



Rendere operativa l'intelligenza artificiale agentica su AWS

AWS Guida prescrittiva



AWS Guida prescrittiva: Rendere operativa l'intelligenza artificiale agentica su AWS

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

I marchi e l'immagine commerciale di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in una qualsiasi modalità che possa causare confusione tra i clienti o in una qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà dei rispettivi proprietari, che possono o meno essere affiliati, collegati o sponsorizzati da Amazon.

Table of Contents

Introduzione	1
Aree di interesse	1
Destinatari principali	2
Obiettivi	2
Informazioni su questa serie di contenuti	3
Fondamenti per l'IA agentic	4
Aree di interesse	5
Intento e ambito	6
Strategia	6
Valore aziendale	8
Componibilità e collaborazione	8
Strategia	9
Valore aziendale	11
Multi-tenancy e controllo	12
Strategia	12
Valore aziendale	13
Autonomia affidabile	14
Strategia	14
Valore aziendale	15
Gestione del ciclo di vita	16
Strategia	16
Valore aziendale	17
Allineamento aziendale	18
Strategia	18
Distribuzione del software	21
Zone di intento	21
Evoluzione dell'SDLC	22
Preparazione delle squadre	24
Preparazione alla scalabilità	26
Team e modelli di proprietà	26
Gestione delle modifiche	27
Interoperabilità e collaborazione	28
Governance	29
Mentalità operativa	30

Dimensionamento	30
Conclusioni	32
Resources	34
Servizi AWS	34
Altre risorse AWS	35
Cronologia dei documenti	37
Glossario	38
#	38
A	39
B	42
C	44
D	47
E	51
F	53
G	55
H	56
I	57
L	60
M	61
O	65
P	68
Q	70
R	71
S	74
T	78
U	79
V	80
W	80
Z	81
.....	lxxxiii

Rendere operativa l'intelligenza artificiale agentica su AWS

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

Agosto [2025](#) (storia del documento)

L'intelligenza artificiale agentica non è una funzionalità, è un nuovo paradigma operativo. Organizzazioni che investono in architetture disciplinate, framework di fiducia e modelli di implementazione allineati al business guideranno la prossima generazione di imprese adattive e intelligenti.

L'intelligenza artificiale agentica rappresenta la convergenza di agenti software autonomi e intelligenza artificiale generativa. Unisce il comportamento decisionale e mirato degli agenti con le capacità di comprensione e generazione del linguaggio di modelli linguistici di grandi dimensioni (LLMs). Questi agenti possono ragionare, agire, adattarsi e collaborare in ambienti aziendali dinamici. Per rendere operativo questo potenziale, le aziende devono spostare la propria mentalità dall'implementazione di modelli all'infrastruttura degli agenti.

Questa guida fornisce una strategia organizzativa per trasformare l'intelligenza artificiale agentica da esperimenti isolati in un'infrastruttura su scala aziendale in grado di generare valore. Può aiutarti a incorporare agenti intelligenti nei flussi di lavoro con governance, scalabilità e allineamento aziendale.

Aree di interesse e raccomandazioni chiave

Questa guida si concentra sulle seguenti aree fondamentali per rendere operativa l'intelligenza artificiale agentica. Vengono fornite raccomandazioni organizzative e commerciali per ciascuna area di interesse:

- [Area di interesse 1: chiarire l'intento e l'ambito dell'agente](#)— Allineare gli agenti alle priorità aziendali e ai problemi cognitivi. Tratta gli agenti come compagni di squadra digitali, non solo come strumenti.
- [Area di interesse 2: Progettazione per la componibilità e la collaborazione](#)— Utilizza sistemi multiagente con architettura modulare, protocolli semantici e delega dinamica tramite agenti arbitri.
- [Area di interesse 3: Architetto per la multi-tenancy e il controllo](#)— Crea un'infrastruttura scalabile e sensibile ai tenant con servizi di agenti condivisi, governance centralizzata e accesso basato sui ruoli.

- [Area di interesse 4: creare fiducia attraverso identità, barriere e osservabilità](#)— Applica la tracciabilità, i controlli di runtime e la spiegabilità per guadagnare la fiducia degli stakeholder.
- [Area di interesse 5: Gestione del ciclo di vita](#)— Stabilisci pipeline di integrazione continua e implementazione continua (CI/CD), controllo rapido delle versioni, telemetria e riqualificazione continua per supportare le prestazioni e l'efficienza dell'intelligenza artificiale agentic.
- [Area di interesse 6: Allineare i modelli degli agenti ai modelli di business](#)— Monetizza le funzionalità degli agenti attraverso modelli basati sull'utilizzo, metriche interne del ROI e offerte commerciali.

Puoi utilizzare i consigli di questa guida per preparare la tua azienda all'intelligenza artificiale agentic su larga scala. Descrive in che modo le organizzazioni devono ristrutturarsi attorno all'IA agentic, tra cui la creazione DevOps di team di agenti (AgentOps), sistemi interoperabili e strategie di gestione del cambiamento che ne consentano l'adozione su larga scala. Sottolinea il pensiero decisionale e l'allineamento con il Well-Architected Framework. AWS

Destinatari principali

Questa guida è destinata agli architetti aziendali, ai responsabili della AI/ML progettazione e agli strateghi della trasformazione digitale che progettano e scalano sistemi agentic, incorporano l'intelligenza artificiale nei flussi di lavoro aziendali principali e rendono operativi gli agenti autonomi negli ambienti di produzione. LLMs Per comprendere i concetti e le raccomandazioni di questa guida, è necessario conoscere le moderne architetture e i sistemi distribuiti nativi del cloud, i modelli linguistici di grandi dimensioni, le funzionalità dei modelli di base e i principi della governance dell'IA e dell'ingegneria della piattaforma. DevOps

Obiettivi

Implementando i consigli contenuti in questa guida, l'organizzazione può ottenere i seguenti risultati di business:

- Accelerazione del processo decisionale e dell'esecuzione del flusso di lavoro tramite agenti autonomi e orientati agli obiettivi che riducono i colli di bottiglia e il carico cognitivo umano.
- Implementazioni scalabili ed economiche di funzionalità intelligenti tra le unità aziendali, attraverso piattaforme di agenti multitenant riutilizzabili.
- Maggiore resilienza, fiducia e governance nei sistemi di intelligenza artificiale, che consentono un'adozione sicura in ambienti regolamentati, mission critical o rivolti ai clienti.

Informazioni su questa serie di contenuti

Questa guida fa parte di una serie sull'intelligenza artificiale agentica su AWS. Per ulteriori informazioni e per visualizzare le altre guide di questa serie, consulta [Agentic AI](#) sul sito Web Prescriptive Guidance. AWS

Fondamenti strategici per l'IA agentica

I sistemi agentici non sono nuovi. Gli agenti software, tra cui l'automazione robotica dei processi (RPA) e i motori decisionali, esistono da decenni. Ma erano semplici e deterministici, progettati per seguire regole predefinite e logica simbolica per eseguire attività ripetitive e con poche variazioni. Con l'avvento dell'intelligenza artificiale generativa, il gioco è cambiato. I modelli linguistici di grandi dimensioni (LLMs) ora possono interpretare input complessi, generare risposte in modo dinamico e sintetizzare rapidamente le conoscenze. Ora puoi scalare l'agenzia senza una logica fragile o codificata. Ora gli agenti possono ragionare, prendere decisioni, richiamare strumenti, adattarsi al contesto e coordinarsi con altri agenti in tutti i flussi di lavoro. Possono operare in modo autonomo verso gli obiettivi, mantenere la memoria e riflettere sui risultati.

Tuttavia, la capacità pura non è sufficiente. L'intelligenza senza integrazione produce novità, non impatto. Per sbloccare il valore dei potenti LLMs, le aziende devono passare da esperimenti isolati a ecosistemi ingegnerizzati. Gli agenti devono essere trattati come servizi di livello produttivo che operano secondo la stessa disciplina di qualsiasi sistema aziendale. Ciò include governance, osservabilità, modelli di identità sicuri e gestione del ciclo di vita. Devono inoltre portare a risultati commerciali reali, non a potenziali speculativi. Questi sistemi dovrebbero essere progettati con confini chiari per il processo decisionale e la tolleranza agli errori. È importante incorporare meccanismi di ripristino automatizzati, monitoraggio delle prestazioni in tempo reale e gestione scalabile delle risorse. Ciò consente di gestire la natura dinamica e non deterministica delle interazioni tra agenti, mantenendo al contempo livelli di servizio coerenti nei flussi di lavoro aziendali.

A livello fondamentale, le aziende devono ripensare al modo in cui l'intelligenza è incorporata nel tessuto delle operazioni. Gli agenti devono essere progettati per integrarsi con i sistemi principali, rispettare le politiche aziendali e fornire un valore misurabile. Devono operare su larga scala, tra reparti, domini e contesti utente. Rendere operativa l'IA agentica è fondamentalmente una questione di utilizzo; è la differenza tra l'implementazione di un'intelligenza artificiale che esegue attività isolate e l'implementazione di agenti che evolvono il modello di business.

Agentic AI rappresenta una nuova filosofia operativa che richiede un cambiamento fondamentale nel modo in cui affrontiamo i sistemi, i processi e le persone per scalare l'intelligenza all'interno dell'organizzazione. Gli agenti diventano risorse strategiche che amplificano le capacità umane. Integrando l'intelligenza artificiale agentica nelle loro operazioni, le organizzazioni possono sbloccare informazioni che promuovono il valore aziendale, aumentano le capacità umane e ottimizzano i flussi di lavoro complessi.

Aree di interesse strategiche per l'intelligenza artificiale agentica

Per passare dai primi prototipi ai sistemi di produzione e generatori di valore, i team hanno bisogno di una strategia coerente che combini architettura, processo e pensiero di prodotto.

