



Creazione di pipeline ML pronte per la produzione su AWS

AWS Guida prescrittiva



AWS Guida prescrittiva: Creazione di pipeline ML pronte per la produzione su AWS

Copyright © 2024 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 delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Introduzione	1
.....	1
.....	3
.....	4
Utilizzo di processi di elaborazione	4
Utilizzo della clausola di protezione principale	6
Test unitari	6
.....	8
Utilizzo dell'SDK Step Functions	8
Estensione dell'SDK Step Functions	9
.....	10
Implementazione conAWS CloudFormation	10
Modifica dell'output dall'SDK Step Functions	11
.....	12
.....	14
Diversi livelli di automazione	14
Piattaforme diverse per carichi di lavoro ML	15
Motori diversi per l'orchestrazione di condotte	15
.....	17
Risorse aggiuntive	18
Cronologia dei documenti	19
Glossario	20
#	20
A	21
B	24
C	25
D	28
E	32
F	34
G	35
H	36
I	37
L	40
M	41

O	44
P	47
Q	49
R	49
S	52
T	55
U	57
V	57
W	58
Z	59
.....	ix

Creazione di pipeline ML pronte per la produzione su AWS

Josiah Davis, Verdi March, Yin Song, Baichuan Sun, Chen Wu e Wei Yih Yap, Amazon Web Services (AWS)

Gennaio 2021([cronologia dei documenti](#))

I progetti di machine learning (ML) richiedono un impegno significativo in più fasi che include modellazione, implementazione e produzione per fornire valore aziendale e risolvere problemi del mondo reale. In ogni fase sono disponibili numerose alternative e opzioni di personalizzazione, che rendono sempre più difficile preparare un modello ML per la produzione entro i limiti delle risorse e del budget. Negli ultimi anni in Amazon Web Services (AWS), il nostro team di Data Science ha lavorato con diversi settori industriali su iniziative di apprendimento automatico. Abbiamo identificato i punti dolenti condivisi da molti AWS clienti, che derivano sia da problemi organizzativi che da sfide tecniche, e abbiamo sviluppato un approccio ottimale per fornire soluzioni ML pronte per la produzione.

Questa guida è destinata ai data scientist e agli ingegneri di ML coinvolti nelle implementazioni delle pipeline di ML. Descrive il nostro approccio per fornire pipeline ML pronte per la produzione. La guida illustra come passare dall'esecuzione di modelli ML in modo interattivo (durante lo sviluppo) alla distribuzione come parte di una pipeline (durante la produzione) per il tuo caso d'uso ML. A tal fine, abbiamo anche sviluppato una serie di modelli di esempio (vedi il [Progetto ML Max](#)), per accelerare la distribuzione di soluzioni ML personalizzate alla produzione, in modo da poter iniziare rapidamente senza dover fare troppe scelte di progettazione.

Panoramica

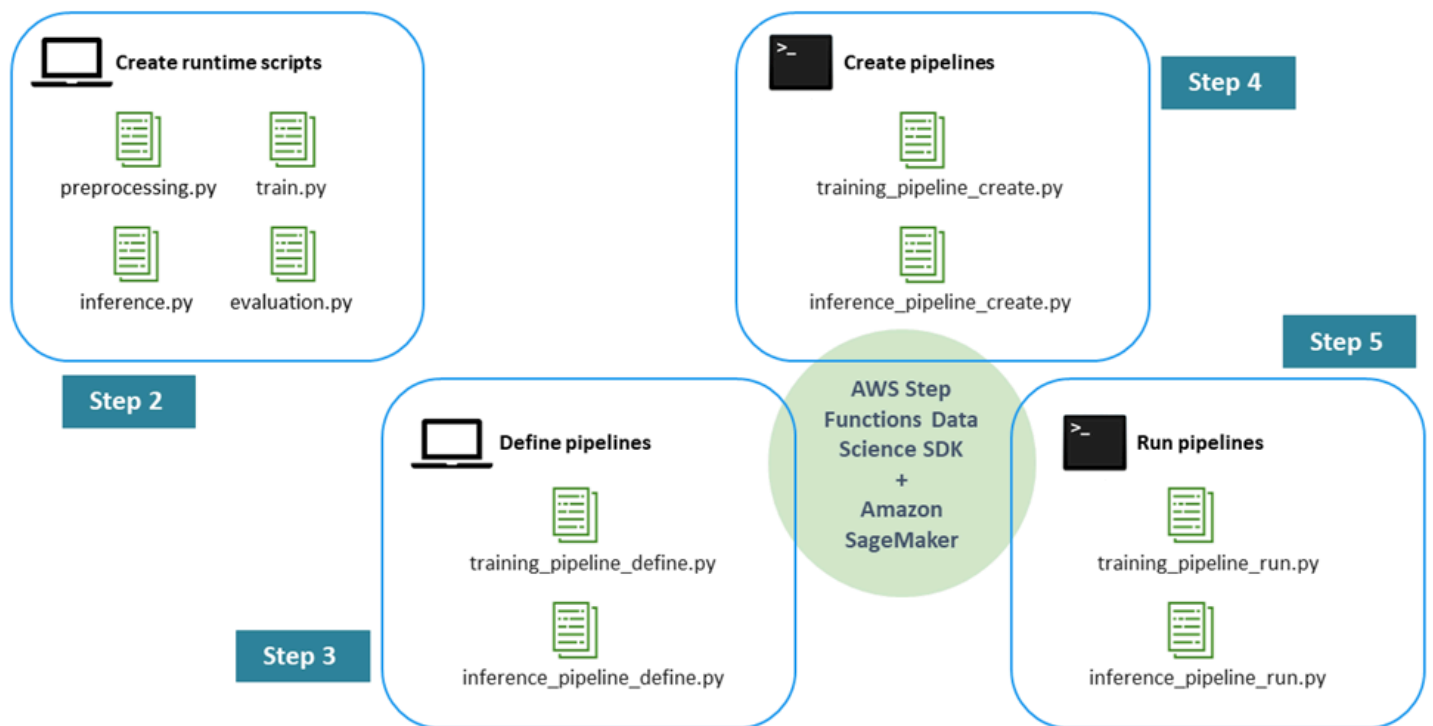
Il processo per creare una pipeline ML pronta per la produzione consiste nei seguenti passaggi:

- [Fase 1](#): Esegui EDA e sviluppa il modello iniziale— I data scientist rendono disponibili i dati grezzi in Amazon Simple Storage Service (Amazon S3), eseguono analisi esplorative dei dati (EDA), sviluppano il modello ML iniziale e ne valutano le prestazioni di inferenza. Puoi condurre queste attività in modo interattivo tramite i notebook Jupyter.
- [Fase 2](#). Creare gli script di runtime— Si integra il modello con script Python di runtime in modo che possa essere gestito e fornito da un framework ML (nel nostro caso, Amazon SageMaker). Questo è il primo passo per passare dallo sviluppo interattivo di un modello autonomo alla

produzione. In particolare, definisci separatamente la logica per la preelaborazione, la valutazione, l'addestramento e l'inferenza.

