud in the AWS Well-Architected Framework.

DML

Vedi linguaggio di manipolazione del database.

progettazione basata sul dominio

Un approccio allo sviluppo di un sistema software complesso collegandone i componenti a domini in evoluzione, o obiettivi aziendali principali, perseguiti da ciascun componente. Questo concetto è stato introdotto da Eric Evans nel suo libro,Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston: Addison-Wesley Professional, 2003). Per informazioni su come utilizzare la progettazione basata sul dominio con il modello del fico strangolatore (Strangler Fig), consulta la sezione Modernizzazione incrementale dei servizi Web Microsoft ASP.NET (ASMX) legacy utilizzando container e il Gateway Amazon API.

DOTT.

Vedi disaster recovery.

rilevamento della deriva

Tracciamento delle deviazioni da una configurazione di base. Ad esempio, è possibile AWS CloudFormation utilizzarlo per rilevare deviazioni nelle risorse di sistema oppure AWS Control Tower per rilevare cambiamenti nella landing zone che potrebbero influire sulla conformità ai requisiti di governance.

DVSM

Vedi la mappatura del flusso di valore dello sviluppo.

# E

EDA

Vedi analisi esplorativa dei dati.

MODIFICA

Vedi scambio elettronico di dati.

edge computing

La tecnologia che aumenta la potenza di calcolo per i dispositivi intelligenti all'edge di una rete IoT. Rispetto al cloud computing, l'edge computing può ridurre la latenza di comunicazione e migliorare i tempi di risposta.

scambio elettronico di dati (EDI)

Lo scambio automatizzato di documenti aziendali tra organizzazioni. Per ulteriori informazioni, vedere Cos'è lo scambio elettronico di dati.

crittografia

Un processo di elaborazione che trasforma i dati in chiaro, leggibili dall'uomo, in testo cifrato.

chiave crittografica

Una stringa crittografica di bit randomizzati generata da un algoritmo di crittografia. Le chiavi possono variare di lunghezza e ogni chiave è progettata per essere imprevedibile e univoca.

endianità

L'ordine in cui i byte vengono archiviati nella memoria del computer. I sistemi big-endian memorizzano per primo il byte più importante. I sistemi little-endian memorizzano per primo il byte meno importante.

endpoint

Vedi service endpoint.

servizio endpoint

Un servizio che puoi ospitare in un cloud privato virtuale (VPC) da condividere con altri utenti. Puoi creare un servizio endpoint con AWS PrivateLink e concedere autorizzazioni ad altri Account AWS o a AWS Identity and Access Management (IAM) principali. Questi account o principali possono connettersi al servizio endpoint in privato creando endpoint VPC di interfaccia. Per ulteriori informazioni, consulta Creazione di un servizio endpoint nella documentazione di Amazon Virtual Private Cloud (Amazon VPC).

pianificazione delle risorse aziendali (ERP)

Un sistema che automatizza e gestisce i processi aziendali chiave (come contabilità, MES e gestione dei progetti) per un'azienda.

crittografia envelope

Il processo di crittografia di una chiave di crittografia con un'altra chiave di crittografia. Per ulteriori informazioni, vedete Envelope encryption nella documentazione AWS Key Management Service (AWS KMS).

ambiente

Un'istanza di un'applicazione in esecuzione. Di seguito sono riportati i tipi di ambiente più comuni nel cloud computing:

- ambiente di sviluppo: un'istanza di un'applicazione in esecuzione disponibile solo per il team principale responsabile della manutenzione dell'applicazione. Gli ambienti di sviluppo vengono utilizzati per testare le modifiche prima di promuoverle negli ambienti superiori. Questo tipo di ambiente viene talvolta definito ambiente di test.
- ambienti inferiori: tutti gli ambienti di sviluppo di un'applicazione, ad esempio quelli utilizzati per le build e i test iniziali.

- ambiente di produzione: un'istanza di un'applicazione in esecuzione a cui gli utenti finali possono accedere. In una CI/CD pipeline, l'ambiente di produzione è l'ultimo ambiente di distribuzione.
- ambienti superiori: tutti gli ambienti a cui possono accedere utenti diversi dal team di sviluppo principale. Si può trattare di un ambiente di produzione, ambienti di preproduzione e ambienti per i test di accettazione da parte degli utenti.

epica

Nelle metodologie agili, categorie funzionali che aiutano a organizzare e dare priorità al lavoro. Le epiche forniscono una descrizione di alto livello dei requisiti e delle attività di implementazione. Ad esempio, le epopee della sicurezza AWS CAF includono la gestione delle identità e degli accessi, i controlli investigativi, la sicurezza dell'infrastruttura, la protezione dei dati e la risposta agli incidenti. Per ulteriori informazioni sulle epiche, consulta la strategia di migrazione AWS , consulta la guida all'implementazione del programma.

ERP

Vedi pianificazione delle risorse aziendali.

analisi esplorativa dei dati (EDA)

Il processo di analisi di un set di dati per comprenderne le caratteristiche principali. Si raccolgono o si aggregano dati e quindi si eseguono indagini iniziali per trovare modelli, rilevare anomalie e verificare ipotesi. L'EDA viene eseguita calcolando statistiche di riepilogo e creando visualizzazioni di dati.

# F

tabella dei fatti

Il tavolo centrale in uno schema a stella. Memorizza dati quantitativi sulle operazioni aziendali. In genere, una tabella dei fatti contiene due tipi di colonne: quelle che contengono misure e quelle che contengono una chiave esterna per una tabella di dimensioni.

fallire velocemente

Una filosofia che utilizza test frequenti e incrementali per ridurre il ciclo di vita dello sviluppo. È una parte fondamentale di un approccio agile.

limite di isolamento dei guasti

Nel Cloud AWS, un limite come una zona di disponibilità Regione AWS, un piano di controllo o un piano dati che limita l'effetto di un errore e aiuta a migliorare la resilienza dei carichi di lavoro. Per ulteriori informazioni, consulta AWS Fault Isolation Boundaries.

ramo di funzionalità

Vedi filiale.

caratteristiche

I dati di input che usi per fare una previsione. Ad esempio, in un contesto di produzione, le caratteristiche potrebbero essere immagini acquisite periodicamente dalla linea di produzione.

importanza delle caratteristiche

Quanto è importante una caratteristica per le previsioni di un modello. Di solito viene espresso come punteggio numerico che può essere calcolato con varie tecniche, come Shapley Additive Explanations (SHAP) e gradienti integrati. Per ulteriori informazioni, consulta Interpretabilità del modello di machine learning con AWS.

trasformazione delle funzionalità

Per ottimizzare i dati per il processo di machine learning, incluso l'arricchimento dei dati con fonti aggiuntive, il dimensionamento dei valori o l'estrazione di più set di informazioni da un singolo campo di dati. Ciò consente al modello di ML di trarre vantaggio dai dati. Ad esempio, se suddividi la data "2021-05-27 00:15:37" in "2021", "maggio", "giovedì" e "15", puoi aiutare l'algoritmo di apprendimento ad apprendere modelli sfumati associati a diversi componenti dei dati.

prompt con pochi scatti

Fornire a un LLM un numero limitato di esempi che dimostrino l'attività e il risultato desiderato prima di chiedergli di eseguire un'attività simile. Questa tecnica è un'applicazione dell'apprendimento contestuale, in cui i modelli imparano da esempi (immagini) incorporati nei prompt. I prompt con pochi passaggi possono essere efficaci per attività che richiedono una formattazione, un ragionamento o una conoscenza del dominio specifici. Vedi anche zero-shot prompting.