Molte organizzazioni si avvicinano ancora all'intelligenza artificiale con una mentalità incentrata sugli strumenti o sui modelli. L'intelligenza artificiale generativa ha amplificato la sperimentazione, ma spesso senza un chiaro allineamento alla strategia aziendale o risultati misurabili. Senza un ruolo strategico definito, gli agenti rischiano di diventare nuovi esperimenti che assorbono risorse anziché fornire valore scalabile. Per stabilire il ruolo strategico dell'IA agentica, le organizzazioni devono iniziare con le priorità aziendali. Identifica le aree di sovraccarico cognitivo, problemi decisionali o flussi di lavoro frammentati in cui l'autonomia può dare sollievo. Utilizza le dichiarazioni relative ai problemi specifiche del dominio per definire le responsabilità degli agenti. Tratta gli agenti come compagni di squadra digitali, non come strumenti, in grado di ragionare, delegare e adattarsi.

Le scienze decisionali sono la disciplina che combina scienza dei dati, analisi e modelli comportamentali per migliorare il processo decisionale. Dovrebbe essere integrato nelle prime fasi del processo di architettura degli agenti per allineare il design ai risultati aziendali. Identificando i modelli decisionali, simulando i compromessi e quantificando l'impatto sul valore, le scienze decisionali possono aiutarvi a individuare dove l'autonomia degli agenti può offrire il massimo valore. Le scienze decisionali possono accelerare le decisioni, ridurre gli errori e consentire adattamenti in tempo reale. Questa base basata sui dati fonda la progettazione degli agenti su informazioni misurabili e consente una più stretta integrazione con le tecnologie aziendali esistenti, come i motori di regole, le piattaforme di analisi e i modelli predittivi.

Per contribuire a stabilire il ruolo strategico degli agenti, questa sezione introduce le aree di interesse fondamentali che costituiscono la spina dorsale dell'operationalizzazione dell'intelligenza artificiale agentica. Ciascuna si riferisce a un lavoro fondamentale da svolgere dal punto di vista di un responsabile tecnico, architetto o proprietario del prodotto responsabile del modo in cui gli agenti vengono concepiti e progettati. Queste aree di interesse non sono fasi sequenziali. Vale la pena rivisitarle durante l'intero ciclo di vita del sistema per coltivare ecosistemi di agenti resilienti, scalabili e monetizzabili.

Questa sezione contiene le seguenti aree di interesse:

- [Area di interesse 1: chiarire l'intento e l'ambito dell'agente](#)

- [Area di interesse 2: Progettazione per la componibilità e la collaborazione](#)
- [Area di interesse 3: Architetto per la multi-tenancy e il controllo](#)
- [Area di interesse 4: creare fiducia attraverso identità, barriere e osservabilità](#)
- [Area di interesse 5: Gestione del ciclo di vita](#)
- [Area di interesse 6: Allineare i modelli degli agenti ai modelli di business](#)

Area di interesse 1: chiarire l'intento e l'ambito dell'agente

Job to done: «Aiutami ad assicurarmi che ogni agente risolva un problema reale con limiti chiari, non solo con una bella demo».

Agentic AI non riguarda solo lo sviluppo di capacità. Si tratta di risolvere il problema giusto, nel modo giusto, per il risultato giusto. Ciò inizia con l'essere completamente chiari sull'intento della soluzione di intelligenza artificiale agentica.

Strategia

Troppo spesso, le organizzazioni partono da ciò che il modello può fare (ad esempio chiamare APIs, rispondere a domande o generare riepiloghi) e adattano un caso d'uso a tale scopo. Ciò porta a problemi di ambito, scarsa integrazione e agenti tecnicamente imponenti ma inutili dal punto di vista operativo. Iniziate invece definendo il ruolo dell'agente attraverso domande specifiche come le seguenti:

- Di quale risultato specifico è responsabile l'agente?
- Di chi agisce per conto?
- Chi ne trae vantaggio?
- Dove inizia e finisce l'autonomia dell'agente?
- Cosa succede quando fallisce?

Un agente competente ha un lavoro chiaro, responsabilità definite e criteri di successo misurabili. Non pensate all'agente come a un assistente o a un chatbot. Invece, assegnagli un titolo professionale. Pensalo come un addetto al successo dei clienti, un addetto alla gestione dei resi dei prodotti o un addetto al monitoraggio della conformità.

Quando coinvolgi le parti interessate o i clienti, enfatizza la scalabilità e l'adattabilità dei sistemi di intelligenza artificiale agentica. Questi agenti si evolvono con l'azienda, migliorando continuamente

attraverso l'apprendimento e il feedback. Per ridurre la resistenza e accelerare l'adozione, evidenzia come gli strumenti agentic siano progettati pensando all'empatia dei lavoratori. Offrono trasparenza, controllo e meccanismi opzionali di esclusione che creano fiducia. Invece di sostituire le persone, gli agenti aumentano le capacità umane e il processo decisionale, aiutando i dipendenti a rimanere aggiornati e concentrarsi su attività di alto valore.

La chiave per un'implementazione di successo è l'allineamento dell'intelligenza artificiale agentica con risultati aziendali specifici e ad alto impatto. Incoraggia i team e i partner a iniziare con progetti pilota mirati che risolvano punti deboli visibili. I successi rapidi generano un ritorno sull'investimento (ROI) misurabile, generano il consenso interno e creano lo slancio per un'adozione più ampia.

Per guidare l'adozione e la maturità, le organizzazioni possono inquadrare la progettazione degli agenti secondo un modello evolutivo. L'autonomia, la complessità e l'impatto aziendale degli agenti aumentano progressivamente. Di seguito sono riportate le fasi di questo modello:

- Gli agenti osservatori estraggono informazioni dal rumore. Un esempio è un agente del sentiment di mercato che monitora la percezione del marchio su tutti i canali digitali.
- Gli assistenti supportano il processo decisionale umano. Un esempio è un agente di consulenza sulle trattative che sintetizza i dati della concorrenza e le condizioni di mercato per i team di vendita.
- Gli agenti autonomi agiscono in modo indipendente entro limiti definiti. Un esempio è un agente di allocazione delle risorse che regola dinamicamente l'infrastruttura cloud in base alla domanda.
- Gli agenti Orchestrator coordinano i flussi di lavoro multiagente. Un esempio è un agente di ottimizzazione della catena di fornitura che gestisce le interazioni tra agenti di inventario, logistica e previsione.
- Gli agenti innovatori generano nuove possibilità strategiche. Un esempio è un agente per l'innovazione dei modelli di business che analizza le tendenze del mercato e consiglia nuovi flussi di entrate.

Inquadrare gli agenti in base a questi risultati strategici e livelli di maturità aumenta l'attenzione, accelera l'adozione e rafforza la fiducia degli stakeholder.

Per supportare l'allineamento in quest'area di interesse Servizi AWS, come [Amazon Quick](#), puoi visualizzare indicatori chiave di performance (KPIs) collegati a risultati basati sugli agenti. Puoi utilizzare [Amazon CloudWatch](#) per monitorare il comportamento degli agenti, le metriche delle prestazioni e lo stato del sistema quasi in tempo reale. Usa il feedback operativo per ottimizzare

le interazioni degli agenti e l'uso delle risorse. [AWS CloudTrail](#) può fornire visibilità sull'attività degli agenti e sui modelli di integrazione durante le prime fasi di sperimentazione e perfezionamento.

Valore aziendale derivante dalla definizione dell'intento e dell'ambito

L'adozione dell'intelligenza artificiale agentica rappresenta un cambiamento fondamentale nel modo in cui le organizzazioni affrontano la trasformazione digitale e l'eccellenza operativa. Non si tratta solo di automazione. Si tratta di consentire un'autonomia intelligente che acceleri il processo decisionale e la realizzazione del valore.

I principali fattori di business includono quanto segue:

- **Vantaggio competitivo:** gli early adopter ottengono un vantaggio strategico grazie a informazioni più rapide, un servizio migliore e operazioni adattive.
- **Miglioramento dell'esperienza del cliente:** gli agenti offrono un supporto in tempo reale, personalizzato e sempre attivo che aumenta la soddisfazione e la fidelizzazione.
- **Efficienza operativa:** Agentic AI riduce in modo significativo il carico cognitivo umano automatizzando attività decisionali complesse e ripetitive. Ciò consente al personale di concentrarsi su attività di maggior valore e di ridurre i costi.

I casi d'uso reali in tutti i settori includono quanto segue:

- **Servizi finanziari:** gli agenti di intelligenza artificiale potrebbero fornire consulenze finanziarie personalizzate e rilevare le frodi.
- **Sanità:** gli agenti addetti al triage e ai piani di trattamento potrebbero migliorare la produttività clinica.
- **Vendita al dettaglio:** gli agenti potrebbero agire come assistenti intelligenti agli acquisti o ottimizzare l'inventario in tempo reale.
- **Produzione:** gli agenti potrebbero eseguire la manutenzione predittiva o coordinare le catene di approvvigionamento.

Area di interesse 2: Progettazione per la componibilità e la collaborazione

Job da fare: «Consentitemi di creare agenti come faccio per creare servizi, modulari e testabili, in modo che possano essere composti e orchestrati secondo necessità».

Molte iniziative di intelligenza artificiale iniziano come progetti pilota monolitici e incentrati sui modelli. Sono utili, ma sono difficili da scalare tra i domini o adattarsi a problemi complessi. Dai valore ai composti quando questi agenti sono progettati per interagire. In tecnologia, la componibilità è l'atto di combinare componenti modulari per creare una soluzione flessibile e scalabile in grado di adattarsi al cambiamento. Senza la componibilità, l'intelligenza rimane bloccata all'interno di flussi di lavoro specifici. Inoltre, la collaborazione tra agenti introduce complessità di orchestrazione, gestione dello stato e negoziazione dei protocolli che i team di automazione tradizionali potrebbero non essere in grado di gestire.