- **Fase 3.** Definisci la pipeline— Definite i segnaposti di input e output per ogni fase della pipeline. I relativi valori concreti verranno forniti in seguito, durante l'esecuzione (fase 5). Ti concentri sulle pipeline per la formazione, l'inferenza, la convalida incrociata e il back-test.
- **Fase 4.** Crea la pipeline— Crei l'infrastruttura sottostante, inclusa laAWS Step Functionsistanza della macchina a stati in modo automatizzato (quasi un clic), utilizzandoAWS CloudFormation.
- **Fase 5.** Esegui la pipeline— Si esegue la pipeline definita nel passaggio 4. Preparate anche i metadati e le posizioni dei dati o dei dati per inserire valori concreti per i segnaposti di input/output definiti nel passaggio 3. Ciò include gli script di runtime definiti nel passaggio 2 e gli iperparametri del modello.
- **Fase 6.** Espandi la pipeline— Implementate processi di integrazione continua e distribuzione continua (CI/CD), riqualificazione automatica, inferenza programmata ed estensioni simili della pipeline.

Il diagramma seguente illustra i passaggi principali.



Steps for creating the training and inference pipeline

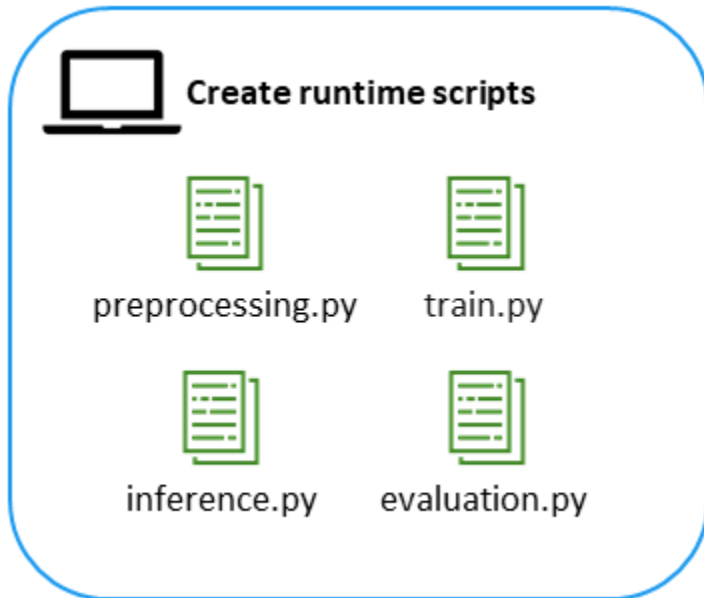
Fase 1. Esegui EDA e sviluppa il modello iniziale

In questo passaggio, i data scientist eseguono analisi esplorative dei dati (EDA) per comprendere il caso d'uso e i dati ML. Sviluppano quindi i modelli ML (ad esempio modelli di classificazione e regressione) per risolvere il problema in un determinato caso d'uso. Durante lo sviluppo del modello, il data scientist spesso fa ipotesi su input e output, come i formati di dati, il ciclo di vita dei dati e le posizioni dell'output intermedio. Queste ipotesi dovrebbero essere documentate in modo che possano essere utilizzate per la verifica durante i test unitari di cui al passaggio 2.

Anche se questo passaggio si concentra sullo sviluppo di modelli, i data scientist spesso devono scrivere una quantità minima di codice di supporto per la preelaborazione, la formazione, la valutazione e l'inferenza. Il data scientist dovrebbe essere in grado di eseguire questo codice nell'ambiente di sviluppo. Consigliamo inoltre di fornire argomenti di runtime opzionali in modo che questo codice helper possa essere configurato dinamicamente per l'esecuzione in altri ambienti senza modifiche manuali estese. Ciò accelererà l'integrazione tra il modello e la pipeline nei passaggi 2 e 3. Ad esempio, il codice per la lettura dei dati grezzi dovrebbe essere incapsulato in funzioni in modo che i dati possano essere preelaborati in modo coerente.

È consigliabile iniziare con un framework come [scikit-learn](#), [XGBoost](#), [PyTorch](#), [Keras](#), oppure [TensorFlow](#) per sviluppare il modello ML e il suo codice helper. Ad esempio, scikit-learn è una libreria ML gratuita scritta in Python. Fornisce una convenzione API uniforme per gli oggetti e include quattro oggetti principali: `estimatrice`, `predittore`, `trasformatore`, `emodello`—che coprono trasformazioni di dati leggere, supportare l'ingegneria di etichette e funzionalità e incapsulare le fasi di preelaborazione e modellazione. Questi oggetti aiutano a evitare la proliferazione del codice boilerplate e impediscono la perdita di dati di convalida e test nel set di dati di allenamento. Analogamente, ogni framework ML ha la sua implementazione di artefatti ML chiave e ti consigliamo di rispettare le convenzioni API del framework selezionato quando sviluppi modelli ML.

Fase 2. Creare gli script runtime



In questo passaggio, integrerai il modello sviluppato nel passaggio 1 e il relativo codice di supporto associato in una piattaforma ML per la formazione e l'inferenza pronte per la produzione. Nello specifico, ciò comporta lo sviluppo di script runtime in modo che il modello possa essere incorporato in SageMaker. Questi script Python standalone includono funzioni di callback SageMaker predefinite e variabili d'ambiente. Sono eseguiti all'interno di un contenitore SageMaker ospitato su un'istanza Amazon Elastic Compute Cloud (Amazon EC2). La [Documentazione di Amazon SageMaker Python SDK](#) fornisce informazioni dettagliate su come questi callback e la configurazione ausiliaria funzionano insieme per la formazione e l'inferenza. (Ad esempio, consulta [Prepara uno script di allenamento scikit-learn](#) [Implementa un modello scikit-learn](#) nella documentazione di SageMaker.) Le sezioni seguenti forniscono ulteriori suggerimenti per lo sviluppo di script di runtime ML, in base alla nostra esperienza di utilizzo AWS clienti.

Utilizzo di processi di elaborazione

SageMaker offre due opzioni per eseguire l'inferenza del modello in modalità batch. È possibile utilizzare un SageMaker lavoro di elaborazione o un processo di trasformazione in batch. Ciascuna opzione presenta vantaggi e svantaggi.

