Applicazione del AWS Well-Architected Framework per Amazon Neptune Analytics

AWS Guida prescrittiva

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Guida prescrittiva: Applicazione del AWS Well-Architected Framework per Amazon Neptune Analytics

Copyright © 2025 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
- Destinatari principali ... 2
- Obiettivi ... 2

Il pilastro dell'eccellenza operativa ... 3
- Automatizza l'implementazione utilizzando un approccio IaC ... 3
- Progettazione delle operazioni ... 4
- Apporta modifiche frequenti, piccole e reversibili ... 5
- Implementa l'osservabilità per ottenere informazioni fruibili ... 5
- Imparate da tutti i fallimenti operativi ... 6
- Utilizza le funzionalità di registrazione per monitorare attività non autorizzate o anomale ... 6

Pilastro della sicurezza ... 8
- Implementare la sicurezza dei ... 8
- Proteggi le tue reti ... 9
- Implementa l'autenticazione e l'autorizzazione ... 9

Pilastro dell'affidabilità ... 11
- Comprendi le quote di servizio di Neptune ... 11
- Comprendi i modelli di implementazione di Neptune ... 12
- Gestisci e ridimensiona i cluster Neptune ... 13
- Gestisci i backup e gli eventi di failover ... 14

Pilastro dell'efficienza delle prestazioni ... 15
- Comprendi la modellazione grafica per l'analisi ... 15
- Ottimizza le interrogazioni ... 17
- Ottimizza le scritture ... 18
- Grafici di dimensioni corrette ... 19

Pilastro dell'ottimizzazione dei costi ... 21
- Comprendi i modelli di utilizzo e i servizi necessari ... 21
- Seleziona le risorse prestando attenzione ai costi ... 23

Pilastro della sostenibilità ... 25
- Considerate la vostra selezione Regione AWS ... 25
- Ottimizza i consumi ... 25
- Ottimizza lo sviluppo del software e i modelli di architettura ... 26

Risorse ... 28
- Riferimenti ... 28
- Post e video del blog ... 28

Addestramento 28
Cronologia dei documenti 29
Glossario 30
..... 30
A 31
B 34
C 36
D 39
E 43
F 45
G 47
H 48
I 50
L 52
M 53
O 58
P 60
Q 63
R 64
S 67
T 71
U 72
V 73
W 73
Z 75
..... lxxvi

Applicazione del AWS Well-Architected Framework per Amazon Neptune Analytics

Michael Havey, Amazon Web Services ()AWS

Dicembre 2024 (cronologia dei documenti)

Puoi creare soluzioni basate su grafici su Amazon Web Services (AWS) utilizzando Amazon Neptune. Neptune include Neptune Analytics, un motore di analisi dei grafici ottimizzato per la memoria che può analizzare rapidamente grandi quantità di dati grafici per ottenere informazioni e trovare tendenze. Può eseguire analisi sui dati nel cluster di database Neptune esistente oppure caricare e analizzare dati da set di dati esterni. Questa guida fornisce una guida prescrittiva per applicare i principi del AWS Well-Architected Framework quando si pianifica l'implementazione di Neptune Analytics. L'applicazione del AWS Well-Architected Framework per Amazon Neptune tratta lo stesso argomento per un database Neptune.

AWS Well-Architected Framework ti aiuta a creare infrastrutture sicure, ad alte prestazioni, resilienti ed efficienti per una varietà di applicazioni e carichi di lavoro. Fornisce inoltre un approccio coerente per valutare le architetture e implementare progetti scalabili.

Il AWS Well-Architected Framework si basa sui seguenti sei pilastri:

- Eccellenza operativa
- Sicurezza
- Affidabilità
- Efficienza delle prestazioni
- Ottimizzazione dei costi
- Sostenibilità

Questa guida fornisce informazioni sui pilastri di progettazione e sulle migliori pratiche di progettazione di Well-Architected Framework e considerazioni da tenere a mente quando si distribuisce Neptune Analytics su. AWS

Destinatari principali

Questa guida è destinata ai data engineer, agli architetti di soluzioni e agli analisti di dati che progettano e implementano soluzioni che utilizzano grafici. AWS

Obiettivi

Questa guida può aiutare te e la tua organizzazione a fare quanto segue:

- Scegli tra le opzioni di distribuzione supportate.
- Segui i modelli di progettazione AWS Well-Architected che ti aiutano a migliorare la resilienza e la sicurezza.
- Progetta le tue query per prestazioni ottimali e risparmi sui costi.
- Scoprite come essere efficienti dal punto di vista operativo nella gestione del grafico di Neptune Analytics in produzione.