Strategia

Abbraccia il paradigma multiagente. Agenti modello come i reparti organizzativi: modulari, specializzati e interoperabili. Definisci interfacce chiare, formati contestuali condivisi e protocolli di comunicazione standard, come [Model Context Protocol \(MCP\)](#) o [Agent2Agent \(A2A\)](#). Adotta modelli di orchestrazione multiagente, come swarm, graph o coordinamento gerarchico. Questi modelli aiutano gli agenti a scoprire funzionalità e a richiedere servizi gli uni agli altri in modo dinamico, in flussi di lavoro paralleli, sequenziali o basati sul consenso, a seconda della struttura delle attività e del livello di fiducia.

Per promuovere una collaborazione scalabile e gestita, utilizza un agente arbitro. Questo tipo di agente è un'autorità neutrale che facilita la delega delle attività sulla base di capacità e strategie di fallback note. Pur non essendo un controllore centralizzato, un agente arbitro svolge un ruolo fondamentale in termini di fiducia e conformità. Garantisce che le attività sensibili o regolamentate vengano indirizzate solo agli agenti che soddisfano i requisiti di identità e policy. Funge da guardiano per i flussi di lavoro vincolati da policy. Implica l'isolamento e consente una delega spiegabile. Fondamentalmente, un agente arbitro non è un collo di bottiglia; coesiste con agenti che si coordinano autonomamente che operano in modo orizzontale. peer-to-peer Questi agenti delegano le attività secondarie, condividono il contesto e risolvono direttamente le dipendenze.

Questo modello ibrido supporta sia l'assegnazione deterministica (tramite l'agente arbitro) sia la collaborazione emergente. Unisce struttura e flessibilità. All'interno di questa architettura, gli agenti possono essere classificati nei seguenti ruoli specializzati:

- Agenti decisionali, come responsabili dell'applicazione delle politiche, allocatori di risorse e valutatori del rischio
- Agenti della conoscenza, come aggregatori di contesto, riconoscitori di schemi e rilevatori di anomalie
- Agenti di esecuzione, come esecutori di attività, controllori di qualità e responsabili dell'integrazione

Per coordinarsi efficacemente, i sistemi multiagente devono supportare solidi protocolli di interazione per la gestione dello stato, il ripristino degli errori e la risoluzione dei conflitti. Ciò favorisce la stabilità e la responsabilità anche quando gli agenti operano in modo indipendente.

Stabilisci regole chiare per la scalabilità, come la creazione di istanze degli agenti in base al carico, l'allocazione delle risorse in base al contesto e l'individuazione e la registrazione automatizzate delle funzionalità. Queste misure aiutano il sistema a crescere dinamicamente in risposta alla domanda o alla complessità.

Progetta gli agenti in modo che siano ready-to-use moduli all'interno di un substrato di messaggistica distribuito. Ad esempio, potresti utilizzare [Amazon EventBridge](#) con A2A o MCP anziché servizi isolati. Adotta modelli di controllo delle versioni, CI/CD pipeline e agenti per supportare la stabilità del sistema accelerando al contempo l'adozione interna e l'evoluzione del ciclo di vita. Incoraggia il riutilizzo e la standardizzazione del codice per ridurre gli attriti legati all'integrazione e promuovere un ecosistema resiliente.

La collaborazione è un moltiplicatore di forza. Sblocca la scalabilità, la specializzazione e la resilienza in ambienti multiagente. Per supportare questa collaborazione dinamica, le organizzazioni dovrebbero progettare un piano di controllo leggero per il coordinamento degli agenti. Questo piano di controllo include quanto segue:

- Registri di funzionalità che definiscono ciò che ogni agente può fare e supportano i metadati con versioni per il peer discovery
- Logica di arbitrato delle attività che utilizza agenti arbitri o supervisori per indirizzare le attività in base al contesto, alla disponibilità e alla politica
- Monitoraggio del ciclo di vita e dello stato che consente un contesto decisionale in tempo reale e passaggi sicuri

I piani di controllo assicurano che i sistemi multiagente rimangano estensibili, allineati alle politiche e tolleranti ai guasti, senza centralizzare l'autorità o rallentare le operazioni.

Tuttavia, gli ambienti con più agenti comportano anche sfide operative. Mantenere il contesto tra le interazioni tra agenti, gestire lo stato condiviso e coordinare le azioni può aumentare la complessità e i costi. I costi possono aumentare se si utilizzano token LLMs che consumano tali token durante la comunicazione tra agenti. Questi costi devono essere confrontati con i vantaggi aziendali combinati dell'autonomia intelligente su larga scala.

Per affrontare queste sfide, prendi in considerazione piattaforme agentiche che astragano le preoccupazioni chiave, come le seguenti:

- Protocolli di comunicazione e formati semantici standardizzati
- Logica di orchestrazione e routing dinamico integrati
- Gestione condivisa del contesto e della memoria tra agenti
- Gestione dei fallback e riduzione graduale in caso di guasto

Per i team che adottano strategie multiagente, l'approccio migliore è iniziare in piccolo e progettare su larga scala. Inizia con soluzioni mirate con un unico agente che risolvono problemi reali. Quindi, componi progressivamente questi agenti in un sistema cooperativo in cui ciascuno può scoprire, coordinare e delegare in base a obiettivi condivisi e a un contesto a livello di sistema.

È importante sottolineare che una solida gestione degli errori e un corretto degrado devono essere i principi di progettazione primari. I sistemi multiagente devono essere in grado di continuare i flussi di lavoro parziali o di avviare una logica di backup quando gli agenti non sono disponibili o si guastano. Ciò favorisce l'affidabilità senza accoppiamenti rigidi.

Servizi AWS offrono funzionalità robuste per supportare questa architettura su larga scala. [Amazon EventBridge](#) e [EventBridge Pipes](#) forniscono la spina dorsale strutturata e basata sugli eventi per la messaggistica multiagente. Per la gestione del comportamento modulare, [AWS AppConfig](#) consente la commutazione sicura e dinamica della configurazione tra le istanze degli agenti. Per supportare la gestione condivisa del contesto e della memoria, usa [Amazon DynamoDB](#) per una persistenza dello stato leggera e sensibile ai tenant e un rapido recupero del contesto tra gli agenti. Puoi utilizzare [Amazon Simple Storage Service \(Amazon S3\)](#) per archiviare cronologie di prompt strutturate, artefatti condivisi o output generati da agenti. Per flussi di lavoro più complessi che richiedono un coordinamento basato sullo stato, puoi orchestrare processi di lunga durata con checkpoint e logica di ripristino degli errori. [AWS Step Functions](#) Insieme, questi servizi consentono di creare sistemi multiagente componibili, resilienti e semanticamente connessi che si adattano alle esigenze aziendali.

Valore aziendale dei sistemi multiagente

Sebbene molte organizzazioni inizino il loro percorso verso l'IA con soluzioni a singolo agente, il pieno potenziale dell'intelligenza artificiale agentica viene sfruttato attraverso sistemi scalabili multiagente. Questi sistemi sono fondamentali per risolvere problemi complessi e distribuiti e creare ecosistemi di intelligenza artificiale robusti e flessibili che si evolvono con le esigenze aziendali.

I vantaggi aziendali principali dei sistemi multiagente includono quanto segue:

- **Scalabilità:** le attività e i carichi di lavoro possono essere distribuiti tra agenti specializzati per aumentare la capacità e le prestazioni.
- **Flessibilità:** gli agenti possono essere aggiunti, sostituiti o modificati con interruzioni minime, garantendo agilità in ambienti dinamici.
- **Resilienza:** la stabilità del sistema viene preservata anche in caso di guasto dei singoli agenti, grazie ai ruoli ridondanti e al failover intelligente.
- **Specializzazione:** gli agenti appositamente progettati eseguono le attività con maggiore efficienza e precisione.
- **Efficienza in termini di costi:** i componenti riutilizzabili degli agenti accelerano lo sviluppo e riducono il costo dell'implementazione di nuove funzionalità.

Sebbene i sistemi multiagente richiedano una pianificazione più anticipata, offrono agilità, velocità e capacità di innovazione a lungo termine. Le aziende che investono in architetture flessibili di collaborazione tra agenti sono in grado di implementare rapidamente nuove funzionalità di intelligenza artificiale, adattarsi alle mutevoli esigenze e guidare un panorama competitivo sempre più basato sugli agenti.

Area di interesse 3: Architetto per la multi-tenancy e il controllo

Job to be done: «Aiutami a scalare l'utilizzo degli agenti su più clienti senza perdere il controllo, la responsabilità o la visibilità».

I primi prototipi sono utili per dimostrare valore in modo isolato, ma la maggior parte delle aziende deve supportare contemporaneamente più clienti, reparti o flussi di lavoro. Ciò significa che ogni agente deve operare entro limiti di policy, dati e identità chiaramente definiti. Senza la multi-tenancy, le operazioni diventano fragili e costose e la governance diventa un mosaico.

Strategia

Segui i principi delle architetture Software as a Service (SaaS). Ad esempio, progettazione per l'isolamento degli inquilini, l'applicazione delle politiche e il controllo delle risorse. Progetta agenti e piattaforme di orchestrazione con memoria, configurazione e identità compatibili con i tenant. Per imporre i limiti, utilizzate i tag, il controllo degli accessi basato sui ruoli (RBAC) e l'ambito della gestione delle identità e degli accessi.

Adotta un livello di osservabilità unificato in cui la telemetria degli agenti viene aggregata per contesto del tenant. Implementa motori di policy centralizzati e commutazione delle funzionalità basata sulla configurazione per applicare regole di comportamento dinamiche.