FGAC

Vedi il controllo granulare degli accessi.

controllo granulare degli accessi (FGAC)

L'uso di più condizioni per consentire o rifiutare una richiesta di accesso.

migrazione flash-cut

Un metodo di migrazione del database che utilizza la replica continua dei dati tramite l'acquisizione dei dati delle modifiche per migrare i dati nel più breve tempo possibile, anziché utilizzare un approccio graduale. L'obiettivo è ridurre al minimo i tempi di inattività.

FM

Vedi modello di base.

modello di fondazione (FM)

Una grande rete neurale di deep learning che si è addestrata su enormi set di dati generalizzati e non etichettati. FMs sono in grado di svolgere un'ampia varietà di attività generali, come comprendere il linguaggio, generare testo e immagini e conversare in linguaggio naturale. Per ulteriori informazioni, consulta Cosa sono i modelli Foundation.

# G

AI generativa

Un sottoinsieme di modelli di intelligenza artificiale che sono stati addestrati su grandi quantità di dati e che possono utilizzare un semplice prompt di testo per creare nuovi contenuti e artefatti, come immagini, video, testo e audio. Per ulteriori informazioni, consulta Cos'è l'IA generativa.

blocco geografico

Vedi restrizioni geografiche.

limitazioni geografiche (blocco geografico)

In Amazon CloudFront, un'opzione per impedire agli utenti di determinati paesi di accedere alle distribuzioni di contenuti. Puoi utilizzare un elenco consentito o un elenco di blocco per specificare i paesi approvati e vietati. Per ulteriori informazioni, consulta Limitare la distribuzione geografica dei contenuti nella CloudFront documentazione.

Flusso di lavoro di GitFlow

Un approccio in cui gli ambienti inferiori e superiori utilizzano rami diversi in un repository di codice di origine. Il flusso di lavoro Gitflow è considerato obsoleto e il flusso di lavoro basato su trunk è l'approccio moderno e preferito.

immagine dorata

Un'istantanea di un sistema o di un software che viene utilizzata come modello per distribuire nuove istanze di quel sistema o software. Ad esempio, nella produzione, un'immagine dorata può essere utilizzata per fornire software su più dispositivi e contribuire a migliorare la velocità, la scalabilità e la produttività nelle operazioni di produzione dei dispositivi.

strategia greenfield

L'assenza di infrastrutture esistenti in un nuovo ambiente. Quando si adotta una strategia greenfield per un'architettura di sistema, è possibile selezionare tutte le nuove tecnologie senza il vincolo della compatibilità con l'infrastruttura esistente, nota anche come brownfield. Per l'espansione dell'infrastruttura esistente, è possibile combinare strategie brownfield e greenfield.

guardrail

Una regola di alto livello che aiuta a governare le risorse, le politiche e la conformità tra le unità organizzative (). OUs I guardrail preventivi applicano le policy per garantire l'allineamento agli standard di conformità. Vengono implementati utilizzando le policy di controllo dei servizi e i limiti delle autorizzazioni IAM. I guardrail di rilevamento rilevano le violazioni delle policy e i problemi di conformità e generano avvisi per porvi rimedio. Sono implementati utilizzando Amazon AWS Config AWS Security Hub GuardDuty AWS Trusted Advisor, Amazon Inspector e controlli personalizzati AWS Lambda .

# H

AH

Vedi disponibilità elevata.

migrazione di database eterogenea

Migrazione del database di origine in un database di destinazione che utilizza un motore di database diverso (ad esempio, da Oracle ad Amazon Aurora). La migrazione eterogenea fa in

genere parte di uno sforzo di riprogettazione e la conversione dello schema può essere un'attività complessa. AWS offre AWS SCT che aiuta con le conversioni dello schema.

alta disponibilità (HA)

La capacità di un carico di lavoro di funzionare in modo continuo, senza intervento, in caso di sfide o disastri. I sistemi HA sono progettati per il failover automatico, fornire costantemente prestazioni di alta qualità e gestire carichi e guasti diversi con un impatto minimo sulle prestazioni.

modernizzazione storica

Un approccio utilizzato per modernizzare e aggiornare i sistemi di tecnologia operativa (OT) per soddisfare meglio le esigenze dell'industria manifatturiera. Uno storico è un tipo di database utilizzato per raccogliere e archiviare dati da varie fonti in una fabbrica.

dati di esclusione

Una parte di dati storici etichettati che viene trattenuta da un set di dati utilizzata per addestrare un modello di apprendimento automatico. È possibile utilizzare i dati di holdout per valutare le prestazioni del modello confrontando le previsioni del modello con i dati di holdout.

migrazione di database omogenea

Migrazione del database di origine in un database di destinazione che condivide lo stesso motore di database (ad esempio, da Microsoft SQL Server ad Amazon RDS per SQL Server). La migrazione omogenea fa in genere parte di un'operazione di rehosting o ridefinizione della piattaforma. Per migrare lo schema è possibile utilizzare le utilità native del database.

dati caldi

Dati a cui si accede frequentemente, come dati in tempo reale o dati di traduzione recenti. Questi dati richiedono in genere un livello o una classe di storage ad alte prestazioni per fornire risposte rapide alle query.

hotfix

Una soluzione urgente per un problema critico in un ambiente di produzione. A causa della sua urgenza, un hotfix viene in genere creato al di fuori del tipico DevOps flusso di lavoro di rilascio.

periodo di hypercare

Subito dopo la conversione, il periodo di tempo in cui un team di migrazione gestisce e monitora le applicazioni migrate nel cloud per risolvere eventuali problemi. In genere, questo periodo dura

da 1 a 4 giorni. Al termine del periodo di hypercare, il team addetto alla migrazione in genere trasferisce la responsabilità delle applicazioni al team addetto alle operazioni cloud.

## I

IaC

Considera l'infrastruttura come codice.

Policy basata su identità

Una policy associata a uno o più principi IAM che definisce le relative autorizzazioni all'interno dell' Cloud AWS ambiente.

applicazione inattiva

Un'applicazione che prevede un uso di CPU e memoria medio compreso tra il 5% e il 20% in un periodo di 90 giorni. In un progetto di migrazione, è normale ritirare queste applicazioni o mantenerle on-premise.

IIoT

Vedi Industrial Internet of Things.

infrastruttura immutabile

Un modello che implementa una nuova infrastruttura per i carichi di lavoro di produzione anziché aggiornare, applicare patch o modificare l'infrastruttura esistente. Le infrastrutture immutabili sono intrinsecamente più coerenti, affidabili e prevedibili delle infrastrutture mutabili. Per ulteriori informazioni, consulta la best practice Deploy using immutable infrastructure in Well-Architected AWS Framework.