Un processo di elaborazione è costituito da un file Python che viene eseguito all'interno di un contenitore SageMaker. Il processo di elaborazione consiste in qualsiasi logica inserita nel file Python. Ha questi vantaggi:

- Quando comprendi la logica di base di un lavoro di formazione, i processi di elaborazione sono semplici da configurare e facili da comprendere. Condividono le stesse astrazioni dei lavori di formazione (ad esempio, l'ottimizzazione del conteggio delle istanze e della distribuzione dei dati).
- I data scientist e gli ingegneri ML hanno il pieno controllo sulle opzioni di manipolazione dei dati.
- Il data scientist non deve gestire alcuna logica dei componenti I/O tranne che per le familiari funzionalità di lettura/scrittura.
- È un po' più semplice eseguire i file in ambienti non Sagemaker, il che facilita lo sviluppo rapido e i test locali.
- Se si verifica un errore, un processo di elaborazione ha esito negativo non appena lo script ha esito negativo e non vi è attesa imprevista di tentativi.

D'altra parte, i processi di trasformazione batch sono un'estensione del concetto di endpoint SageMaker. In fase di runtime, questi processi importano funzioni di callback, che quindi gestiscono l'I/O per la lettura dei dati, il caricamento del modello e le previsioni. I processi di trasformazione Batch presentano i seguenti vantaggi:

- Usano un'astrazione di distribuzione dei dati che differisce dall'astrazione utilizzata dai lavori di formazione.
- Usano lo stesso file principale e la stessa struttura di funzioni sia per l'inferenza batch che per l'inferenza in tempo reale, il che è conveniente.
- Dispongono di un meccanismo di tolleranza ai guasti integrato basato su retrò. Ad esempio, se si verifica un errore su un batch di record, verrà riprovato più volte prima che il processo venga terminato come errore.

Grazie alla sua trasparenza, alla facilità d'uso in più ambienti e alla sua astrazione condivisa con i lavori di formazione, abbiamo deciso di utilizzare il processo di elaborazione anziché il processo di trasformazione batch nell'architettura di riferimento presentata in questa guida.

È necessario eseguire script runtime Python localmente prima di distribuirli nel cloud. In particolare, si consiglia di utilizzare la clausola di protezione principale quando si strutturano gli script Python e si eseguono test unitari.

Utilizzo della clausola di protezione principale

Utilizzare una clausola di protezione principale per supportare l'importazione dei moduli e per eseguire lo script Python. L'esecuzione di script Python singolarmente è utile per il debug e l'isolamento dei problemi nella pipeline ML. È consigliabile eseguire le operazioni seguenti:

- Utilizzare un parser di argomenti nei file di elaborazione Python per specificare i file di input/output e la loro posizione.
- Fornisci una guida principale e le funzioni di test per ogni file Python.
- Dopo aver testato un file Python, incorporarlo nelle diverse fasi della pipeline ML, indipendentemente dal fatto che si stia utilizzando un file PythonAWS Step Functionsmodello o processo di elaborazione SageMaker.
- UtilizzaAsserzioneistruzioni nelle sezioni critiche dello script per facilitare il test e il debug. Puoi ad esempio usare unAsserzioneistruzione per garantire che il numero di funzioni del set di dati sia coerente dopo il caricamento.

Test unitari

Il test unitario degli script runtime scritti per la pipeline è un compito importante che viene spesso ignorato nello sviluppo della pipeline ML. Questo perché l'apprendimento automatico e la scienza dei dati sono campi relativamente nuovi e sono stati lenti ad adottare pratiche di ingegneria software consolidate come il test unitario. Poiché la pipeline ML verrà utilizzata nell'ambiente di produzione, è essenziale testare il codice della pipeline prima di applicare il modello ML alle applicazioni reali.

Il test unitario dello script runtime fornisce inoltre i seguenti vantaggi esclusivi per i modelli ML:

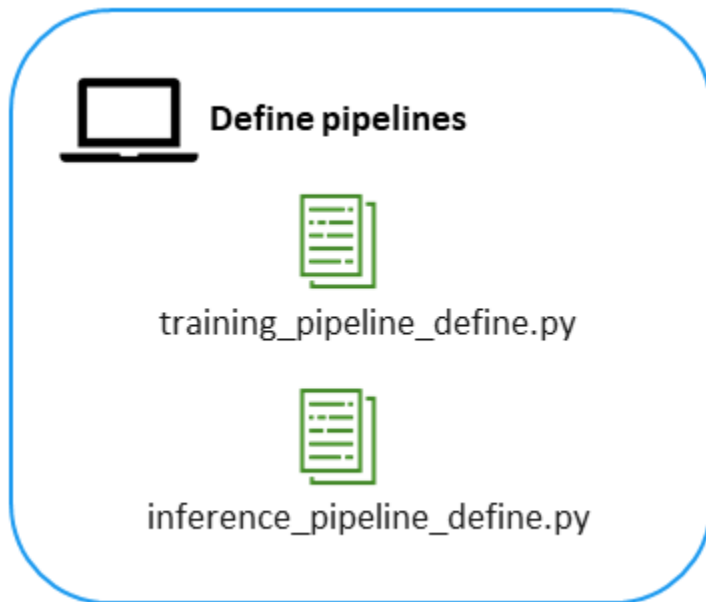
- Previene trasformazioni impreviste di dati. La maggior parte delle pipeline ML coinvolge molte trasformazioni di dati, quindi è fondamentale che queste trasformazioni funzionino come previsto.
- Convalida la riproducibilità del codice. Qualsiasi casualità nel codice può essere rilevata mediante test unitari con diversi casi d'uso.
- Applica la modularità del codice. I test unitari sono solitamente associati alla misura di copertura del test, ovvero il grado in cui una particolare suite di test (una raccolta di casi di test) esegue il codice sorgente di un programma. Per ottenere un'elevata copertura dei test, gli sviluppatori modularizzano il codice, perché è difficile scrivere test unitari per una grande quantità di codice senza suddividerlo in funzioni o classi.
- Impedisce l'introduzione di codici o errori di bassa qualità nella produzione.

È consigliabile utilizzare un framework maturo di test unitario, ad esempio [pytest](#) per scrivere i casi di test unitari, perché è più facile gestire test unitari estesi all'interno di un framework.

 Important

Il test unitario non garantisce che tutte le custodie angolari siano testate, ma può aiutarti a evitare errori in modo proattivo prima di distribuire il modello. Si consiglia inoltre di monitorare il modello dopo l'installazione, per garantire l'eccellenza operativa.