Crea la distribuzione degli agenti come servizio. Consenti ai team o ai clienti interni di utilizzare le funzionalità degli agenti in modo scalabile e gestito APIs. AWS fornisce una solida base per questi modelli. Puoi utilizzare [Amazon Cognito](#) per gestire l'identità di utenti e tenant [AWS Organizatione le politiche di controllo del servizio \(SCPs\)](#) per la governance tra account e [AWS Resource Access Manager \(AWS RAM\)](#) per la condivisione sicura delle funzionalità. Inoltre, [AWS AppConfig](#) può gestire dinamicamente il comportamento degli agenti per tenant o ambiente. Questi servizi aiutano a far rispettare limiti e politiche supportando al contempo l'infrastruttura condivisa.

Questa transizione dall'implementazione statica al provisioning dinamico trasforma l'intelligenza artificiale agentic in una piattaforma a livello aziendale.

Valore aziendale delle piattaforme con agenti multi-tenant

La multi-tenancy è più di una comodità architettonica: è un acceleratore di business. Con la proliferazione di agenti intelligenti tra reparti e team, le organizzazioni devono supportare la crescita senza duplicare l'infrastruttura o frammentare la governance.

I principali vantaggi aziendali dei sistemi multi-tenant includono quanto segue:

- **Scalabilità:** una piattaforma multi-tenant per agenti consente ai team interni, alle unità aziendali o ai clienti di integrare più rapidamente le funzionalità di intelligenza artificiale senza bisogno di ambienti personalizzati.
- **Efficienza in termini di costi:** l'infrastruttura condivisa riduce al minimo le implementazioni ridondanti, consolida i costi operativi e semplifica la manutenzione tra gli ambienti.
- **Governance e riduzione dei rischi:** i controlli centralizzati delle policy, i modelli di identità e l'osservabilità aiutano gli agenti a operare in modo più sicuro e conforme, tra tutti i tenant.
- **Riutilizzabilità dei servizi:** per promuovere il riutilizzo e ridurre la duplicazione, è possibile offrire agenti sensibili ai tenant come servizi interni, ad esempio per l'arricchimento, la conformità o il riepilogo.

Di seguito sono riportati alcuni esempi di casi d'uso per sistemi multi-tenant:

- Un agente di conformità distribuito tra le filiali adatta la propria logica alle normative locali attraverso una configurazione specifica per il tenant. Ciò elimina la necessità di creare agenti separati per ogni regione.
- Un agente interno di automazione del flusso di lavoro serve più reparti con limiti di dati e autorizzazioni diversi. Mantiene l'isolamento accelerando al contempo l'adempimento delle attività.

Progettando gli agenti come multi-tenant-aware servizi, le organizzazioni evitano il sovraccarico di iniziative di intelligenza artificiale isolate. Al contrario, promuovono una piattaforma di intelligence unificata. Questa architettura consente l'implementazione scalabile, la coerenza operativa e un migliore ROI. Inoltre, semplifica l'espansione dell'adozione dell'IA in tutta l'azienda.

Area di interesse 4: creare fiducia attraverso identità, barriere e osservabilità

Job da fare: «Dammi la certezza che gli agenti agiranno in modo sicuro e prevedibile, soprattutto quando nessuno sta guardando».

Gli agenti autonomi sfidano i modelli di controllo tradizionali. La loro capacità di ragionare e agire in modo indipendente comporta dei rischi se non vengono gestiti correttamente. Senza chiari vincoli di proprietà, verificabilità o policy, possono allontanarsi dal comportamento previsto. Costruire la fiducia nelle organizzazioni richiede molto più della semplice affidabilità tecnica. Richiede spiegabilità, responsabilità e coerenza.

Strategia

Crea un sistema di controllo incentrato sull'identità come spina dorsale di un'autonomia affidabile. Ogni agente deve operare con un'identità verificabile, autorizzazioni specifiche e una cronologia di esecuzione tracciabile. Gli agenti devono essere incorporati in un [framework zero-trust](#) che includa l'associazione dei tenant, l'ereditarietà degli accessi contestuali e l'applicazione del runtime tramite guardrail e motori di policy. Ciò consente di controllare, annullare o limitare le azioni degli agenti in base alle regole organizzative e alla posizione di rischio.

Integra l'applicazione della fiducia in fase di esecuzione attraverso barriere intelligenti. Ciò include il controllo della frequenza e la limitazione in base ai modelli comportamentali o alle condizioni del carico di lavoro, ai limiti delle risorse applicati insieme all'auto-scaling e al punteggio decisionale per valutare il rischio. Crea trigger per coinvolgere i flussi di lavoro quando vengono superate le soglie. human-in-the-loop

Ogni agente deve inoltre essere trasparente e spiegabile. Incorpora la telemetria strutturata tramite registrazioni, tracce e riepiloghi di ragionamento per esporre la logica decisionale. Supporta percorsi decisionali e tracciamento dell'impatto. Questo ti aiuta a ricollegare le azioni degli agenti alle metriche o ai risultati chiave. Implementa meccanismi di rilevamento delle deviazioni che monitorino le deviazioni dal comportamento o dalle politiche previsti.

Introduci agenti riflettenti che osservano continuamente il comportamento degli agenti e i modelli del sistema. Dovrebbero segnalare anomalie o incongruenze in tempo reale. Questi agenti contribuiscono ai circuiti di feedback sulla governance che possono avviare la riconvalida, l'adattamento o la disattivazione delle funzionalità.

Istituisce consigli di amministrazione che esaminino le politiche degli agenti, approvino le modifiche alle funzionalità e supervisionino i protocolli di risposta agli incidenti. La fiducia deve essere guadagnata, misurata e continuamente rafforzata.

AWS fornisce una solida base per l'implementazione di questo quadro di fiducia:

- [AWS Identity and Access Management \(IAM\)](#) impone limiti di esecuzione e autorizzazione basati sui ruoli
- [Amazon CloudWatch](#) e [AWS X-Ray](#) il supporto offrono visibilità e tracciabilità complete.
- [Amazon GuardDuty](#) e [AWS Config](#) rileva anomalie di sicurezza o variazioni delle politiche.

Insieme, questi servizi consentono l'applicazione delle identità, la sicurezza in fase di esecuzione e una governance basata sulla fiducia su larga scala. Possono contribuire a rendere i sistemi autonomi potenti e affidabili.

Il valore aziendale dell'autonomia affidabile

Man mano che gli agenti diventano più autonomi, la fiducia diventa un fattore fondamentale per l'adozione, la governance e le prestazioni operative aziendali. Stabilire una base di identità, osservabilità e barriere aiuta le organizzazioni a scalare l'intelligenza artificiale agentica in domini sensibili, senza sacrificare la governance o il controllo.

I principali fattori di business includono quanto segue:

- Garanzia della governance: solidi modelli di identità, audit trail e limiti di autorizzazione riducono il rischio di conformità e supportano l'allineamento normativo.
- Continuità operativa: le barriere di runtime e il rilevamento delle anomalie aiutano a prevenire comportamenti indesiderati e supportano il ripristino automatico in caso di guasti estremi.

- **Fiducia delle parti interessate:** la spiegabilità delle decisioni e la telemetria creano fiducia tra le parti interessate interne, i gestori del rischio e i revisori esterni.
- **Resilienza agli incidenti:** l'osservabilità integrata accelera l'analisi delle cause alla radice e i tempi di risposta in caso di problemi.

Ecco alcuni esempi di casi d'uso:

- Nei servizi finanziari, gli addetti al rilevamento delle frodi devono esporre le proprie motivazioni, registrare ogni azione con un'identità tracciabile e operare secondo ruoli IAM ben definiti.
- Nel settore sanitario, gli agenti di triage autonomi devono applicare controlli di sicurezza in fase di esecuzione, passare alla revisione umana quando vengono raggiunte le soglie e fornire registri completi per la supervisione clinica.

Integrando meccanismi di fiducia nel ciclo di vita degli agenti, le organizzazioni possono consentire ai propri sistemi di funzionare in modo autonomo e responsabile. Questa base riduce i rischi e consente agli agenti di agire per conto dell'azienda con trasparenza e integrità.

In definitiva, l'autonomia affidabile accelera l'adozione offrendo sia agli utenti che ai dirigenti la fiducia necessaria per scalare gli agenti intelligenti tra le operazioni principali.

Area di interesse 5: Gestione del ciclo di vita

Job da fare: «Assicurati che il mio team possa migliorare gli agenti nel tempo, senza caos o atti eroici».

A differenza delle applicazioni tradizionali che sono modellate solo dal codice, il comportamento degli agenti è modellato anche dai prompt, dalla memoria, dagli strumenti e dal contesto di addestramento. Questi fattori variano nel tempo. La deriva erode l'affidabilità, aumenta i costi e rende quasi impossibile il debugging. Senza controlli sul ciclo di vita, gli agenti smettono di fornire valore e iniziano ad accumulare rischi.

Strategia

Stabilisci DevOps per gli agenti (AgentOps) come prassi. Integra CI/CD pipeline personalizzate per gli agenti. Utilizza queste pipeline per testare i risultati tempestivi, convalidare le integrazioni degli strumenti e profilare il comportamento in termini di costi e prestazioni. Conserva la cronologia delle versioni di prompt, policy e interazioni tra modelli.

Utilizza i cicli di feedback derivanti dai dati di osservabilità per avviare la riqualificazione, la tempestiva ottimizzazione o il ritiro degli agenti. Incorpora meccanismi di riflessione a livello di sistema, come un registro dei miglioramenti, per istituzionalizzare l'apprendimento.