VPC in ingresso (ingress)

In un'architettura AWS multi-account, un VPC che accetta, ispeziona e indirizza le connessioni di rete dall'esterno di un'applicazione. La AWS Security Reference Architecture consiglia di configurare l'account di rete con funzionalità in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e Internet in generale.

migrazione incrementale

Una strategia di conversione in cui si esegue la migrazione dell'applicazione in piccole parti anziché eseguire una conversione singola e completa. Ad esempio, inizialmente potresti spostare

solo alcuni microservizi o utenti nel nuovo sistema. Dopo aver verificato che tutto funzioni correttamente, puoi spostare in modo incrementale microservizi o utenti aggiuntivi fino alla disattivazione del sistema legacy. Questa strategia riduce i rischi associati alle migrazioni di grandi dimensioni.

Industria 4.0

Un termine introdotto da Klaus Schwab nel 2016 per riferirsi alla modernizzazione dei processi di produzione attraverso progressi in termini di connettività, dati in tempo reale, automazione, analisi e AI/ML.

infrastruttura

Tutte le risorse e gli asset contenuti nell'ambiente di un'applicazione.

infrastruttura come codice (IaC)

Il processo di provisioning e gestione dell'infrastruttura di un'applicazione tramite un insieme di file di configurazione. Il processo IaC è progettato per aiutarti a centralizzare la gestione dell'infrastruttura, a standardizzare le risorse e a dimensionare rapidamente, in modo che i nuovi ambienti siano ripetibili, affidabili e coerenti.

IIoInternet delle cose industriale (T)

L'uso di sensori e dispositivi connessi a Internet nei settori industriali, come quello manifatturiero, energetico, automobilistico, sanitario, delle scienze della vita e dell'agricoltura. Per ulteriori informazioni, vedere Creazione di una strategia di trasformazione digitale per l'Internet of Things (IIoT) industriale.

VPC di ispezione

In un'architettura AWS multi-account, un VPC centralizzato che gestisce le ispezioni del traffico di rete tra VPCs (nello stesso o in modo diverso Regioni AWS), Internet e le reti locali. La AWS Security Reference Architecture consiglia di configurare l'account di rete con informazioni in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e Internet in generale.

Internet of Things (IoT)

La rete di oggetti fisici connessi con sensori o processori incorporati che comunicano con altri dispositivi e sistemi tramite Internet o una rete di comunicazione locale. Per ulteriori informazioni, consulta Cos'è l'IoT?

interpretabilità

Una caratteristica di un modello di machine learning che descrive il grado in cui un essere umano è in grado di comprendere in che modo le previsioni del modello dipendono dai suoi input. Per ulteriori informazioni, vedere Interpretabilità del modello di machine learning con. AWS

IoT

Vedi Internet of Things.

libreria di informazioni IT (ITIL)

Una serie di best practice per offrire servizi IT e allinearli ai requisiti aziendali. ITIL fornisce le basi per ITSM.

gestione dei servizi IT (ITSM)

Attività associate alla progettazione, implementazione, gestione e supporto dei servizi IT per un'organizzazione. Per informazioni sull'integrazione delle operazioni cloud con gli strumenti ITSM, consulta la guida all'integrazione delle operazioni.

ITIL

Vedi la libreria di informazioni IT.

ITSM

Vedi Gestione dei servizi IT.

# L

controllo degli accessi basato su etichette (LBAC)

Un'implementazione del controllo di accesso obbligatorio (MAC) in cui agli utenti e ai dati stessi viene assegnato esplicitamente un valore di etichetta di sicurezza. L'intersezione tra l'etichetta di sicurezza utente e l'etichetta di sicurezza dei dati determina quali righe e colonne possono essere visualizzate dall'utente.

zona di destinazione

Una landing zone è un AWS ambiente multi-account ben progettato, scalabile e sicuro. Questo è un punto di partenza dal quale le organizzazioni possono avviare e distribuire rapidamente carichi di lavoro e applicazioni con fiducia nel loro ambiente di sicurezza e infrastruttura. Per ulteriori

informazioni sulle zone di destinazione, consulta la sezione Configurazione di un ambiente AWS multi-account sicuro e scalabile.

modello linguistico di grandi dimensioni (LLM)

Un modello di intelligenza artificiale di deep learning preaddestrato su una grande quantità di dati. Un LLM può svolgere più attività, come rispondere a domande, riepilogare documenti, tradurre testo in altre lingue e completare frasi. Per ulteriori informazioni, consulta Cosa sono. LLMs

migrazione su larga scala

Una migrazione di 300 o più server.

BIANCO

Vedi controllo degli accessi basato su etichette.

Privilegio minimo

La best practice di sicurezza per la concessione delle autorizzazioni minime richieste per eseguire un'attività. Per ulteriori informazioni, consulta Applicazione delle autorizzazioni del privilegio minimo nella documentazione di IAM.

eseguire il rehosting (lift and shift)

Vedi 7 R.

sistema little-endian

Un sistema che memorizza per primo il byte meno importante. Vedi anche endianità.

LLM

Vedi modello linguistico di grandi dimensioni.

ambienti inferiori

Vedi ambiente.

# M

machine learning (ML)

Un tipo di intelligenza artificiale che utilizza algoritmi e tecniche per il riconoscimento e l'apprendimento di schemi. Il machine learning analizza e apprende dai dati registrati, come i dati

dell'Internet delle cose (IoT), per generare un modello statistico basato su modelli. Per ulteriori informazioni, consulta la sezione Machine learning.

ramo principale

Vedi filiale.

malware

Software progettato per compromettere la sicurezza o la privacy del computer. Il malware potrebbe interrompere i sistemi informatici, divulgare informazioni sensibili o ottenere accessi non autorizzati. Esempi di malware includono virus, worm, ransomware, trojan horse, spyware e keylogger.

servizi gestiti

Servizi AWS per cui AWS gestisce il livello di infrastruttura, il sistema operativo e le piattaforme e si accede agli endpoint per archiviare e recuperare i dati. Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3) e Amazon DynamoDB sono esempi di servizi gestiti. Questi sono noti anche come servizi astratti.

sistema di esecuzione della produzione (MES)

Un sistema software per tracciare, monitorare, documentare e controllare i processi di produzione che convertono le materie prime in prodotti finiti in officina.

MAP

Vedi Migration Acceleration Program.

meccanismo

Un processo completo in cui si crea uno strumento, si promuove l'adozione dello strumento e quindi si esaminano i risultati per apportare le modifiche. Un meccanismo è un ciclo che si rafforza e si migliora man mano che funziona. Per ulteriori informazioni, consulta Creazione di meccanismi nel AWS Well-Architected Framework.

account membro

Tutti gli account Account AWS diversi dall'account di gestione che fanno parte di un'organizzazione in. AWS Organizations Un account può essere membro di una sola organizzazione alla volta.

MEH

Vedi sistema di esecuzione della produzione.