Fase 3. Definisci la pipeline



In questo passaggio vengono definite la sequenza e la logica delle azioni che la pipeline eseguirà. Ciò include passaggi discreti, così come i loro ingressi e uscite logici. Ad esempio, qual è lo stato dei dati all'inizio della pipeline? Proviene da più file che si trovano a diversi livelli di granularità o da un singolo file flat? Se i dati provengono da più file, è necessario un solo passaggio per tutti i file o passaggi separati per ciascun file per definire la logica di preelaborazione? La decisione dipende dalla complessità delle origini dati e dalla misura in cui vengono preelaborate.

Nella nostra implementazione di riferimento, utilizziamo [AWS Step Functions](#), che è un orchestratore di funzioni serverless, per definire i passaggi del flusso di lavoro. Tuttavia, il [Quadro ML Max](#) supporta anche altri sistemi di tubazioni o macchine statali come Apache AirFlow (vedere il [Motori diversi per l'orchestrazione di condotte](#) sezione) per favorire lo sviluppo e l'implementazione di pipeline ML.

Utilizzo dell'SDK Step Functions

Per definire la pipeline ML, utilizziamo innanzitutto l'API Python di alto livello fornita dal [AWS Step Functions Data Science SDK](#) (Step Functions SDK) per definire due componenti chiave della pipeline: `pass` e `task`. Se si pensa a una pipeline come a un grafico aciclico diretto (DAG), i passaggi rappresentano i nodi sul grafico e i dati vengono visualizzati come bordi diretti che collegano un nodo (passo) al successivo. Esempi tipici di passaggi ML includono la preelaborazione, la formazione e la valutazione. L'SDK Step Functions fornisce una serie di passaggi integrati (ad esempio [Fase di](#)

[formazione](#)) che puoi utilizzare. Esempi di dati includono input, output e molti set di dati intermedi prodotti da alcuni passaggi nella pipeline. Quando si progetta una pipeline ML, non si conoscono i valori concreti degli elementi di dati. È possibile definire segnaposti dei dati che fungono da modello (simile ai parametri di funzione) e contengono solo il nome dell'elemento di dati e i tipi di dati primitivi. In questo modo, è possibile progettare un progetto completo della pipeline senza conoscere in anticipo i valori concreti dei dati che viaggiano sul grafico. A tale scopo, è possibile utilizzare le classi segnaposto nell'SDK Step Functions per modellare esplicitamente questi modelli di dati.

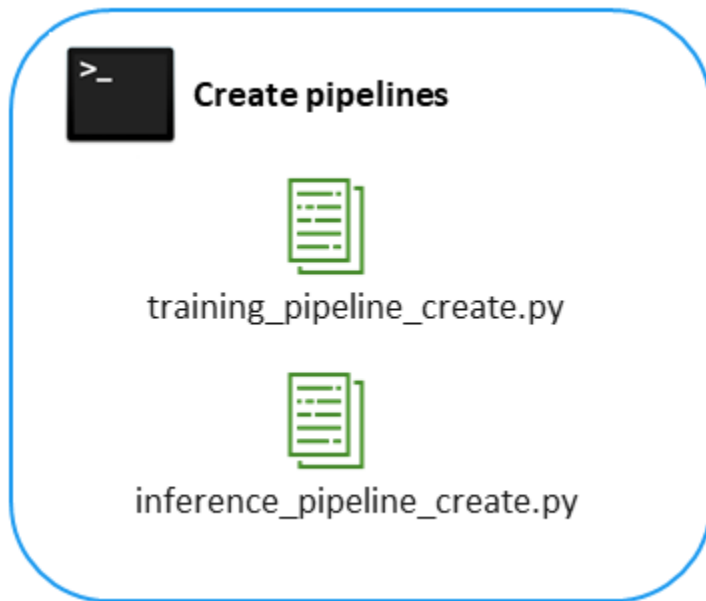
Le tubazioni ML richiedono inoltre parametri di configurazione per eseguire un controllo a grana fine sul comportamento di ogni fase ML. Questi segnaposti per dati speciali sono chiamati segnaposti dei parametri. Molti dei loro valori sono sconosciuti quando si definisce la pipeline. Esempi di segnaposti dei parametri includono i parametri correlati all'infrastruttura definiti durante la progettazione della pipeline (ad esempio, l'URL dell'immagine di AWS Region o contenitore) e i parametri relativi alla modellazione ML (ad esempio i parametri) definiti quando si esegue la pipeline.

Estensione dell'SDK Step Functions

Nella nostra implementazione di riferimento, un requisito era quello di separare le definizioni delle tubazioni ML dalla creazione e distribuzione di pipeline ML concrete utilizzando impostazioni specifiche dei parametri. Tuttavia, alcuni passaggi integrati nell'SDK Step Functions non ci hanno permesso di passare tutti questi parametri segnaposto. Invece, i valori dei parametri dovevano essere ottenuti direttamente durante il tempo di progettazione della pipeline tramite le chiamate API di configurazione SageMaker. Funziona bene se l'ambiente di progettazione SageMaker è identico all'ambiente runtime di SageMaker, ma raramente accade nelle impostazioni reali. Un accoppiamento così stretto tra il tempo di progettazione della pipeline e il runtime, e il presupposto che l'infrastruttura della piattaforma ML rimanga costante, ostacola notevolmente l'applicabilità della pipeline progettata. Infatti, la pipeline ML si interrompe immediatamente quando la piattaforma di distribuzione sottostante subisce anche il minimo cambiamento.

Per superare questa sfida e produrre una robusta pipeline ML (che volevamo progettare una volta ed eseguire ovunque), abbiamo implementato i nostri passaggi personalizzati estendendo alcuni dei passaggi integrati, inclusi Fase di formazione, Modello Step, e Passaggio trasformatore. Queste estensioni sono fornite nella [Progetto ML Max](#). Le interfacce di questi passaggi personalizzati supportano molti più segnaposti di parametri che possono essere popolati con valori concreti durante la creazione della pipeline o quando la pipeline è in esecuzione.

Fase 4. Per creare la pipeline



Dopo aver definito logicamente la pipeline, è ora di creare l'infrastruttura per supportare la pipeline. Questo passaggio richiede almeno le seguenti funzionalità:

- Storage, per ospitare e gestire input e output della pipeline, inclusi codice, artifact del modello e dati utilizzati nelle esecuzioni di formazione e inferenza.
- Calcolo (GPU o CPU), per la modellazione e l'inferenza, nonché la preelaborazione e la postelaborazione dei dati.
- Orchestrazione, per gestire le risorse utilizzate e per pianificare qualsiasi esecuzione regolare. Ad esempio, il modello potrebbe essere riaddestrato periodicamente man mano che i nuovi dati diventano disponibili.
- Registrazione e avvisi, per monitorare l'accuratezza del modello di pipeline, per l'utilizzo delle risorse e per la risoluzione dei problemi.