Crea una dashboard di telemetria delle prestazioni che mostri l'accuratezza delle decisioni, la latenza, i costi e l'affidabilità. Per semplificare e accelerare la gestione del ciclo di vita utilizzando l'AWS infrastruttura, i team possono utilizzare i toolkit per gli agenti. Un esempio è lo [Strands Agents SDK](#), che fornisce strumenti strutturati per il controllo rapido delle versioni, la registrazione degli strumenti e l'integrazione CI/CD con, ad esempio, e. Servizi AWS [AWS CodePipeline](#) [AWS Cloud Development Kit \(AWS CDK\)](#) [AWS Lambda](#). Inoltre, utilizza [Amazon S3](#) e Amazon [Elastic File System \(Amazon EFS\)](#) per archiviare gli artefatti degli agenti e i dati di addestramento. Utilizzalo [AWS Step Functions](#) per automatizzare flussi di lavoro complessi di riqualificazione o convalida. Puoi utilizzare [Amazon SageMaker AI](#) quando gli agenti richiedono l'ottimizzazione di modelli personalizzati o l'ottimizzazione di flussi di lavoro oltre l'orchestrazione LLM. La disciplina del ciclo di vita trasforma gli agenti da esperimenti in risorse durevoli e in continua evoluzione.

Nel tempo, questo sistema del ciclo di vita costituisce la spina dorsale dell'innovazione. Ti aiuta a ricomporre, riqualificare e ridistribuire le funzionalità con agilità. Questo trasforma il livello degli agenti in un sistema vivente, in grado di evolversi in risposta sia al feedback che alle opportunità.

Valore aziendale della gestione del ciclo di vita

Una gestione efficace del ciclo di vita è un fattore chiave per le prestazioni degli agenti e l'efficienza dei costi. Garantisce che gli agenti intelligenti continuino a fornire risultati accurati, affidabili e allineati al valore man mano che si evolvono. Per impostazione predefinita, gli agenti non rimangono preziosi. Devono evolversi in sintonia con i mutevoli requisiti aziendali, i flussi di lavoro e gli ambienti di dati. Un AgentOps team disciplinato aiuta gli agenti a rimanere accurati, efficienti e allineati agli obiettivi aziendali nel tempo.

I principali fattori di business includono quanto segue:

- **Coerenza delle prestazioni:** test continui, convalida tempestiva e riqualificazione aiutano gli agenti a mantenere la qualità delle decisioni in base a condizioni e set di dati mutevoli.
- **Ottimizzazione dei costi:** la profilazione basata sulla telemetria identifica strumenti inefficienti, richieste ad alto numero di token o esecuzioni non necessarie. Quindi, puoi eseguire la regolazione per ridurre i costi operativi.
- **Iterazione più rapida:** automazione del ciclo di vita che CI/CD accelera i cicli di sviluppo, aiutando i team a sperimentare, implementare e migliorare gli agenti con sicurezza.

- Riduzione del rischio: il rapido controllo delle versioni, il supporto per il rollback e i meccanismi di valutazione strutturati aiutano a prevenire le regressioni e supportano una gestione delle modifiche sicura e affidabile.

I casi d'uso di esempio includono quanto segue:

- Un agente dell'assistenza clienti viene monitorato per quanto riguarda la latenza, il costo del modello e il feedback degli utenti. L'osservabilità rivela un aumento dei costi, che richiede una nuova ottimizzazione dei prompt incorporati e della logica del modello di fallback.
- Un agente di riepilogo dei contratti viene aggiornato in base al feedback dei team legali. Le istruzioni con versioni diverse vengono testate in ambienti sandbox prima del rilascio in produzione, a supporto della sicurezza e della qualità.

Con la gestione strutturata del ciclo di vita, le organizzazioni vanno oltre la manutenzione reattiva per passare al miglioramento proattivo e continuo. Gli agenti diventano risorse digitali adattive che vengono misurate, perfezionate e riconvalidate rispetto agli obiettivi aziendali. Questa pratica trasforma gli ecosistemi di agenti in sistemi ad alte prestazioni, attenti ai costi e resilienti che offrono valore duraturo rimanendo al passo con il cambiamento.

Area di interesse 6: Allineare i modelli degli agenti ai modelli di business

Job to done: «Mostratemi l'impatto, in modo da poter giustificare un investimento continuo».

Anche gli agenti tecnicamente capaci diventano responsabili se non sono legati ai risultati aziendali. Gli agenti devono puntare all'efficienza, alla monetizzazione o alla differenziazione strategica. Tuttavia, la maggior parte delle aziende fatica a definire in che modo gli agenti si adattano ai prezzi, alla confezione o ai modelli di utilizzo. Senza un chiaro allineamento al valore aziendale, è difficile giustificare la scalabilità o addirittura il mantenimento dell'investimento.

Strategia

Adotta pratiche di gestione dei prodotti. Tratta gli agenti come servizi monetizzabili con un ROI misurabile. Definisci strategie di prezzo basate su decisioni, sessioni o risultati. Quindi, raccogli le funzionalità degli agenti in offerte a più livelli allineate ai segmenti di clienti o alle unità aziendali interne.

Per promuovere la sostenibilità, le organizzazioni devono sfruttare sia il valore diretto che i moltiplicatori di crescita attraverso l'impiego di agenti. Prendi in considerazione l'utilizzo delle seguenti metriche del ROI per misurare il valore immediato:

- Costo per decisione: confronta i costi di elaborazione degli agenti con quelli umani equivalenti.
- Compressione dei tempi: quantifica il valore dei cicli accelerati, come vendite o approvazioni più rapide.
- Riduzione degli errori: misura i risparmi derivanti da una maggiore precisione, coerenza e conformità.

Oltre a questi vantaggi immediati, gli agenti possono sbloccare le seguenti opportunità di crescita a lungo termine:

- Capability Stacking: combina i servizi degli agenti per creare soluzioni verticali specifiche per il dominio.
- Effetti di rete: aumenta il valore attraverso ecosistemi multiagenti in cui il coordinamento si unisce all'utilità.
- Estensione del mercato: generazione di nuovi flussi di entrate attraverso servizi utilizzabili esternamente e abilitati agli agenti.

Crea cicli di feedback sulla base di metriche aziendali (come il risparmio sui costi, l'aumento delle conversioni o) per promuovere l'evoluzione continua degli agenti. time-to-resolution Analizza la telemetria di utilizzo e i punteggi di soddisfazione degli utenti per affinare l'allineamento dei valori e le priorità della roadmap. Collegando le funzionalità degli agenti direttamente ai modelli di business, le organizzazioni si posizionano per acquisire valore sostenibile e cumulabile, non solo risultati tecnici.

Quanto segue Servizi AWS supporta questo allineamento fornendo solidi framework di tracciamento e monetizzazione:

- [AWS Cost Explorer](#) e [Amazon CloudWatch](#) forniscono informazioni dettagliate sui costi e sull'efficienza operativa per agente.
- [Amazon API Gateway](#) consente l'accesso misurato, la limitazione della velocità e la determinazione dei prezzi su più livelli per gli endpoint degli agenti.
- [Marketplace AWS](#) fornisce un canale per agenti editoriali e soluzioni agentiche come prodotti commerciali.

Questi servizi consentono di trasformare le funzionalità degli agenti in offerte digitali scalabili e basate sul valore, in linea con le strategie di crescita e monetizzazione aziendali.

Distribuzione di software in evoluzione per l'intelligenza artificiale agentica

La moderna distribuzione del software si è basata su un semplice presupposto: il controllo dei sistemi spediti è assicurato dal cliente. Definisci i requisiti, scrivi la logica, esegui test in base ai risultati previsti e distribuisce servizi prevedibili. Anche Agile e DevOps gli approcci si basano ancora sul principio secondo cui ogni sprint offre qualcosa di deterministico, verificabile e in gran parte soggetto alla supervisione umana.

L'intelligenza artificiale agentica stravolge queste fondamenta. I sistemi agentici interpretano, ragionano e si adattano anziché seguire gli script. Il loro comportamento dipende dal codice che scrivi, dal contesto in cui operano, dagli input che ricevono, dagli strumenti a cui possono accedere e dagli obiettivi assegnati. In breve, non seguono gli ordini; perseguono i risultati.

Ciò rende la consegna meno incentrata sul controllo e più sull'allineamento. Invece di fornire istruzioni, è necessario modellare il modo in cui si comporta. Ciò significa che il tradizionale ciclo di vita dello sviluppo del software (SDLC) non è più adatto perché è stato progettato per sistemi basati sulla logica e controllati dall'uomo.

Questa sezione contiene i seguenti argomenti:

- [Zone di intento per l'intelligenza artificiale agentica](#)
- [Evoluzione del ciclo di vita di distribuzione per l'IA agentica](#)
- [Preparazione dei team per l'IA agentica](#)

Zone di intento per l'intelligenza artificiale agentica

Invece di fasi rigide, come la definizione, la costruzione, il test e il rilascio, abbiamo bisogno di un modello che abbracci l'autonomia, l'incertezza e l'emergenza. Al contrario, si utilizzano zone di intento. Una zona di intento definisce uno spazio limitato in cui un agente può operare con autonomia, entro limiti. L'obiettivo è passare dalla microgestione di ogni attività alla progettazione di ambienti in cui gli agenti possano agire, apprendere e collaborare in sicurezza. Siete voi a specificare il cosa (il risultato desiderato), il perché (l'intento) e i limiti (i vincoli, le politiche e i limiti di fiducia). Dati questi limiti e queste informazioni, l'agente capisce come.

Invece di una catena di montaggio, pensate all'ambiente come a uno spazio aereo. Sei tu a controllare chi può entrare, cosa può fare e dove può andare. Ma una volta dentro, sono liberi di navigare secondo necessità. È così che i sistemi agentici scalano senza caos.