Il pilastro dell'eccellenza operativa

Il pilastro dell'eccellenza operativa del AWS Well-Architected Framework si concentra sull'esecuzione e il monitoraggio dei sistemi e sul miglioramento continuo di processi e procedure. Include la capacità di supportare lo sviluppo ed eseguire i carichi di lavoro in modo efficace, ottenere informazioni dettagliate sul loro funzionamento e migliorare continuamente i processi e le procedure di supporto per offrire valore aziendale. È possibile ridurre la complessità operativa attraverso carichi di lavoro con riparazione automatica, che rilevano e risolvono la maggior parte dei problemi senza l'intervento umano. Puoi raggiungere questo obiettivo seguendo le best practice descritte in questa sezione e utilizzare i parametri APIs e i meccanismi di Amazon Neptune Analytics per rispondere correttamente quando il carico di lavoro si discosta dal comportamento previsto.

Questa discussione sul pilastro dell'eccellenza operativa si concentra sulle seguenti aree chiave:

- Infrastructure as code (IaC)
- Gestione delle modifiche
- Strategie di resilienza
- Gestione degli incidenti
- Segnalazione di audit per la conformità
- Registrazione di log e monitoraggio

Automatizza l'implementazione utilizzando un approccio IaC

Le migliori pratiche per automatizzare l'implementazione su Neptune utilizzando IaC includono quanto segue:

- Applica IaC per distribuire i grafici di Neptune Analytics e le risorse correlate. Per una configurazione coerente dell'ambiente, utilizza il supporto per Neptune Analytics fornito AWS CloudFormationda per fornire grafici ed endpoint privati.
- Utilizzalo CloudFormation per effettuare il provisioning delle istanze di notebook Neptune su Amazon AI. SageMaker Puoi usare i taccuini per interrogare e visualizzare i dati in un grafico di Neptune Analytics.
- Quando crei un grafico di Neptune Analytics da una fonte esistente, ad esempio un cluster o uno snapshot di database Neptune, o file di dati archiviati in Amazon Simple Storage Service (Amazon S3), monitora l'attività di importazione in blocco.

- Automatizza le procedure operative di Neptune Analytics, come il ridimensionamento del grafico, l'eliminazione e la creazione di un'istantanea del grafico, il ripristino del grafico da un'istantanea e il ripristino e il ricaricamento del grafico. Utilizza l'API Neptune Analytics, disponibile tramite AWS CLI() o AWS Command Line Interface SDK s.
- Valuta il tempo di attività richiesto per il tuo grafico. L'analisi è spesso effimera; il grafico è necessario solo per il tempo necessario per eseguire gli algoritmi. In tal caso, utilizzate AWS CLI o SDKs per creare un'istantanea ed eliminare il grafico quando non è più necessario. È quindi possibile ripristinarlo da un'istantanea in un secondo momento, se necessario.
- Archivia le stringhe di connessione esternamente dal client. Puoi archiviare le stringhe di connessione in AWS Secrets ManagerAmazon DynamoDB o in qualsiasi posizione in cui possano essere modificate dinamicamente.
- Usa i tag per aggiungere metadati alle tue risorse di Neptune Analytics e monitora l'utilizzo in base ai tag. I tag aiutano a organizzare le risorse. Ad esempio, puoi applicare un tag comune alle risorse in un ambiente o un'applicazione specifici. Puoi anche utilizzare i tag per analizzare la fatturazione dell'utilizzo delle risorse; per ulteriori informazioni, consulta Organizzazione e monitoraggio dei costi utilizzando AWS i tag di allocazione dei costi nella AWS Billing User Guide. Inoltre, puoi utilizzare le condizioni delle tue policy AWS Identity and Access Management (IAM) per controllare l'accesso alle AWS risorse in base ai tag utilizzati su quella risorsa. Puoi farlo utilizzando la chiave di condizione `aws:ResourceTag/tag-key` globale. Per ulteriori informazioni, consulta Controlling access to AWS resources nella IAM User Guide.

Progettazione delle operazioni

Adotta approcci per migliorare il modo in cui utilizzi i grafici di Neptune Analytics:

- Mantieni grafici di Neptune Analytics separati per lo sviluppo, il test e l'uso in produzione. Questi grafici potrebbero avere set di dati, utenti e controlli operativi diversi.
- Mantieni grafici di Neptune Analytics separati per usi diversi. Ad esempio, se due gruppi di utenti analitici richiedono grafici separati con tempistiche, modelli, prestazioni e disponibilità e modelli di utilizzo diversi SLAs, mantieni grafici separati per ogni gruppo.
- Prepara gli utenti e il personale operativo per gli aggiornamenti di manutenzione di Neptune Analytics.