Message Queuing Telemetry Transport (MQTT)

Un protocollo di comunicazione machine-to-machine (M2M) leggero, basato sul modello di pubblicazione/sottoscrizione, per dispositivi IoT con risorse limitate.

microservizio

Un servizio piccolo e indipendente che comunica tramite canali ben definiti ed è in genere di proprietà di piccoli team autonomi. APIs Ad esempio, un sistema assicurativo potrebbe includere microservizi che si riferiscono a funzionalità aziendali, come vendite o marketing, o sottodomini, come acquisti, reclami o analisi. I vantaggi dei microservizi includono agilità, dimensionamento flessibile, facilità di implementazione, codice riutilizzabile e resilienza. Per ulteriori informazioni, consulta Integrazione dei microservizi utilizzando servizi serverless. AWS

architettura di microservizi

Un approccio alla creazione di un'applicazione con componenti indipendenti che eseguono ogni processo applicativo come microservizio. Questi microservizi comunicano attraverso un'interfaccia ben definita utilizzando sistemi leggeri. APIs Ogni microservizio in questa architettura può essere aggiornato, distribuito e dimensionato per soddisfare la richiesta di funzioni specifiche di un'applicazione. Per ulteriori informazioni, vedere Implementazione dei microservizi su. AWS

Programma di accelerazione della migrazione (MAP)

Un AWS programma che fornisce consulenza, supporto, formazione e servizi per aiutare le organizzazioni a costruire una solida base operativa per il passaggio al cloud e per contribuire a compensare il costo iniziale delle migrazioni. MAP include una metodologia di migrazione per eseguire le migrazioni precedenti in modo metodico e un set di strumenti per automatizzare e accelerare gli scenari di migrazione comuni.

migrazione su larga scala

Il processo di trasferimento della maggior parte del portfolio di applicazioni sul cloud avviene a ondate, con più applicazioni trasferite a una velocità maggiore in ogni ondata. Questa fase utilizza le migliori pratiche e le lezioni apprese nelle fasi precedenti per implementare una fabbrica di migrazione di team, strumenti e processi per semplificare la migrazione dei carichi di lavoro attraverso l'automazione e la distribuzione agile. Questa è la terza fase della strategia di migrazione AWS.

fabbrica di migrazione

Team interfunzionali che semplificano la migrazione dei carichi di lavoro attraverso approcci automatizzati e agili. I team di Migration Factory in genere includono addetti alle operazioni,

analisti e proprietari aziendali, ingegneri addetti alla migrazione, sviluppatori e DevOps professionisti che lavorano nell'ambito degli sprint. Tra il 20% e il 50% di un portfolio di applicazioni aziendali è costituito da schemi ripetuti che possono essere ottimizzati con un approccio di fabbrica. Per ulteriori informazioni, consulta la discussione sulle fabbriche di migrazione e la Guida alla fabbrica di migrazione al cloud in questo set di contenuti.

metadati di migrazione

Le informazioni sull'applicazione e sul server necessarie per completare la migrazione. Ogni modello di migrazione richiede un set diverso di metadati di migrazione. Esempi di metadati di migrazione includono la sottorete, il gruppo di sicurezza e l'account di destinazione. AWS

modello di migrazione

Un'attività di migrazione ripetibile che descrive in dettaglio la strategia di migrazione, la destinazione della migrazione e l'applicazione o il servizio di migrazione utilizzati. Esempio: riorganizza la migrazione su Amazon EC2 con AWS Application Migration Service.

Valutazione del portfolio di migrazione (MPA)

Uno strumento online che fornisce informazioni per la convalida del business case per la migrazione a. Cloud AWS MPA offre una valutazione dettagliata del portfolio (dimensionamento corretto dei server, prezzi, confronto del TCO, analisi dei costi di migrazione) e pianificazione della migrazione (analisi e raccolta dei dati delle applicazioni, raggruppamento delle applicazioni, prioritizzazione delle migrazioni e pianificazione delle ondate). Lo strumento MPA (richiede l'accesso) è disponibile gratuitamente per tutti i AWS consulenti e i consulenti dei partner APN.

valutazione della preparazione alla migrazione (MRA)

Il processo di acquisizione di informazioni sullo stato di preparazione al cloud di un'organizzazione, l'identificazione dei punti di forza e di debolezza e la creazione di un piano d'azione per colmare le lacune identificate, utilizzando il CAF. AWS Per ulteriori informazioni, consulta la guida di preparazione alla migrazione. MRA è la prima fase della strategia di migrazione AWS.

strategia di migrazione

L'approccio utilizzato per migrare un carico di lavoro verso. Cloud AWS Per ulteriori informazioni, consulta la voce 7 R in questo glossario e consulta Mobilita la tua organizzazione per accelerare le migrazioni su larga scala.

ML

Vedi machine learning.

modernizzazione

Trasformazione di un'applicazione obsoleta (legacy o monolitica) e della relativa infrastruttura in un sistema agile, elastico e altamente disponibile nel cloud per ridurre i costi, aumentare l'efficienza e sfruttare le innovazioni. Per ulteriori informazioni, vedere Strategia per la modernizzazione delle applicazioni in. Cloud AWS

valutazione della preparazione alla modernizzazione

Una valutazione che aiuta a determinare la preparazione alla modernizzazione delle applicazioni di un'organizzazione, identifica vantaggi, rischi e dipendenze e determina in che misura l'organizzazione può supportare lo stato futuro di tali applicazioni. Il risultato della valutazione è uno schema dell'architettura di destinazione, una tabella di marcia che descrive in dettaglio le fasi di sviluppo e le tappe fondamentali del processo di modernizzazione e un piano d'azione per colmare le lacune identificate. Per ulteriori informazioni, vedere Valutazione della preparazione alla modernizzazione per le applicazioni in. Cloud AWS

applicazioni monolitiche (monoliti)

Applicazioni eseguite come un unico servizio con processi strettamente collegati. Le applicazioni monolitiche presentano diversi inconvenienti. Se una funzionalità dell'applicazione registra un picco di domanda, l'intera architettura deve essere dimensionata. L'aggiunta o il miglioramento delle funzionalità di un'applicazione monolitica diventa inoltre più complessa man mano che la base di codice cresce. Per risolvere questi problemi, puoi utilizzare un'architettura di microservizi. Per ulteriori informazioni, consulta la sezione Scomposizione dei monoliti in microservizi.

MAPPA

Vedi Migration Portfolio Assessment.

MQTT

Vedi Message Queuing Telemetry Transport.

classificazione multiclasse

Un processo che aiuta a generare previsioni per più classi (prevedendo uno o più di due risultati). Ad esempio, un modello di machine learning potrebbe chiedere "Questo prodotto è un libro, un'auto o un telefono?" oppure "Quale categoria di prodotti è più interessante per questo cliente?"

infrastruttura mutabile

Un modello che aggiorna e modifica l'infrastruttura esistente per i carichi di lavoro di produzione. Per migliorare la coerenza, l'affidabilità e la prevedibilità, il AWS Well-Architected Framework consiglia l'uso di un'infrastruttura immutabile come best practice.

# O

OAC

Vedi Origin Access Control.

QUERCIA

Vedi Origin Access Identity.

OCM

Vedi gestione delle modifiche organizzative.

migrazione offline

Un metodo di migrazione in cui il carico di lavoro di origine viene eliminato durante il processo di migrazione. Questo metodo prevede tempi di inattività prolungati e viene in genere utilizzato per carichi di lavoro piccoli e non critici.