Questo non è solo un cambiamento filosofico, è pratico. L'output non deterministico dei sistemi basati su agenti non può essere completamente testato tramite test unitari. Non può avere versioni come i binari statici. Gli agenti cambiano nel tempo, si adattano ai nuovi dati e interagiscono con altri sistemi in modi imprevedibili. Cercare di distribuirli utilizzando modelli tradizionali porta a architetture fragili e non scalabili. Nel peggiore dei casi, porta a una falsa fiducia in sistemi che in realtà non è possibile governare.

Quando i team adottano una distribuzione basata sugli intenti, ottengono due vantaggi:

- Controlla dove è più importante: definiscono i confini anziché i risultati.
- Scalabilità tramite delega: consentono agli agenti di gestire complessità che gli umani non possono codificare.

In questo modo si passa da prototipi isolati a sistemi agentici reali di livello di produzione in grado di fornire valore in modo ripetuto e affidabile.

Evoluzione del ciclo di vita di distribuzione per l'IA agentica

Per supportare un comportamento intelligente e adattivo, l'SDLC deve essere riformulato dal controllo deterministico all'intento adattivo. Di seguito sono riportate le modifiche necessarie per far evolvere il tradizionale SDLC per l'intelligenza artificiale agentica:

- La pianificazione diventa progettazione d'intenti. I team definiscono obiettivi, vincoli e comportamenti previsti degli agenti. Le politiche e i criteri di successo sono formulati in termini di allineamento, non di logica.
- L'architettura diventa impalcatura. I team si concentrano sulla definizione di ruoli, interfacce, barriere, meccanismi di riserva e osservabilità piuttosto che sulla creazione di script per ogni percorso decisionale.
- Il test diventa una valutazione comportamentale. Piuttosto che affermare risultati specifici, i team convalidano se gli agenti rispettano i limiti accettabili e soddisfano gli intenti con input diversi.
- L'implementazione diventa un'orchestrazione continua. I sistemi Agentic vengono implementati con controlli di runtime, monitoraggio in tempo reale e canali di feedback che consentono la regolazione in tempo reale.

- L'iterazione diventa feedback e adattamento. Invece dei tradizionali cicli di modifica del codice, i team osservano come si evolvono gli agenti, dove hanno successo o quando vanno alla deriva. Se necessario, i team intervengono aggiornando i vincoli, riqualificando e aggiungendo o modificando meccanismi di controllo.

Le pratiche esistenti incentrate sull'iterazione, sulla sperimentazione e sul feedback rapido sono a metà strada. Il passaggio ai sistemi agentici non è un rifiuto dei principi Agile. In effetti, è una loro evoluzione naturale. Il pensiero agile enfatizza l'adattabilità, il feedback e le soluzioni di lavoro rispetto ai piani rigidi. Ciò si allinea perfettamente con la natura dei sistemi agentici, che apprendono, si adattano e rispondono al contesto in tempo reale. Se state già eseguendo cicli brevi, convalidando rapidamente le ipotesi e gestendo l'incertezza attraverso la distribuzione continua, siete ben attrezzati per guidare questa transizione.

Ma ci sono differenze fondamentali. L'approccio Agile tradizionale presuppone che ciò che viene fornito sia deterministico. Si presuppone che, una volta costruito, l'oggetto si comporti in modo coerente e prevedibile, con risultati ripetibili per gli stessi input. Questa ripetibilità consente di eseguire il debug, il test e l'iterazione con sicurezza. I sistemi agentici rompono questo modello. Sono probabilistici, sensibili al contesto e in grado di evolversi indipendentemente. Ciò significa che alcune pratiche Agile diventano meno utili, come il monitoraggio della velocità basato sul completamento della storia, criteri di accettazione rigorosi o la pianificazione deterministica degli sprint.

I seguenti aspetti dell'SDLC tradizionale si applicano all'intelligenza artificiale agentica:

- Sviluppo e distribuzione iterativi
- Il feedback dei clienti come segnale principale
- Collaborazione interfunzionale
- Integrazione e implementazione continue

I seguenti aspetti dell'SDLC tradizionale devono evolversi per l'intelligenza artificiale agentica:

- Ridefinisci il termine in base all'intento. Concentratevi sul fatto che il comportamento dell'agente soddisfi l'obiettivo prefissato entro i vincoli definiti.
- Passa dai criteri di accettazione ai limiti comportamentali.
- Espandi la definizione di fatto per includere la prontezza in fase di esecuzione, che include meccanismi di osservabilità, spiegabilità e feedback che supportano l'apprendimento continuo e la fiducia.

- Dai priorità ai cicli di feedback in tempo reale e al monitoraggio del comportamento rispetto alla pianificazione anticipata

La buona notizia è che non è necessario buttare via il playbook SDLC. Devi solo passare dalla gestione del codice alla definizione della condotta. Nei sistemi agentici, il successo non dipende solo dal fatto che il software funzioni, ma anche dal modo in cui si comporta.

Preparazione dei team per l'IA agentica

L'ingegneria del software non scomparirà. Si sta evolvendo. Il lavoro passa dalla scrittura di funzioni alla definizione di framework e meccanismi di controllo per un comportamento intelligente. Nel mondo dell'intelligenza artificiale agentica, la costruzione non è più la parte difficile, ma la gestione delle emergenze lo è. Per la maggior parte dei team di ingegneri, l'evoluzione sembra un cambiamento di mentalità piuttosto che un salto tecnico. Invece di chiedere «Cosa farà il sistema?» la domanda diventa «Cosa l'abbiamo autorizzata a perseguire e come faremo a sapere se proseguirà sulla rotta giusta?»

Per i team di ingegneri, l'evoluzione verso l'intelligenza artificiale degli agenti richiede le seguenti modifiche:

- Un cambiamento culturale: i team devono abituarsi all'incertezza e all'autonomia in sistemi che non controllano completamente.
- Nuovi ruoli: i progettisti di intenti, i tester comportamentali e gli ingegneri dell'osservabilità diventano fondamentali per l'erogazione.
- Linguaggio condiviso: i team hanno bisogno di una comprensione chiara e condivisa degli obiettivi, dei limiti e dei segnali di successo, proprio come una volta avevano bisogno di specifiche e casi di test.

Man mano che l'intelligenza artificiale generativa matura, vedremo sempre più sistemi agentici interagire con clienti, prodotti e operazioni. Le organizzazioni che avranno successo non saranno quelle con i modelli migliori. Saranno quelle in grado di integrare gli agenti nei flussi di lavoro del mondo reale con sicurezza, controllo e velocità. Ciò significa che i modelli di distribuzione e i team di progettazione devono evolversi insieme. Le zone di intento ti danno l'astrazione necessaria per farlo. Ti aiutano a rendere operativa l'autonomia senza rinunciare alla responsabilità. Offrono anche un framework condiviso tra i team per aiutare a governare sistemi che non possono essere codificati.

Per ulteriori informazioni sulla preparazione dei team per l'IA agentic, consulta la sezione [Preparare il business per l'IA agentic](#) su larga scala di questa guida.

Preparare l'azienda all'IA agentica su larga scala

Man mano che le [aree di interesse](#) descritte in questa guida convergono, l'intelligenza artificiale agentica passa da funzioni isolate a un livello di intelligenza unificato che può essere inteso come una piattaforma di funzionalità. Questa piattaforma non si limita a eseguire attività. Si evolve, si adatta e si coordina tra i domini. Gli agenti diventano servizi modulari, riutilizzabili e individuabili che accelerano l'innovazione, riducono il carico cognitivo e promuovono risultati misurabili in tutta l'azienda. Questa visione della piattaforma pone le basi per un'intelligenza scalabile integrata in tutto il modello operativo.

L'operationalizzazione dell'IA agentica richiede molto più della semplice implementazione di agenti intelligenti. Richiede una trasformazione fondamentale del modo in cui l'azienda organizza i team, progetta i processi e governa la tecnologia. Proprio come il passaggio al cloud o a modelli operativi DevOps ridefiniti, l'intelligenza artificiale agentica introduce una nuova era di automazione delle decisioni, apprendimento continuo e coordinamento autonomo. Il successo dipende dall'allineamento dei sistemi, delle persone e dei processi attorno a questa nuova filosofia operativa.

Questa sezione contiene i seguenti argomenti:

- [Allineamento dei team e dei modelli di proprietà](#)
- [Gestione del cambiamento e preparazione organizzativa](#)
- [Progettazione per l'interoperabilità e la collaborazione](#)
- [Integrare la governance in un tessuto agentico](#)
- [Adottare una mentalità operativa che mette al primo posto le decisioni](#)
- [Scalabilità mirata e mirata](#)

Allineamento dei team e dei modelli di proprietà

Il primo passo verso la maturità è l'allineamento interfunzionale. Le aziende devono creare AgentOps team che includano AI/ML professionisti e specialisti di settore, ad esempio architetti di sistemi distribuiti, ingegneri del software, proprietari dei prodotti, responsabili della conformità e architetti di piattaforme. Questi team gestiscono congiuntamente l'intero ciclo di vita di un agente, dalla progettazione e implementazione alla riqualificazione e al monitoraggio.

Il provisioning e il rilascio degli agenti devono seguire pratiche native del cloud, come l'utilizzo di and for infrastructure as code [AWS Cloud Development Kit \(AWS CDK\)](#) e [AWS CodePipeline](#)

distribuzione automatizzata. Questa struttura favorisce la responsabilità condivisa e accelera l'iterazione. Così come DevOps unifica lo sviluppo e le operazioni, AgentOps collega l'intelligence con la governance e l'esecuzione.

Per essere efficaci, questi team hanno anche bisogno di un linguaggio condiviso. Gli stakeholder aziendali devono capire [cosa sono gli agenti](#), [come operano](#) e [quali risultati generano](#). La formazione e l'abilitazione interna sono essenziali. Demistificando gli agenti e incorporando questo modello mentale nelle conversazioni quotidiane, le organizzazioni sbloccano una partecipazione più ampia e un'innovazione più allineata.