Implementazione con AWS CloudFormation

Per creare la pipeline AWS CloudFormation, che è un AWS Servizio per la distribuzione e la gestione di infrastrutture come codice. La AWS CloudFormationi modelli includono la definizione Step Functions creata nel passaggio precedente con l'SDK Step Functions. Questo passaggio include la creazione dell'istanza di Step Functions gestita da AWS, denominata Step Functions a stati. In questa

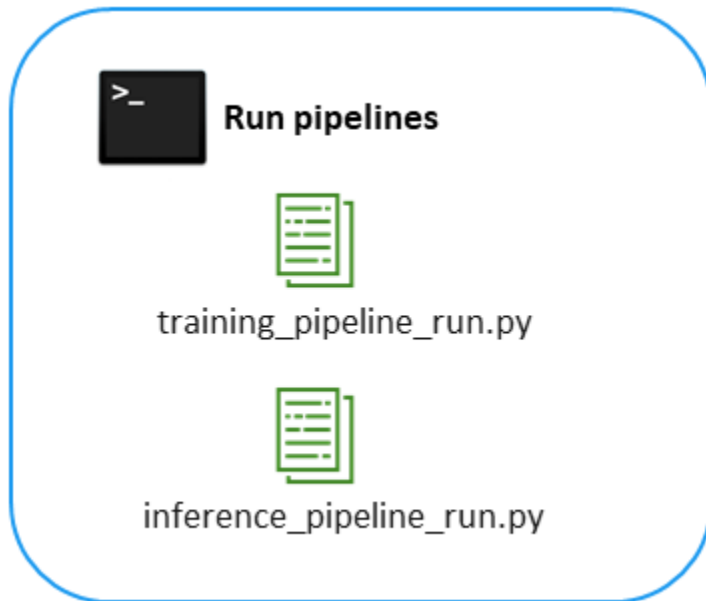
fase non vengono create risorse per la formazione e l'inferenza, perché i processi di formazione e inferenza vengono eseguiti su richiesta, solo quando sono necessari, come lavori di SageMaker. Questo passaggio include anche la creazione di AWS Identity and Access Management (IAM) ruoli per eseguire Step Functions, eseguire SageMaker e leggere e scrivere da Amazon S3.

Modifica dell'output dall'SDK Step Functions

Abbiamo dovuto apportare alcune piccole modifiche all'output dell'AWS CloudFormation della sezione precedente. Abbiamo usato una semplice corrispondenza di stringhe Python per fare quanto segue:

- Abbiamo aggiunto la logica per creare il `Parameters` sezione della AWS CloudFormation template. Questo perché vogliamo creare due ruoli e definire il nome della pipeline come parametro insieme all'ambiente di distribuzione. Questo passaggio copre anche tutte le risorse e i ruoli aggiuntivi che potresti voler creare, come illustrato nel passaggio 6.
- Abbiamo riformattato tre campi per avere il necessario! Subprefisso e virgolette in modo che possano essere aggiornati dinamicamente come parte del processo di distribuzione:
 - `LaStateMachineName` proprietà, che nomina la macchina a stati.
 - `LaDefinitionString` proprietà, che definisce la macchina a stati.
 - `LaRoleArn` proprietà, che viene restituita dalla macchina statale.

Fase 5: Esegui la pipeline



Questo passaggio esegue la pipeline di addestramento o inferenza creata nell'AWS CloudFormation si imitano nella fase 4. La pipeline non può essere eseguita fino a quando i parametri del segnaposto interno non sono stati compilati con valori concreti. Questa azione di assegnazione di valori ai parametri segnaposto è l'attività principale del passaggio 5. I parametri segnaposto di esempio includono:

- La posizione dei set di dati di input, output e intermedi
- La posizione di Amazon S3 degli script di runtime e altri codici di preelaborazione o valutazione sviluppati nel passaggio 2 (ad esempio, `sm_submit_url` per la pipeline di formazione)
- Il nome della `AWSRegion`

È necessario assicurarsi che questi valori di percorso puntino a dati o codice validi prima di eseguire la pipeline. Ad esempio, se si inserisce il parametro segnaposto che rappresenta l'URL Amazon S3 degli script runtime Python, è necessario caricare tali script su tale URL. La persona che gestisce la pipeline è responsabile del controllo della coerenza e del caricamento dei dati. Le persone che definiscono o creano la pipeline non devono preoccuparsi di tutto questo.

A seconda della scadenza della pipeline, questo passaggio può essere automatizzato per funzionare regolarmente (settimanalmente o mensilmente). L'automazione richiede anche un monitoraggio solido, che è un'area importante ma non rientra nell'ambito di questa guida. Per l'esecuzione della

pipeline di formazione, sarebbe opportuno monitorare le metriche di valutazione. Per la pipeline di inferenza, sarebbe opportuno monitorare la deriva della distribuzione dei dati di input e, se possibile, raccogliere periodicamente le etichette e misurare la deriva nell'accuratezza delle previsioni. Questi record provenienti dalle esecuzioni di addestramento e inferenza devono essere registrati in un database per l'analisi in un secondo momento.

Fase 6: Espandere la pipeline

Questa guida spiega come iniziare a costruire pipeline ML suAWSrapidamente, con architettura concreta. Esistono ulteriori considerazioni sulla maturazione della pipeline, come la gestione dei metadati, il monitoraggio degli esperimenti e il monitoraggio. Si tratta di argomenti importanti che non rientrano nell'ambito di questa guida. Le sezioni seguenti discutono un altro aspetto della gestione delle pipeline, ovvero l'automazione delle pipeline.

Diversi livelli di automazione

Sebbene sia possibile impostare manualmente una pipeline di formazione nella console di SageMaker, in pratica si consiglia di ridurre al minimo i punti di contatto manuali nella distribuzione delle pipeline di formazione ML per garantire che i modelli ML vengano distribuiti in modo coerente e ripetuto. A seconda delle esigenze e dei problemi aziendali che stai affrontando, puoi determinare e implementare una strategia di implementazione su tre livelli: semi-automatica, completamente automatizzata e completamente gestita.