OI

Vedi l'integrazione delle operazioni.

OLA

Vedi accordo a livello operativo.

migrazione online

Un metodo di migrazione in cui il carico di lavoro di origine viene copiato sul sistema di destinazione senza essere messo offline. Le applicazioni connesse al carico di lavoro possono continuare a funzionare durante la migrazione. Questo metodo comporta tempi di inattività pari a zero o comunque minimi e viene in genere utilizzato per carichi di lavoro di produzione critici.

OPC-UA

Vedi Open Process Communications - Unified Architecture.

Comunicazioni a processo aperto - Architettura unificata (OPC-UA)

Un protocollo di comunicazione machine-to-machine (M2M) per l'automazione industriale. OPC-UA fornisce uno standard di interoperabilità con schemi di crittografia, autenticazione e autorizzazione dei dati.

accordo a livello operativo (OLA)

Un accordo che chiarisce quali sono gli impegni reciproci tra i gruppi IT funzionali, a supporto di un accordo sul livello di servizio (SLA).

revisione della prontezza operativa (ORR)

Un elenco di domande e best practice associate che aiutano a comprendere, valutare, prevenire o ridurre la portata degli incidenti e dei possibili guasti. Per ulteriori informazioni, vedere Operational Readiness Reviews (ORR) nel Well-Architected AWS Framework.

tecnologia operativa (OT)

Sistemi hardware e software che interagiscono con l'ambiente fisico per controllare le operazioni, le apparecchiature e le infrastrutture industriali. Nella produzione, l'integrazione di sistemi OT e di tecnologia dell'informazione (IT) è un obiettivo chiave per le trasformazioni dell'Industria 4.0.

integrazione delle operazioni (OI)

Il processo di modernizzazione delle operazioni nel cloud, che prevede la pianificazione, l'automazione e l'integrazione della disponibilità. Per ulteriori informazioni, consulta la guida all'integrazione delle operazioni.

trail organizzativo

Un percorso creato da noi AWS CloudTrail che registra tutti gli eventi di un'organizzazione per tutti Account AWS . AWS Organizations Questo percorso viene creato in ogni Account AWS che fa parte dell'organizzazione e tiene traccia dell'attività in ogni account. Per ulteriori informazioni, consulta Creazione di un percorso per un'organizzazione nella CloudTrail documentazione.

gestione del cambiamento organizzativo (OCM)

Un framework per la gestione di trasformazioni aziendali importanti e che comportano l'interruzione delle attività dal punto di vista delle persone, della cultura e della leadership. OCM aiuta le organizzazioni a prepararsi e passare a nuovi sistemi e strategie accelerando l'adozione del cambiamento, affrontando i problemi di transizione e promuovendo cambiamenti culturali e organizzativi. Nella strategia di AWS migrazione, questo framework si chiama accelerazione delle

persone, a causa della velocità di cambiamento richiesta nei progetti di adozione del cloud. Per ulteriori informazioni, consultare la Guida OCM.

controllo dell'accesso all'origine (OAC)

In CloudFront, un'opzione avanzata per limitare l'accesso per proteggere i contenuti di Amazon Simple Storage Service (Amazon S3). OAC supporta tutti i bucket S3 in generale Regioni AWS, la crittografia lato server con AWS KMS (SSE-KMS) e le richieste dinamiche e dirette al bucket S3. `PUT` `DELETE`

identità di accesso origine (OAI)

Nel CloudFront, un'opzione per limitare l'accesso per proteggere i tuoi contenuti Amazon S3. Quando usi OAI, CloudFront crea un principale con cui Amazon S3 può autenticarsi. I principali autenticati possono accedere ai contenuti in un bucket S3 solo tramite una distribuzione specifica. CloudFront Vedi anche OAC, che fornisce un controllo degli accessi più granulare e avanzato.

ORR

Vedi la revisione della prontezza operativa.

- NON

Vedi la tecnologia operativa.

VPC in uscita (egress)

In un'architettura AWS multi-account, un VPC che gestisce le connessioni di rete avviate dall'interno di un'applicazione. La AWS Security Reference Architecture consiglia di configurare l'account di rete con funzionalità in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e Internet in generale.

# P

limite delle autorizzazioni

Una policy di gestione IAM collegata ai principali IAM per impostare le autorizzazioni massime che l'utente o il ruolo possono avere. Per ulteriori informazioni, consulta Limiti delle autorizzazioni nella documentazione di IAM.

informazioni di identificazione personale (PII)

Informazioni che, se visualizzate direttamente o abbinate ad altri dati correlati, possono essere utilizzate per dedurre ragionevolmente l'identità di un individuo. Esempi di informazioni personali includono nomi, indirizzi e informazioni di contatto.

Informazioni che consentono l'identificazione personale degli utenti

Visualizza le informazioni di identificazione personale.

playbook

Una serie di passaggi predefiniti che raccolgono il lavoro associato alle migrazioni, come l'erogazione delle funzioni operative principali nel cloud. Un playbook può assumere la forma di script, runbook automatici o un riepilogo dei processi o dei passaggi necessari per gestire un ambiente modernizzato.

PLC

Vedi controllore logico programmabile.

PLM

Vedi la gestione del ciclo di vita del prodotto.

policy

Un oggetto in grado di definire le autorizzazioni (vedi politica basata sull'identità), specificare le condizioni di accesso (vedi politicabasata sulle risorse) o definire le autorizzazioni massime per tutti gli account di un'organizzazione in (vedi politica di controllo dei servizi). AWS Organizations

persistenza poliglotta

Scelta indipendente della tecnologia di archiviazione di dati di un microservizio in base ai modelli di accesso ai dati e ad altri requisiti. Se i microservizi utilizzano la stessa tecnologia di archiviazione di dati, possono incontrare problemi di implementazione o registrare prestazioni scadenti. I microservizi vengono implementati più facilmente e ottengono prestazioni e scalabilità migliori se utilizzano l'archivio dati più adatto alle loro esigenze. Per ulteriori informazioni, consulta la sezione Abilitazione della persistenza dei dati nei microservizi.

valutazione del portfolio

Un processo di scoperta, analisi e definizione delle priorità del portfolio di applicazioni per pianificare la migrazione. Per ulteriori informazioni, consulta la pagina Valutazione della preparazione alla migrazione.

predicate

Una condizione di interrogazione che restituisce o, in genere, si trova in una clausola`WHERE`. `true` `false`

predicato pushdown

Una tecnica di ottimizzazione delle query del database che filtra i dati della query prima del trasferimento. Ciò riduce la quantità di dati che devono essere recuperati ed elaborati dal database relazionale e migliora le prestazioni delle query.

controllo preventivo

Un controllo di sicurezza progettato per impedire il verificarsi di un evento. Questi controlli sono la prima linea di difesa per impedire accessi non autorizzati o modifiche indesiderate alla rete. Per ulteriori informazioni, consulta Controlli preventivi in Implementazione dei controlli di sicurezza in AWS.

principale

Un'entità in AWS grado di eseguire azioni e accedere alle risorse. Questa entità è in genere un utente root per un Account AWS ruolo IAM o un utente. Per ulteriori informazioni, consulta Principali in Termini e concetti dei ruoli nella documentazione di IAM.