Per accelerare lo sviluppo e l'integrazione degli agenti che utilizzano Servizi AWS, i team possono adottare framework come [Strands Agents SDK, che offre strumenti basati su CLI per lo scaffolding, la configurazione e gli agenti](#) di confezionamento. Strands Agents è progettato per funzionare perfettamente con AWS infrastrutture come [Amazon Bedrock](#), [AWS Lambda](#) [Amazon EventBridge](#), and. AWS CDK AWS CodePipeline Consente la prototipazione e l'implementazione rapide mantenendo standard di livello di produzione.

Ma la struttura e gli strumenti da soli non bastano. La scalabilità dell'intelligenza artificiale agentica richiede una preparazione culturale, educativa e di leadership deliberata per garantire che l'adozione si radichi in tutta l'organizzazione.

Gestione del cambiamento e preparazione organizzativa

Per scalare con successo l'intelligenza artificiale agentica non basta implementare infrastrutture o agenti intelligenti. Richiede un approccio strutturato al cambiamento organizzativo. Ciò include la preparazione culturale, lo sviluppo delle competenze, cicli di feedback basati sulle metriche e l'allineamento dei dirigenti per garantire che l'adozione sia sia intenzionale che sostenibile.

Promuovere l'evoluzione culturale

- Posiziona gli agenti come compagni di squadra, non come sostituti, per ridurre la resistenza e creare fiducia.
- Comunica in modo trasparente sulle capacità e i limiti degli agenti per stabilire aspettative realistiche.
- Stabilisci protocolli di trasferimento chiari per quando gli agenti devono affidare le decisioni a un'autorità superiore o delegare parti del processo a un collaboratore umano.

Stabilire un quadro per lo sviluppo delle competenze

- Offri una formazione basata sui ruoli su misura per ingegneri, responsabili di prodotto, responsabili di dominio e responsabili della conformità.
- Crea centri di eccellenza per condividere best practice, modelli di strumenti e risorse riutilizzabili.
- Associa specialisti di intelligenza artificiale a esperti di settore attraverso programmi di tutoraggio per colmare le lacune di conoscenza.

Definisci metriche e cicli di feedback

- Ancorate gli aspetti tecnici e commerciali KPIs al valore strategico per valutare l'impatto. Alcuni esempi di valore includono la latenza decisionale, l'accuratezza della risoluzione e il risparmio sui costi.
- Acquisisci in modo sistematico e continuo il feedback degli utenti per evidenziare i punti di attrito e le sfide di adozione.
- Conduci periodiche retrospettive per valutare le prestazioni degli agenti, le tendenze di utilizzo e le opportunità di miglioramento.

Allinea la leadership dall'alto

- Ottieni la sponsorizzazione dei dirigenti collegando le iniziative degli agenti ai risultati strategici e al ROI.
- Forma comitati di governance interfunzionali che includano la leadership tecnica e aziendale.
- Personalizza le strategie di comunicazione per garantire chiarezza e coinvolgimento a tutti i livelli organizzativi.

Questo approccio sistematico alla gestione del cambiamento assicura che l'implementazione della tecnologia sia accompagnata dalla maturità organizzativa. Crea una base per la fiducia, l'adozione e il valore aziendale a lungo termine.

Progettazione per l'interoperabilità e la collaborazione

Le implementazioni di agenti isolati offrono vantaggi locali. Ma il valore aziendale emerge quando gli agenti possono scoprire, richiamare e collaborare tra loro in modo dinamico. Ciò significa definire standard per la registrazione, l'autenticazione e lo scambio di funzionalità degli agenti. Dal punto di

vista architettonico, ciò rispecchia il passaggio dai monoliti ai microservizi, che sono unità componibili, riutilizzabili e liberamente accoppiate che risolvono insieme problemi complessi.

[I protocolli emergenti, come A2A e MCP, sono fondamentali.](#) Consentono l'interoperabilità semantica tra agenti, strumenti e sistemi di memoria. A2A supporta l'interazione tra pari, che consente agli agenti di negoziare la proprietà delle attività, condividere il contesto e coordinare i flussi di lavoro. MCP completa questo aspetto offrendo schemi condivisi per lo scambio di dati contestuali tra gli agenti e i loro ambienti. Standardizza il modo in cui le funzioni vengono richiamate, sono accessibili e gli stati APIs vengono mantenuti. Insieme, questi protocolli promuovono l'estensibilità, la coerenza e la manutenibilità a lungo termine in tutto l'ecosistema degli agenti.

La governance rimane fondamentale. I livelli di controllo, come gli arbiter agent, consentono una delega basata sulle policy senza introdurre colli di bottiglia centralizzati. Questi agenti agiscono come broker fiduciari. Fanno rispettare i limiti lasciando che gli altri agenti si organizzino da soli. La collaborazione tra agenti aiuta le organizzazioni a scalare i propri ecosistemi di intelligenza artificiale agentica con agilità e fiducia.

Integrare la governance in un tessuto agentico

Una maggiore autonomia comporta maggiori rischi. La governance deve essere integrata nell'architettura degli agenti sin dal primo giorno. Ciò include la definizione dei limiti delle politiche che definiscono ciò che gli agenti sono autorizzati a fare, l'applicazione di modelli di identità che determinano per chi agiscono e l'implementazione della spiegabilità e della tracciabilità. I sistemi di osservabilità devono acquisire la telemetria sul comportamento degli agenti utilizzando servizi come [Amazon CloudWatch](#) e [AWS X-Ray](#), che forniscono registrazione centralizzata e tracciamento distribuito tra i flussi di lavoro degli agenti. Gli agenti Reflective possono controllare e valutare continuamente le prestazioni sulla base di questi feed di telemetria.

La governance deve inoltre evolversi man mano che l'ecosistema degli agenti matura. Man mano che gli agenti diventano più capaci e più autonomi, i meccanismi di supervisione devono diventare più adattivi. Gli aggiornamenti delle policy, il capacity gating e i vincoli comportamentali di runtime devono essere dinamici e applicabili su larga scala. La fiducia non è una funzionalità introduttiva. Viene continuamente rafforzata attraverso l'architettura, il comportamento e il processo. [AWS Identity and Access Management \(IAM\)](#) e [AWS AppConfig](#) svolgono un ruolo fondamentale nell'applicazione di identità sicure, limiti di autorizzazione di runtime e alternanze di comportamento specifiche dell'ambiente tra gli agenti.

Adottare una mentalità operativa che mette al primo posto le decisioni

L'automazione tradizionale si concentra sull'efficienza dei processi, che consiste nell'esecuzione di script o flussi di lavoro predefiniti in modo più rapido e affidabile. L'intelligenza artificiale agentica, al contrario, introduce l'automazione che mette al primo posto le decisioni. Gli agenti valutano il contesto, valutano le opzioni e adattano il comportamento in tempo reale. Questo passaggio da una mentalità incentrata sull'esecuzione a quella decisionale richiede una nuova riflessione sulle metriche e sui risultati di successo. Invece di misurare il successo esclusivamente in base al completamento delle attività, il successo dell'intelligenza artificiale agentica viene misurato in base all'allineamento della decisione con l'intento, le politiche e le condizioni in evoluzione.

Piuttosto che misurare solo il completamento delle attività o la durata del ciclo, le organizzazioni devono valutare la qualità delle decisioni e la reattività al cambiamento time-to-action. KPIs dovrebbero includere metriche come:

- Qualità delle decisioni: in che modo l'agente ha personalizzato la sua risposta in base all'utente o allo scenario specifico? Ha preso decisioni dettagliate in linea con gli obiettivi aziendali e il contesto dell'utente?
- Time-to-action — Con quale rapidità e intelligenza l'agente ha valutato una situazione e ha risposto? La latenza era sufficientemente bassa da sembrare adattiva e simile a quella umana?
- Scarico cognitivo: quanto è stato in grado di gestire l'analisi manuale, il triage o il processo decisionale di routine l'agente per conto di un essere umano? Ha ridotto lo sforzo o lo ha semplicemente spostato?

Le aziende che adottano una mentalità che mette al primo posto le decisioni possono diventare più resilienti, adattive e in grado di operare a un nuovo livello di complessità.

Scalabilità mirata e mirata

Scalare con successo l'intelligenza artificiale agentica non significa sperimentare altri strumenti. Si tratta di creare un livello durevole di intelligenza aziendale. Ciò richiede investimenti nell'infrastruttura della piattaforma, nella cultura operativa, nei quadri di governance e nell'allineamento strategico. Le aziende devono adottare un approccio intenzionale. Devono trattare gli agenti non come esperimenti ma come componenti fondamentali del loro modello operativo digitale.

L'allineamento con [AWS Well-Architected Framework](#) aiuta i sistemi a soddisfare gli standard aziendali di affidabilità, sicurezza, efficienza delle prestazioni e ottimizzazione dei costi. Strumenti come [Strands Agents SDK](#) possono accelerare questo percorso fornendo prompt strutturati, registrazione degli strumenti e preparazione CI/CD. Questo aiuta i team a passare dalla sperimentazione alla distribuzione scalabile utilizzando flussi di lavoro familiari. AWS

L'intelligenza artificiale agentic non è uno strumento; è un cambiamento nel modo in cui l'intelligenza è incorporata nelle operazioni. Organizations che si preparano di conseguenza possono automatizzare di più, operare in modo più intelligente, adattarsi più velocemente e creare vantaggi duraturi in un mondo sempre più complesso.

Conclusione per rendere operativa l'IA agentica

L'intelligenza artificiale agentica rappresenta più di un cambiamento tecnologico. Segna l'emergere di un nuovo sistema operativo per l'azienda. Le organizzazioni che adottano questa trasformazione vanno oltre i ristretti casi d'uso dell'automazione e integrano l'intelligenza alla base delle loro operazioni. Questo cambiamento riguarda la riprogettazione del modo in cui vengono prese le decisioni, il modo in cui i sistemi si adattano e il modo in cui i risultati vengono realizzati su larga scala.