- **Semi-automatico:** per impostazione predefinita, i passaggi descritti nella sezione precedente seguono un approccio semi-automatico, poiché implementano la pipeline di formazione e inferenza utilizzandoAWS CloudFormationmodelli. Ciò consente di garantire la riproducibilità della pipeline e consente di modificarla e aggiornarla facilmente.
- **Completamente automatizzato:** un'opzione più avanzata consiste nell'utilizzare l'integrazione e l'implementazione continua (CI/CD) negli ambienti di sviluppo, staging e produzione. L'integrazione di pratiche CI/CD nell'implementazione della pipeline di formazione può garantire che l'automazione includa la tracciabilità e i cancelli di qualità.
- **Completamente gestito:** in definitiva, è possibile sviluppare un sistema completamente gestito in modo da poter distribuire una pipeline di formazione ML con una serie di semplici manifesti e il sistema può autoconfigurare e coordinare le richiesteAWSServizi .

In questa guida abbiamo scelto di presentare un'architettura concreta. Tuttavia, ci sono tecnologie alternative che puoi considerare. Le due sezioni successive discutono di alcune scelte alternative per la piattaforma e il motore di orchestrazione.

Piattaforme diverse per carichi di lavoro ML

[Amazon SageMaker](#) è il AWS servizio gestito per la formazione e la fornitura di modelli ML. Molti utenti apprezzano la sua vasta gamma di funzionalità integrate e le numerose opzioni offerte per l'esecuzione di carichi di lavoro ML. SageMaker è particolarmente utile se hai appena iniziato a implementare ML nel cloud. Le caratteristiche principali di SageMaker includono:

- Tracciabilità integrata (inclusi etichettatura, formazione, tracciamento dei modelli, ottimizzazione e inferenza).
- Opzioni integrate con un clic per l'addestramento e l'inferenza con un'esperienza minima in Python e ML.
- Sintonizzazione iperparametrica avanzata.
- Supporto per tutti i principali framework di intelligenza artificiale e apprendimento automatico (ML/AI) e contenitori Docker personalizzati.
- Funzionalità di monitoraggio integrate.
- Tracciamento integrato delle cronologie, inclusi processi di formazione, processi di elaborazione, processi di trasformazione batch, modelli, endpoint e ricercabilità. Alcune cronologie, come la formazione, l'elaborazione e la trasformazione in batch, sono immutabili e di sola aggiunta.

Una delle alternative all'utilizzo di SageMaker è [AWS Batch](#). AWS Batch fornisce un livello inferiore di controllo sull'elaborazione e l'orchestrazione per il tuo ambiente, ma non è progettato su misura per il machine learning. Alcune delle sue caratteristiche principali includono:

- Scalabilità automatica pronta all'uso delle risorse di elaborazione in base al carico di lavoro.
- Supporto pronto all'uso per priorità del lavoro, tentativi e dipendenze di lavoro.
- Approccio basato sulla coda che supporta la creazione di processi ricorrenti e on-demand.
- Supporto per carichi di lavoro CPU e GPU. La possibilità di utilizzare la GPU per la creazione di modelli ML è fondamentale, poiché la GPU può velocizzare notevolmente il processo di formazione, soprattutto per i modelli di deep learning.
- Possibilità di definire una personalizzazione [Amazon Machine Image](#) (AMI) per l'ambiente di calcolo.

Motori diversi per l'orchestrazione di condotte

Il secondo componente principale è il livello di orchestrazione della pipeline. AWS fornisce [Step Functions](#) per un'esperienza di orchestrazione completamente gestita. Un'alternativa popolare alle

Step Functions è Apache Airflow. Quando prendi una decisione tra i due aspetti, considera i seguenti aspetti:

- **Infrastruttura richiesta**— AWS Step Functions è un servizio completamente gestito ed è senza server, mentre Airflow richiede la gestione della propria infrastruttura e si basa su software open source. Di conseguenza, Step Functions fornisce un'elevata disponibilità pronta all'uso, mentre l'amministrazione di Apache Airflow richiede ulteriori passaggi.
- **Funzionalità di pianificazione**: sia Step Functions che Airflow forniscono funzionalità comparabili.
- **Funzionalità di visualizzazione e interfaccia utente**: sia Step Functions che Airflow forniscono funzionalità comparabili.
- **Passaggio di variabili all'interno del grafico computazionale** — Step Functions fornisce funzionalità limitate per l'utilizzo di funzioni AWS Lambda, mentre Airflow fornisce interfacce XCom.
- **Utilizzo** — Step Functions è molto popolare tra i clienti AWS e Airflow è stato ampiamente adottato dalla comunità di data engineering.

Conclusioni

Mentre il machine learning passa da una disciplina di ricerca a un campo applicato, abbiamo assistito a una crescita annuale del 25% nello sviluppo, nella distribuzione e nel funzionamento delle pipeline ML in vari settori. Il valore aziendale di ML è realizzato attraverso operazioni e pipeline ML quotidiane che, a loro volta, guidano la ricerca e lo sviluppo di modelli e algoritmi ML. Tuttavia, l'implementazione di ML in produzione presenta numerose sfide, perché intreccia attività e artefatti significativamente diversi, come la gestione dei dati, l'elaborazione, l'analisi, la modellazione, la verifica e la sicurezza. Attraverso numerosi impegni AI/ML con AWS clienti, il nostro team di Data Science ha osservato che una sfida fondamentale è la mancanza di un flusso di lavoro end-to-end che fornisca una serie di modelli per fondere o separare in modo ottimale diverse attività e artefatti ML DevOps. In questa guida, abbiamo presentato il [Flusso di lavoro ML Max](#) per risolvere questo problema urgente. ML Max fornisce linee guida dettagliate e una serie di modelli di programmazione. L'obiettivo è quello di consentire una transizione rapida ed economica da una fase di sviluppo interattiva del modello a una configurazione completa e scalabile della pipeline ML pronta per la produzione.

Risorse aggiuntive

Guide e modelli correlati

- [Quantificare l'incertezza nei sistemi di deep learning](#)
- [Migrazione dei carichi di lavoro di compilazione, addestramento e distribuzione di ML ad Amazon SageMaker utilizzando AWS Strumenti per developer](#)
- [AWS Sito web di Prescriptive Guidance](#)

AWS servizi

- [AWS Batch](#)
- [AWS CloudFormation](#)
- [IAM](#)
- [Amazon SageMaker](#)
- [AWS Step Functions](#)

AWS risorse generali

- [AWS documentazione](#)
- [AWS riferimento generale](#)
- [AWS Glossario](#)

Cronologia dei documenti

La tabella seguente descrive le modifiche importanti apportate a questa guida. Se desideri ricevere una notifica sugli aggiornamenti future, puoi sottoscrivere un [Avanzamento RSS](#).