privacy fin dalla progettazione

Un approccio di ingegneria dei sistemi che tiene conto della privacy durante l'intero processo di sviluppo.

zone ospitate private

Un contenitore che contiene informazioni su come desideri che Amazon Route 53 risponda alle query DNS per un dominio e i relativi sottodomini all'interno di uno o più. VPCs Per ulteriori informazioni, consulta Utilizzo delle zone ospitate private nella documentazione di Route 53.

controllo proattivo

Un controllo di sicurezza progettato per impedire l'implementazione di risorse non conformi. Questi controlli analizzano le risorse prima del loro provisioning. Se la risorsa non è conforme al controllo, non viene fornita. Per ulteriori informazioni, consulta la guida di riferimento sui controlli nella AWS Control Tower documentazione e consulta Controlli proattivi in Implementazione dei controlli di sicurezza su. AWS

gestione del ciclo di vita del prodotto (PLM)

La gestione dei dati e dei processi di un prodotto durante l'intero ciclo di vita, dalla progettazione, sviluppo e lancio, attraverso la crescita e la maturità, fino al declino e alla rimozione.

Ambiente di produzione

Vedi ambiente.

controllore logico programmabile (PLC)

Nella produzione, un computer altamente affidabile e adattabile che monitora le macchine e automatizza i processi di produzione.

concatenamento rapido

Utilizzo dell'output di un prompt LLM come input per il prompt successivo per generare risposte migliori. Questa tecnica viene utilizzata per suddividere un'attività complessa in sottoattività o per perfezionare o espandere iterativamente una risposta preliminare. Aiuta a migliorare l'accuratezza e la pertinenza delle risposte di un modello e consente risultati più granulari e personalizzati.

pseudonimizzazione

Il processo di sostituzione degli identificatori personali in un set di dati con valori segnaposto. La pseudonimizzazione può aiutare a proteggere la privacy personale. I dati pseudonimizzati sono ancora considerati dati personali.

publish/subscribe (pub/sub)

Un modello che consente comunicazioni asincrone tra microservizi per migliorare la scalabilità e la reattività. Ad esempio, in un MES basato su microservizi, un microservizio può pubblicare messaggi di eventi su un canale a cui altri microservizi possono abbonarsi. Il sistema può aggiungere nuovi microservizi senza modificare il servizio di pubblicazione.

# Q

Piano di query

Una serie di passaggi, come le istruzioni, utilizzati per accedere ai dati in un sistema di database relazionale SQL.

regressione del piano di query

Quando un ottimizzatore del servizio di database sceglie un piano non ottimale rispetto a prima di una determinata modifica all'ambiente di database. Questo può essere causato da modifiche a statistiche, vincoli, impostazioni dell'ambiente, associazioni dei parametri di query e aggiornamenti al motore di database.

# R

Matrice RACI

Vedi responsabile, responsabile, consultato, informato (RACI).

STRACCIO

Vedi Retrieval Augmented Generation.

ransomware

Un software dannoso progettato per bloccare l'accesso a un sistema informatico o ai dati fino a quando non viene effettuato un pagamento.

Matrice RASCI

Vedi responsabile, responsabile, consultato, informato (RACI).

RCAC

Vedi controllo dell'accesso a righe e colonne.

replica di lettura

Una copia di un database utilizzata per scopi di sola lettura. È possibile indirizzare le query alla replica di lettura per ridurre il carico sul database principale.

riprogettare

Vedi 7 Rs.

obiettivo del punto di ripristino (RPO)

Il periodo di tempo massimo accettabile dall'ultimo punto di ripristino dei dati. Questo determina ciò che si considera una perdita di dati accettabile tra l'ultimo punto di ripristino e l'interruzione del servizio.

obiettivo del tempo di ripristino (RTO)

Il ritardo massimo accettabile tra l'interruzione del servizio e il ripristino del servizio.

rifattorizzare

Vedi 7 R.

Regione

Una raccolta di AWS risorse in un'area geografica. Ciascuna Regione AWS è isolata e indipendente dalle altre per fornire tolleranza agli errori, stabilità e resilienza. Per ulteriori informazioni, consulta Specificare cosa può usare Regioni AWS il tuo account.

regressione

Una tecnica di ML che prevede un valore numerico. Ad esempio, per risolvere il problema "A che prezzo verrà venduta questa casa?" un modello di ML potrebbe utilizzare un modello di regressione lineare per prevedere il prezzo di vendita di una casa sulla base di dati noti sulla casa (ad esempio, la metratura).

riospitare

Vedi 7 R.

rilascio

In un processo di implementazione, l'atto di promuovere modifiche a un ambiente di produzione.

trasferisco

Vedi 7 Rs.

ripiattaforma

Vedi 7 Rs.

riacquisto

Vedi 7 Rs.

resilienza

La capacità di un'applicazione di resistere o ripristinare le interruzioni. L'elevata disponibilità e il disaster recovery sono considerazioni comuni quando si pianifica la resilienza in. Cloud AWSPer ulteriori informazioni, vedere Cloud AWS Resilience.

policy basata su risorse

Una policy associata a una risorsa, ad esempio un bucket Amazon S3, un endpoint o una chiave di crittografia. Questo tipo di policy specifica a quali principali è consentito l'accesso, le azioni supportate e qualsiasi altra condizione che deve essere soddisfatta.

matrice di assegnazione di responsabilità (RACI)

Una matrice che definisce i ruoli e le responsabilità di tutte le parti coinvolte nelle attività di migrazione e nelle operazioni cloud. Il nome della matrice deriva dai tipi di responsabilità definiti nella matrice: responsabile (R), responsabile (A), consultato (C) e informato (I). Il tipo di supporto (S) è facoltativo. Se includi il supporto, la matrice viene chiamata matrice RASCI e, se la escludi, viene chiamata matrice RACI.

controllo reattivo

Un controllo di sicurezza progettato per favorire la correzione di eventi avversi o deviazioni dalla baseline di sicurezza. Per ulteriori informazioni, consulta Controlli reattivi in Implementazione dei controlli di sicurezza in AWS.

retain

Vedi 7 R.

andare in pensione

Vedi 7 Rs.

Retrieval Augmented Generation (RAG)

Una tecnologia di intelligenza artificiale generativa in cui un LLM fa riferimento a una fonte di dati autorevole esterna alle sue fonti di dati di formazione prima di generare una risposta. Ad esempio, un modello RAG potrebbe eseguire una ricerca semantica nella knowledge base o nei dati personalizzati di un'organizzazione. Per ulteriori informazioni, consulta Cos'è il RAG.

rotazione

Processo di aggiornamento periodico di un segreto per rendere più difficile l'accesso alle credenziali da parte di un utente malintenzionato.

controllo dell'accesso a righe e colonne (RCAC)

L'uso di espressioni SQL di base e flessibili con regole di accesso definite. RCAC è costituito da autorizzazioni di riga e maschere di colonna.

RPO

Vedi l'obiettivo del punto di ripristino.