In un'epoca caratterizzata dalla crescente complessità, dalla domanda in tempo reale e dal sovraccarico di informazioni, il modello tradizionale di automazione basata su script ha raggiunto i suoi limiti. Il successo ora dipende dalla capacità di incorporare l'intelligenza direttamente nei flussi di lavoro per creare sistemi che percepiscono, ragionano, agiscono ed evolvono. L'intelligenza artificiale agentica può allineare l'autonomia allo scopo, il processo decisionale alla governance e l'adattabilità alla responsabilità.

Questa transizione richiede il passaggio da un approccio incentrato sull'esecuzione a un approccio incentrato sulle decisioni. I sistemi agentici non si limitano a seguire le istruzioni. Interpretano gli obiettivi, ponderano i compromessi e perseguono i risultati entro limiti definiti. In questo contesto, il successo non si misura solo in base al completamento delle attività. Viene misurato anche in base alla qualità, all'agilità e alla spiegabilità delle decisioni prese in tempo reale. Le organizzazioni devono ripensare le metriche, gli incentivi e la progettazione dei sistemi per supportare gli agenti che operano in modo intelligente in condizioni di incertezza.

Rendere operativa l'intelligenza artificiale agentica non è un aggiornamento. plug-and-play È una trasformazione architettonica e culturale. Richiede pratiche disciplinate in materia di gestione del ciclo di vita, applicazione della fiducia, interoperabilità e allineamento ai modelli di business. Richiede inoltre l'evoluzione dei modelli di distribuzione, come la definizione delle zone di intento, l'integrazione di barriere di runtime e l'allineamento continuo del comportamento degli agenti ai risultati strategici. I team devono adottare un linguaggio condiviso, una proprietà condivisa e una responsabilità condivisa per le prestazioni e la sicurezza degli agenti.

La preparazione aziendale può determinare chi prospera in questo nuovo ambiente. Le organizzazioni devono investire in strumenti interni, AgentOps funzionalità e framework di governance che scalino e creino valore a lungo termine. Chi ha successo può creare sistemi più intelligenti e può anche creare aziende più adattabili, resilienti e basate sulle informazioni.

Questa guida pone le basi. Collega la strategia all'esecuzione e prepara le organizzazioni a creare piattaforme scalabili di agenti intelligenti. La più ampia serie di contenuti sull'intelligenza artificiale agentica fornisce indicazioni complementari. AWS Per visualizzare le altre guide di questa serie, consulta [Agentic AI](#) sul sito Web Prescriptive Guidance. AWS Questa serie di contenuti offre una tabella di marcia per rendere operativa l'autonomia con disciplina e intento.

Per iniziare, identifica uno spazio decisionale ad alto impatto in cui gli agenti possano offrire miglioramenti misurabili in termini di velocità, precisione o reattività. Quindi schiera un agente pilota mirato che disponga di strumentazione, governance e circuiti di feedback. Utilizzatelo per convalidare l'ipotesi del valore, generare slancio interno e creare fiducia nell'approccio. Lo slancio aumenta attraverso l'apprendimento.

L'intelligenza artificiale agentica non è una destinazione; è un livello di capacità che si evolve insieme al business. Rappresenta un passaggio a lungo termine verso l'intelligenza come infrastruttura. Organizations leader in questo settore possono automatizzare di più, rispondere più rapidamente, adattarsi meglio e creare modelli operativi in grado di affrontare la complessità su scala aziendale.

Risorse per rendere operativa l'IA agentic

Servizi AWS

Le seguenti funzionalità possono aiutarti a creare Servizi AWS e rendere operativi sistemi di intelligenza artificiale agentic nei seguenti settori: Cloud AWS

- [Amazon API Gateway](#) può presentare le funzionalità degli agenti come scalabili e offre prezzi basati sull'utilizzo.
- [AWS AppConfig](#) offre la gestione della configurazione in fase di esecuzione e l'alternanza delle funzionalità per gli agenti tra tenant o ambienti.
- [Amazon Bedrock](#) è un servizio modello di base che gli agenti possono utilizzare per il ragionamento, la generazione e l'esecuzione tempestiva.
- [AWS Cloud Development Kit \(AWS CDK\)](#) è un servizio di infrastruttura come codice che puoi utilizzare per distribuire e gestire stack di agenti.
- [AWS CloudTrail](#) registra la cronologia degli eventi in modo da poter tenere traccia delle attività degli agenti, degli audit trail e dei comportamenti di integrazione.
- [Amazon CloudWatch](#) può gestire log, metriche e allarmi per monitorare le prestazioni degli agenti e il comportamento di collaborazione tra più agenti.
- [AWS CodePipeline](#) fornisce CI/CD l'automazione che puoi utilizzare per testare, convalidare e distribuire il codice dell'agente.
- [Amazon Cognito](#) è un servizio di identità che puoi utilizzare per gestire l'autenticazione di utenti e tenant in sistemi multiagente.
- [AWS Config](#) offre conformità e rilevamento delle deviazioni per la configurazione delle politiche degli agenti e dell'ambiente.
- [AWS Cost Explorer](#) può tracciare l'utilizzo a livello di agente e contribuire ad allineare i costi per massimizzare il ROI.
- [Amazon DynamoDB](#) è un servizio di storage che puoi utilizzare per la memoria degli agenti, i log di miglioramento e lo stato contestuale.
- [Amazon Elastic File System \(Amazon EFS\)](#) è un file system condiviso che puoi utilizzare per la collaborazione tra agenti o l'elaborazione intermedia tra flussi di lavoro.
- [Amazon EventBridge](#) è un bus di eventi principale che puoi utilizzare per indirizzare le attività e orchestrare le comunicazioni nella struttura degli agenti.

- [Amazon EventBridge Pipes](#) può semplificare l'inserimento e il routing degli eventi per connettere agenti e servizi.
- [Amazon GuardDuty](#) offre il rilevamento delle minacce e il monitoraggio delle anomalie in grado di supportare l'esecuzione sicura degli agenti.
- [AWS Identity and Access Management \(IAM\)](#) ti aiuta a definire autorizzazioni granulari per l'esecuzione degli agenti e l'accesso ai dati.
- [AWS Lambda](#) è un servizio di elaborazione stateless in grado di eseguire la logica degli agenti e i droni swarm.
- [Marketplace AWS](#) è una piattaforma di distribuzione esterna che è possibile utilizzare per offrire funzionalità agli agenti come prodotti commerciali.
- [AWS Organizations](#) è un servizio di governance e applicazione delle politiche interaccount che può aiutarti a gestire l'infrastruttura di agenti multi-tenant.
- [AWS Organizations le politiche di controllo dei servizi](#) fungono da barriere per il controllo delle autorizzazioni a livello di account o di unità organizzativa.
- [Amazon Quick](#) è una piattaforma generativa di business intelligence (BI) basata sull'intelligenza artificiale che ti aiuta ad analizzare dati, creare visualizzazioni, automatizzare i flussi di lavoro e collaborare con altri all'interno dell'organizzazione.
- [AWS Resource Access Manager \(AWS RAM\)](#) può aiutarti a condividere funzionalità tra account e servizi per agenti.
- [Amazon SageMaker AI](#) è un servizio che puoi utilizzare per la formazione dei modelli, la messa a punto e l'inferenza oltre ai modelli fondamentali.
- [Amazon Simple Storage Service \(Amazon S3\) Simple Storage Service \(Amazon S3\)](#) offre lo storage di oggetti per librerie di prompt, artefatti di modelli e dati generati da agenti.
- [AWS Step Functions](#) è un motore di flusso di lavoro che può aiutarti a coordinare i flussi multiagente e le pipeline di riqualificazione.
- [AWS X-Ray](#) offre un tracciamento distribuito che è possibile utilizzare per tenere traccia dei flussi decisionali degli agenti e delle dipendenze dei servizi.

Altre risorse 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 AWS](#)

- [Creazione di architetture serverless per l'intelligenza artificiale agentica su AWS](#)
- [Creazione di architetture multi-tenant per l'intelligenza artificiale agentica su AWS](#)

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	—	12 agosto 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 tuo database Oracle locale a Oracle su un'istanza EC2 in 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. Esegui la migrazione dei server da una piattaforma locale a un servizio cloud per la stessa piattaforma. Esempio: migra un'applicazione su Microsoft 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.

implementazione blu/verde

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 [Cloud Adoption AWS Framework](#).

implementazione canaria

Il rilascio lento e incrementale di una versione agli utenti finali. Quando sei sicuro, distribuisce 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 riproduzione 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

IA generativa

Un sottoinsieme di modelli di [intelligenza artificiale](#) che sono stati addestrati su grandi quantità di dati e che possono utilizzare un semplice messaggio 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 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 CSPM 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 blocco

[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

Vedi [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.

IIo/Internet 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 includono in genere 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 AWS con 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.

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 `true`. `false` `WHERE`

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 \(RACI\)](#).

RAG

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.](#)

Region

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 alle interruzioni o di ripristinarle. [L'elevata disponibilità e il disaster recovery](#) sono considerazioni comuni quando si pianifica la resilienza in Cloud AWS. [Per 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 [obiettivo del punto di ripristino](#).

VERSO

Vedi [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 Console di gestione AWS 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 Gestione dei segreti AWS, 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 Amazon EC2 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

tag

Coppie chiave-valore che fungono da metadati per l'organizzazione delle risorse. AWS Con i tag è possibile a gestire, identificare, organizzare, cercare e filtrare le risorse. Per ulteriori informazioni, consulta [Tagging delle 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.

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 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 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.