Modifica	Descrizione	Data
Pubblicazione iniziale	—	29 gennaio 2021

Glossario del Prontuario AWS

I seguenti termini sono comunemente utilizzati in strategie, guide e pattern forniti dal Prontuario AWS. 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: esegui la migrazione del database Oracle on-premise ad Amazon Aurora edizione compatibile con 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: esegui la migrazione del database Oracle on-premise ad Amazon Relational Database Service (Amazon RDS) per Oracle nel 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: esegui la migrazione del 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: esegui la migrazione del tuo database Oracle on-premise su Oracle su un'istanza EC2 nel 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. Questo scenario di migrazione è specifico di VMware Cloud su AWS, che supporta la compatibilità delle macchine virtuali (VM) e la portabilità del carico di lavoro tra l'ambiente on-premise e AWS. È possibile utilizzare le tecnologie VMware Cloud Foundation dai data center on-premise durante la migrazione dell'infrastruttura a VMware Cloud su AWS. Esempio: trasferisci l'hypervisor che ospita il database Oracle su VMware Cloud su 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 effettuare 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 di 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 aggregata

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 di frequente 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 viene utilizzato AIOps nella strategia di migrazione AWS, 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 per AWS](#) nella documentazione di 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

Posizione separata all'interno di una Regione AWS isolata dagli errori che si verificano in altre zone di disponibilità che offre connettività di rete non costosa e a bassa latenza ad altre zone di disponibilità nella stessa regione.

Framework per l'adozione del cloud AWS (AWS CAF)

Un framework di linee guida e buone pratiche di 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: azienda, 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. Per questa prospettiva, AWS CAF fornisce linee guida per lo sviluppo del personale, la formazione e le comunicazioni per aiutare l'organizzazione nell'adozione efficace del cloud. Per ulteriori informazioni, consulta il [sito web di AWS CAF](#) e il [white paper AWS CAF](#).

AWS Workload Qualification Framework (AWS WQF)

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

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.

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, vedere [About branch](#) (GitHub documentazione).

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-ArchitectedAWS.

strategia brownfield

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

cache del buffer

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

capacità di business

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

pianificazione della continuità operativa (BCP)

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

C

CAF

Vedi [AWS Cloud Adoption Framework](#).

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 in locale, prima che vengano ricevuti dal Servizio AWS di destinazione.

centro di eccellenza del 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 i [post sul CCoE](#) sul blog AWS Cloud Enterprise Strategy.

cloud computing

La tecnologia cloud generalmente utilizzata per l'archiviazione remota di dati e la gestione dei dispositivi IoT. Il cloud computing è comunemente 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 il cloud AWS:

- Progetto: esecuzione di alcuni progetti relativi al cloud per scopi di dimostrazione e apprendimento
- Fondamento: effettuare investimenti fondamentali per dimensionare l'adozione del cloud (ad esempio, creazione di una zona di destinazione, definizione di un CCoE, 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 del blog [The Journey Toward Cloud-First & the Stages of Adoption](#) sul blog AWS Cloud Enterprise Strategy. Per informazioni su come si relazionano alla strategia di migrazione AWS, consulta la [guida di 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 o AWS CodeCommit. 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

Un campo dell'intelligenza artificiale utilizzato dalle macchine per identificare persone, luoghi e cose nelle immagini con una precisione pari o superiore ai livelli umani. Spesso costruito con modelli di deep learning, automatizza l'estrazione, l'analisi, la classificazione e la comprensione di informazioni utili da una singola immagine o da una sequenza di immagini.

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 serie di regole di AWS Config e azioni correttive che puoi riunire per personalizzare i controlli di conformità e sicurezza. Puoi distribuire un pacchetto di conformità come singola entità in un Account AWS e in una regione, o all'interno di un'organizzazione, utilizzando un modello YAML. Per ulteriori informazioni, consulta [Pacchetti di conformità](#) nella documentazione di AWS Config.

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

Il processo di automazione delle fasi di origine, creazione, test, gestione temporanea e produzione del processo di rilascio del software. Il processo CI/CD è comunemente descritto come una pipeline. CI/CD può aiutare ad automatizzare i processi, migliorare la produttività, migliorare la qualità del codice e velocizzare le distribuzioni. 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](#).

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 Framework AWS Well-Architected. 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.

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'AWSambiente 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 adotti questa strategia in AWS, puoi aggiungere più controlli a diversi livelli della struttura AWS Organizations 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 membro di AWS 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 suAWS: 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](#).

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.

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 le autorizzazioni ad altri Account AWS o ai principali AWS Identity and Access Management (IAM). 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).

crittografia envelope

Il processo di crittografia di una chiave di crittografia con un'altra chiave di crittografia. Per ulteriori informazioni, consulta [Crittografia a busta](#) nella documentazione di 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 pipeline CI/CD, l'ambiente di produzione è l'ultimo ambiente di implementazione.
- ambienti superiori: tutti gli ambienti a cui possono accedere utenti diversi dal team di sviluppo principale. Si può trattare di un ambiente di produzione, ambienti di preproduzione e ambienti per i test di accettazione da parte degli utenti.

epica

Nelle metodologie agili, categorie funzionali che aiutano a organizzare e dare priorità al lavoro. Le epiche forniscono una descrizione di alto livello dei requisiti e delle attività di implementazione. Ad esempio le epiche di 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](#).

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

NelCloud AWS, un limite come una zona di disponibilitàRegione 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 [AWSFault](#) 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, vedere [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.

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.

G

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](#).

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 livello elevato che consente di governare risorse, policy e conformità tra le unità organizzative (OU). 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 GuardDutyAWS Trusted Advisor, Amazon Inspector e controlli personalizzatiAWS 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.

migrazione di database omogenea

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

dati caldi

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

hotfix

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

periodo di hypercare

Subito dopo la conversione, il periodo di tempo in cui un team di migrazione gestisce e monitora le applicazioni migrate nel cloud per risolvere eventuali problemi. In genere, questo periodo dura da 1 a 4 giorni. Al termine del periodo di hypercare, il team addetto alla migrazione in genere trasferisce la responsabilità delle applicazioni al team addetto alle operazioni cloud.

I

IaC

Considera [l'infrastruttura come codice](#).

Policy basata su identità

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

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 multi-account AWS, un VPC che accetta, ispeziona e instrada le connessioni di rete dall'esterno di un'applicazione. Nel documento [Architettura di riferimento per la sicurezza di AWS](#) si consiglia di configurare l'account di rete con VPC in entrata, in uscita e di ispezione 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.

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.

Internet delle cose industriale (IIoT)

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, consulta [Creazione di una strategia di trasformazione digitale dell'Internet delle cose industriale \(IIoT\)](#).