RTO

Vedi l'obiettivo del tempo di ripristino.

runbook

Un insieme di procedure manuali o automatizzate necessarie per eseguire un'attività specifica. In genere sono progettati per semplificare operazioni o procedure ripetitive con tassi di errore elevati.

# S

SAML 2.0

Uno standard aperto utilizzato da molti provider di identità (IdPs). Questa funzionalità abilita il single sign-on (SSO) federato, in modo che gli utenti possano accedere AWS Management Console o chiamare le operazioni AWS API senza che tu debba creare un utente in IAM per tutti i membri dell'organizzazione. Per ulteriori informazioni sulla federazione basata su SAML 2.0, consulta Informazioni sulla federazione basata su SAML 2.0 nella documentazione di IAM.

SCADA

Vedi controllo di supervisione e acquisizione dati.

SCP

Vedi la politica di controllo del servizio.

Secret

In AWS Secrets Manager, informazioni riservate o riservate, come una password o le credenziali utente, archiviate in forma crittografata. È costituito dal valore segreto e dai relativi metadati. Il valore segreto può essere binario, una stringa singola o più stringhe. Per ulteriori informazioni, consulta Cosa c'è in un segreto di Secrets Manager? nella documentazione di Secrets Manager.

sicurezza fin dalla progettazione

Un approccio di ingegneria dei sistemi che tiene conto della sicurezza durante l'intero processo di sviluppo.

controllo di sicurezza

Un guardrail tecnico o amministrativo che impedisce, rileva o riduce la capacità di un autore di minacce di sfruttare una vulnerabilità di sicurezza. Esistono quattro tipi principali di controlli di sicurezza: preventivi, investigativi, reattivi e proattivi.

rafforzamento della sicurezza

Il processo di riduzione della superficie di attacco per renderla più resistente agli attacchi. Può includere azioni come la rimozione di risorse che non sono più necessarie, l'implementazione di best practice di sicurezza che prevedono la concessione del privilegio minimo o la disattivazione di funzionalità non necessarie nei file di configurazione.

sistema di gestione delle informazioni e degli eventi di sicurezza (SIEM)

Strumenti e servizi che combinano sistemi di gestione delle informazioni di sicurezza (SIM) e sistemi di gestione degli eventi di sicurezza (SEM). Un sistema SIEM raccoglie, monitora e analizza i dati da server, reti, dispositivi e altre fonti per rilevare minacce e violazioni della sicurezza e generare avvisi.

automazione della risposta alla sicurezza

Un'azione predefinita e programmata progettata per rispondere o porre rimedio automaticamente a un evento di sicurezza. Queste automazioni fungono da controlli di sicurezza investigativi o reattivi che aiutano a implementare le migliori pratiche di sicurezza. AWS Esempi di azioni di risposta automatizzate includono la modifica di un gruppo di sicurezza VPC, l'applicazione di patch a un'istanza EC2 Amazon o la rotazione delle credenziali.

Crittografia lato server

Crittografia dei dati a destinazione, da parte di chi li riceve. Servizio AWS

Policy di controllo dei servizi (SCP)

Una politica che fornisce il controllo centralizzato sulle autorizzazioni per tutti gli account di un'organizzazione in. AWS Organizations SCPs definire barriere o fissare limiti alle azioni che un amministratore può delegare a utenti o ruoli. È possibile utilizzarli SCPs come elenchi consentiti o elenchi di rifiuto, per specificare quali servizi o azioni sono consentiti o proibiti. Per ulteriori informazioni, consulta le politiche di controllo del servizio nella AWS Organizations documentazione.

endpoint del servizio

L'URL del punto di ingresso per un Servizio AWS. Puoi utilizzare l'endpoint per connetterti a livello di programmazione al servizio di destinazione. Per ulteriori informazioni, consulta Endpoint del Servizio AWS nei Riferimenti generali di AWS.

accordo sul livello di servizio (SLA)

Un accordo che chiarisce ciò che un team IT promette di offrire ai propri clienti, ad esempio l'operatività e le prestazioni del servizio.

indicatore del livello di servizio (SLI)

Misurazione di un aspetto prestazionale di un servizio, ad esempio il tasso di errore, la disponibilità o la velocità effettiva.

obiettivo a livello di servizio (SLO)

Una metrica target che rappresenta lo stato di un servizio, misurato da un indicatore del livello di servizio.

Modello di responsabilità condivisa

Un modello che descrive la responsabilità condivisa AWS per la sicurezza e la conformità del cloud. AWS è responsabile della sicurezza del cloud, mentre tu sei responsabile della sicurezza nel cloud. Per ulteriori informazioni, consulta Modello di responsabilità condivisa.

SIEM

Vedi il sistema di gestione delle informazioni e degli eventi sulla sicurezza.

punto di errore singolo (SPOF)

Un guasto in un singolo componente critico di un'applicazione che può disturbare il sistema.

SLAM

Vedi il contratto sul livello di servizio.

SLI

Vedi l'indicatore del livello di servizio.

LENTA

Vedi obiettivo del livello di servizio.

split-and-seed modello

Un modello per dimensionare e accelerare i progetti di modernizzazione. Man mano che vengono definite nuove funzionalità e versioni dei prodotti, il team principale si divide per creare nuovi team di prodotto. Questo aiuta a dimensionare le capacità e i servizi dell'organizzazione, migliora la produttività degli sviluppatori e supporta una rapida innovazione. Per ulteriori informazioni, vedere Approccio graduale alla modernizzazione delle applicazioni in. Cloud AWS

SPOF

Vedi punto di errore singolo.

schema a stella

Una struttura organizzativa di database che utilizza un'unica tabella dei fatti di grandi dimensioni per archiviare i dati transazionali o misurati e utilizza una o più tabelle dimensionali più piccole per memorizzare gli attributi dei dati. Questa struttura è progettata per l'uso in un data warehouse o per scopi di business intelligence.

modello del fico strangolatore

Un approccio alla modernizzazione dei sistemi monolitici mediante la riscrittura e la sostituzione incrementali delle funzionalità del sistema fino alla disattivazione del sistema legacy. Questo modello utilizza l'analogia di una pianta di fico che cresce fino a diventare un albero robusto e alla fine annienta e sostituisce il suo ospite. Il modello è stato introdotto da Martin Fowler come metodo per gestire il rischio durante la riscrittura di sistemi monolitici. Per un esempio di come applicare questo modello, consulta Modernizzazione incrementale dei servizi Web legacy di Microsoft ASP.NET (ASMX) mediante container e Gateway Amazon API.

sottorete

Un intervallo di indirizzi IP nel VPC. Una sottorete deve risiedere in una singola zona di disponibilità.

controllo di supervisione e acquisizione dati (SCADA)

Nella produzione, un sistema che utilizza hardware e software per monitorare gli asset fisici e le operazioni di produzione.

crittografia simmetrica

Un algoritmo di crittografia che utilizza la stessa chiave per crittografare e decrittografare i dati.

test sintetici

Test di un sistema in modo da simulare le interazioni degli utenti per rilevare potenziali problemi o monitorare le prestazioni. Puoi usare Amazon CloudWatch Synthetics per creare questi test.

prompt di sistema

Una tecnica per fornire contesto, istruzioni o linee guida a un LLM per indirizzarne il comportamento. I prompt di sistema aiutano a impostare il contesto e stabilire regole per le interazioni con gli utenti.