VPC di ispezione

In un'architettura multi-account AWS, un VPC centralizzato che gestisce le ispezioni del traffico di rete tra VPC (in Regioni AWS uguali o diverse), Internet e le reti on-premise. Nel documento [Architettura di riferimento per la sicurezza di AWS](#) si consiglia di configurare l'account di rete con VPC in entrata, in uscita e di ispezione 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, consulta la sezione [Interpretabilità dei modelli 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 zona di destinazione è un ambiente AWS 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](#).

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](#).

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](#).

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.

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 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.

microservizio

Un piccolo servizio indipendente che comunica tramite API ben definite ed è in genere di proprietà di piccoli team autonomi. 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 la sezione [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 tramite un'interfaccia ben definita utilizzando API leggere. Ogni microservizio in questa architettura può essere aggiornato, distribuito e dimensionato per soddisfare la richiesta di funzioni specifiche di un'applicazione. Per ulteriori informazioni, consulta la sezione [Implementazione di microservizi su AWS](#).

Programma di accelerazione della migrazione (MAP)

Un programma AWS che offre consulenza, 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 negli 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 di destinazione, il gruppo di sicurezza e l'account 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: eseguire il rehosting della migrazione ad Amazon EC2 con AWS Application Migration Service.

Valutazione del portfolio di migrazione (MPA)

Uno strumento online che fornisce informazioni per la convalida del business case per la migrazione al 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 il login) è disponibile gratuitamente per tutti i consulenti AWS e i consulenti partner APN.

valutazione della preparazione alla migrazione (MRA)

Il processo di acquisizione di informazioni sullo stato di idoneità 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 AWS CAF. 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 eseguire la migrazione di un carico di lavoro verso il 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.](#)

MAPPA

Vedi [Migration Portfolio Assessment](#).

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, consulta la sezione [Strategia per modernizzare le applicazioni nel 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, consulta la sezione [Valutazione della preparazione alla modernizzazione per le applicazioni nel 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](#).

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.

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.

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 trail creato da AWS CloudTrail che registra tutti gli eventi per tutti gli Account AWS in un'organizzazione di 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, vedere [Creazione di un percorso per un'organizzazione](#) nella documentazione. CloudTrail

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 migrazione AWS, 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 tutte le Regioni AWS, crittografia lato server con AWS KMS (SSE-KMS) e richieste PUT e DELETE dinamiche al bucket S3.

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.

O

Vedi la revisione della [prontezza operativa](#).

VPC in uscita (egress)

In un'architettura multi-account AWS, un VPC che gestisce le connessioni di rete avviate dall'interno di un'applicazione. Nel documento [Architettura di riferimento per la sicurezza di AWS](#) si consiglia di configurare l'account di rete con VPC in entrata, in uscita e di ispezione 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.

policy

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

persistenza poliglotta

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

valutazione del portfolio

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

predicate

Una condizione di interrogazione che restituisce o, in genere, si trova in una clausola `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 che può eseguire operazioni e accedere alle risorse. Questa entità è in genere un utente root per un Account AWS, un 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 all'ingegneria dei sistemi che tiene conto della privacy durante l'intero processo di progettazione.

zone ospitate private

Un container che contiene informazioni su come si desidera che Amazon Route 53 risponda alle query DNS per un dominio e i relativi sottodomini all'interno di uno o più VPC. 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

Ambiente di produzione

Vedi [ambiente](#).

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.

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).

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. Ciò determina quella che viene considerata una perdita di dati accettabile tra l'ultimo punto di ripristino e l'interruzione del servizio.

obiettivo del tempo di ripristino (RTO)

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

rifattorizzare

Vedi [7 R.](#)

Regione

Una raccolta di risorse AWS in un'area geografica. Ogni Regione AWS è isolata e indipendente dalle altre per fornire tolleranza agli errori, stabilità e resilienza. Per ulteriori informazioni, consulta [Gestione delle Regioni AWS](#) nei Riferimenti generali di AWS.

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.](#)

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.](#)

rotazione

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

controllo dell'accesso a righe e colonne (RCAC)

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

RPO

Vedi l'obiettivo del punto [di ripristino](#).

RTO

Vedi l'[obiettivo del tempo di ripristino](#).

runbook

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

S

SAML 2.0

Uno standard aperto utilizzato da molti provider di identità (IdPs). Questa funzionalità consente l'autenticazione unica (SSO) federata, grazie alla quale gli utenti possono accedere alla AWS Management Console o eseguire chiamate alle operazioni delle API AWS. In questo modo non è necessario creare un utente IAM per tutti gli utenti nell'organizzazione. Per ulteriori informazioni sulla federazione basata su SAML 2.0, consulta [Informazioni sulla federazione basata su SAML 2.0](#) nella documentazione di IAM.

SCP

Vedi la [politica di controllo del servizio](#).

Secret

In AWS Secrets Manager, informazioni riservate o riservate, come una password o le credenziali utente, archiviate in forma crittografata. È costituito dal valore segreto e dai relativi metadati. Il valore segreto può essere binario, una stringa singola o più stringhe. Per ulteriori informazioni, [consulta Secret](#) nella documentazione di Secrets Manager.

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 del Servizio AWS che li riceve.

Policy di controllo dei servizi (SCP)

Una policy che fornisce il controllo centralizzato sulle autorizzazioni per tutti gli account di un'organizzazione in AWS Organizations. Le SCP definiscono i guardrail o fissano i limiti alle azioni che un amministratore può delegare a utenti o ruoli. Puoi utilizzare le SCP come elenchi consentiti o elenchi di rifiuto, per specificare quali servizi o azioni sono consentiti o proibiti. Per ulteriori informazioni, consulta [Policy di sicurezza dei servizi](#) nella documentazione di AWS Organizations.

endpoint del servizio

L'URL del punto di accesso 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 con AWS per la sicurezza e la conformità del cloud. AWS è responsabile della sicurezza del cloud, mentre l'utente è 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.

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.

T

tags

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 è possibile utilizzare per collegare i VPC e le reti on-premise. Per ulteriori informazioni, consulta [Che cos'è un Transit Gateway?](#) nella documentazione di AWS Transit Gateway.

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

La concessione di autorizzazioni a un servizio specificato dall'utente per eseguire attività all'interno dell'organizzazione in AWS Organizations e nei relativi 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 di AWS Organizations con altri servizi AWS](#) nella documentazione di AWS Organizations.

regolazione

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

team da due pizze

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

U

incertezza

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

compiti indifferenziati

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

ambienti superiori

[Vedi ambiente.](#)

V

vacuum

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

controllo delle versioni

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

Peering VPC

Una connessione tra due VPC che consente di instradare il traffico tramite 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 interrogazioni 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.

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.