# T

tags

Coppie chiave-valore che fungono da metadati per l'organizzazione delle risorse. AWS I tag possono aiutarti a gestire, identificare, organizzare, cercare e filtrare le risorse. Per ulteriori informazioni, consulta l'articolo relativo all'assegnazione di tag alle risorse AWS.

variabile di destinazione

Il valore che stai cercando di prevedere nel machine learning supervisionato. Questo è indicato anche come variabile di risultato. Ad esempio, in un ambiente di produzione la variabile di destinazione potrebbe essere un difetto del prodotto.

elenco di attività

Uno strumento che viene utilizzato per tenere traccia dei progressi tramite un runbook. Un elenco di attività contiene una panoramica del runbook e un elenco di attività generali da completare. Per ogni attività generale, include la quantità stimata di tempo richiesta, il proprietario e lo stato di avanzamento.

Ambiente di test

Vedi ambiente.

training

Fornire dati da cui trarre ispirazione dal modello di machine learning. I dati di training devono contenere la risposta corretta. L'algoritmo di apprendimento trova nei dati di addestramento i pattern che mappano gli attributi dei dati di input al target (la risposta che si desidera prevedere). Produce un modello di ML che acquisisce questi modelli. Puoi quindi utilizzare il modello di ML per creare previsioni su nuovi dati di cui non si conosce il target.

Transit Gateway

Un hub di transito di rete che puoi utilizzare per interconnettere le tue reti VPCs e quelle locali. Per ulteriori informazioni, consulta Cos'è un gateway di transito nella AWS Transit Gateway documentazione.

flusso di lavoro basato su trunk

Un approccio in cui gli sviluppatori creano e testano le funzionalità localmente in un ramo di funzionalità e quindi uniscono tali modifiche al ramo principale. Il ramo principale viene quindi integrato negli ambienti di sviluppo, preproduzione e produzione, in sequenza.

Accesso attendibile

Concessione delle autorizzazioni a un servizio specificato dall'utente per eseguire attività all'interno dell'organizzazione AWS Organizations e nei suoi account per conto dell'utente. Il servizio attendibile crea un ruolo collegato al servizio in ogni account, quando tale ruolo è necessario, per eseguire attività di gestione per conto dell'utente. Per ulteriori informazioni, consulta Utilizzo AWS Organizations con altri AWS servizi nella AWS Organizations documentazione.

regolazione

Modificare alcuni aspetti del processo di training per migliorare la precisione del modello di ML. Ad esempio, puoi addestrare il modello di ML generando un set di etichette, aggiungendo etichette e quindi ripetendo questi passaggi più volte con impostazioni diverse per ottimizzare il modello.

team da due pizze

Una piccola DevOps squadra che puoi sfamare con due pizze. Un team composto da due persone garantisce la migliore opportunità possibile di collaborazione nello sviluppo del software.

# U

incertezza

Un concetto che si riferisce a informazioni imprecise, incomplete o sconosciute che possono minare l'affidabilità dei modelli di machine learning predittivi. Esistono due tipi di incertezza: l'incertezza epistemica, che è causata da dati limitati e incompleti, mentre l'incertezza aleatoria è causata dal rumore e dalla casualità insiti nei dati. Per ulteriori informazioni, consulta la guida Quantificazione dell'incertezza nei sistemi di deep learning.

compiti indifferenziati

Conosciuto anche come sollevamento di carichi pesanti, è un lavoro necessario per creare e far funzionare un'applicazione, ma che non apporta valore diretto all'utente finale né offre vantaggi competitivi. Esempi di attività indifferenziate includono l'approvvigionamento, la manutenzione e la pianificazione della capacità.

ambienti superiori

Vedi ambiente.

# V

vacuum

Un'operazione di manutenzione del database che prevede la pulizia dopo aggiornamenti incrementali per recuperare lo spazio di archiviazione e migliorare le prestazioni.

controllo delle versioni

Processi e strumenti che tengono traccia delle modifiche, ad esempio le modifiche al codice di origine in un repository.

Peering di VPC

Una connessione tra due VPCs che consente di indirizzare il traffico utilizzando indirizzi IP privati. Per ulteriori informazioni, consulta Che cos'è il peering VPC? nella documentazione di Amazon VPC.

vulnerabilità

Un difetto software o hardware che compromette la sicurezza del sistema.

# W

cache calda

Una cache del buffer che contiene dati correnti e pertinenti a cui si accede frequentemente. L'istanza di database può leggere dalla cache del buffer, il che richiede meno tempo rispetto alla lettura dalla memoria dal disco principale.

dati caldi

Dati a cui si accede raramente. Quando si eseguono interrogazioni di questo tipo di dati, in genere sono accettabili query moderatamente lente.

funzione finestra

Una funzione SQL che esegue un calcolo su un gruppo di righe che si riferiscono in qualche modo al record corrente. Le funzioni della finestra sono utili per l'elaborazione di attività, come il calcolo di una media mobile o l'accesso al valore delle righe in base alla posizione relativa della riga corrente.

Carico di lavoro

Una raccolta di risorse e codice che fornisce valore aziendale, ad esempio un'applicazione rivolta ai clienti o un processo back-end.

flusso di lavoro

Gruppi funzionali in un progetto di migrazione responsabili di una serie specifica di attività. Ogni flusso di lavoro è indipendente ma supporta gli altri flussi di lavoro del progetto. Ad esempio, il flusso di lavoro del portfolio è responsabile della definizione delle priorità delle applicazioni, della pianificazione delle ondate e della raccolta dei metadati di migrazione. Il flusso di lavoro del portfolio fornisce queste risorse al flusso di lavoro di migrazione, che quindi migra i server e le applicazioni.

VERME

Vedi scrivere una volta, leggere molti.

WQF

Vedi AWS Workload Qualification Framework.

scrivi una volta, leggi molte (WORM)

Un modello di storage che scrive i dati una sola volta e ne impedisce l'eliminazione o la modifica. Gli utenti autorizzati possono leggere i dati tutte le volte che è necessario, ma non possono modificarli. Questa infrastruttura di archiviazione dei dati è considerata immutabile.

# Z

exploit zero-day

Un attacco, in genere malware, che sfrutta una vulnerabilità zero-day.

vulnerabilità zero-day

Un difetto o una vulnerabilità assoluta in un sistema di produzione. Gli autori delle minacce possono utilizzare questo tipo di vulnerabilità per attaccare il sistema. Gli sviluppatori vengono spesso a conoscenza della vulnerabilità causata dall'attacco.

prompt zero-shot

Fornire a un LLM le istruzioni per eseguire un'attività, ma non fornire esempi (immagini) che possano aiutarla. Il LLM deve utilizzare le sue conoscenze pre-addestrate per gestire l'attività. L'efficacia del prompt zero-shot dipende dalla complessità dell'attività e dalla qualità del prompt. Vedi anche few-shot prompting.

applicazione zombie

Un'applicazione che prevede un utilizzo CPU e memoria inferiore al 5%. In un progetto di migrazione, è normale ritirare queste applicazioni.

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.