Apporta modifiche frequenti, piccole e reversibili

I seguenti consigli si concentrano su piccole modifiche reversibili che è possibile apportare per ridurre al minimo la complessità e ridurre la probabilità di interruzione del carico di lavoro:

- Archivia modelli e script IAc in un servizio di controllo del codice sorgente come o. GitHub GitLab

> **Important**
>
> Non memorizzate AWS le credenziali nel controllo del codice sorgente.

- Richiedi che le implementazioni IaC utilizzino un servizio di integrazione e distribuzione continua (CI/CD) come o. AWS CodeDeployAWS CodeBuild Compila, testa e distribuisci il codice in un ambiente Neptune Analytics non di produzione prima di trasformarlo in un grafico di produzione.

Implementa l'osservabilità per ottenere informazioni fruibili

Ottieni una comprensione completa del comportamento, delle prestazioni, dell'affidabilità, dei costi e dello stato del carico di lavoro. I seguenti consigli ti aiutano ad acquisire quel livello di comprensione in Neptune Analytics:

- Monitora i CloudWatch parametri di Amazon per Neptune Analytics. In base a questi parametri, puoi determinare la dimensione di un grafico (numero di nodi, bordi e vettori, più la dimensione totale dei byte), l'utilizzo della CPU e le richieste di query e i tassi di errore.
- Crea CloudWatch dashboard e allarmi per metriche chiave come`NumQueuedRequestsPerSec`,, `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, e `CPUUtilization` le risposte dei client Neptune presenti nei registri delle applicazioni.
- Imposta notifiche per monitorare lo stato del grafico di Neptune Analytics, ad esempio quando la dimensione del grafico, la frequenza delle richieste o l'utilizzo della CPU superano la soglia. Ad esempio, se su un grafico `GraphStorageUsagePercent` è salito al 90 percento e intendi crescere in modo significativo, decidi se aumentare la capacità della Neptune Capacity Unit (m-NCU) ottimizzata per la memoria. Se l'attuale m-NCU è 128, aumentandolo a 256 si ridurrà lo storage di circa il 45 percento. Se `NumQueuedRequestsPerSec` è spesso maggiore di zero, valuta la possibilità di aumentare la capacità m-NCU per fornire una maggiore capacità di elaborazione. In alternativa, è possibile ridurre la concorrenza lato client.

Imparate da tutti i fallimenti operativi

Un'infrastruttura che si ripara automaticamente è uno sforzo a lungo termine che si sviluppa in iterazioni man mano che si verificano problemi rari o le risposte non sono così efficaci come desiderato. L'adozione delle seguenti pratiche favorisce l'attenzione verso tale obiettivo:

- Promuovi il miglioramento imparando da tutti i fallimenti.
- Condividi ciò che viene appreso tra i team e l'organizzazione. Se più team all'interno della tua organizzazione utilizzano Neptune, crea una chat room o un gruppo di utenti comune per condividere conoscenze e best practice.

Utilizza le funzionalità di registrazione per monitorare attività non autorizzate o anomale

Utilizza la registrazione per osservare prestazioni e modelli di attività anomali. Considerate le seguenti best practice:

- Neptune Analytics supporta la registrazione delle azioni del piano di controllo utilizzando. AWS CloudTrail Per ulteriori informazioni, consulta Registrazione delle chiamate API di Neptune Analytics utilizzando. AWS CloudTrail Grazie a questa funzionalità, puoi monitorare la creazione, l'aggiornamento e l'eliminazione delle risorse di Neptune Analytics. Per un monitoraggio e un sistema di avvisi affidabili, puoi anche integrare CloudTrail gli eventi con Amazon CloudWatch Logs. Per migliorare l'analisi dell'attività del servizio Neptune Analytics e identificare i cambiamenti nelle attività di Account AWS un utente, puoi CloudTrail interrogare i log utilizzando Amazon Athena. Ad esempio, è possibile utilizzare le query per identificare le tendenze e isolare con maggiore precisione le attività in base ad attributi specifici, ad esempio l'indirizzo IP di origine o un utente.
- È inoltre possibile utilizzarlo CloudTrail per abilitare la registrazione delle attività del piano dati di Neptune Analytics come le esecuzioni di query. È possibile visualizzare quali query vengono eseguite, la loro frequenza e la loro origine. Per impostazione predefinita, CloudTrail non registra gli eventi relativi ai dati. Per gli eventi di dati sono previsti costi aggiuntivi. Per ulteriori informazioni, consulta Prezzi di AWS CloudTrail.
- È inoltre possibile registrare le chiamate alle applicazioni a Neptune Analytics nel piano di controllo o nel piano dati. Ad esempio, se si utilizza il per AWS SDK per Python (Boto3)effettuare interrogazioni, è possibile abilitare la registrazione a livello di debug per ottenere una traccia delle

query sulla console o sul file. Ciò è utile durante lo sviluppo. Ti consigliamo inoltre di catturare e registrare le eccezioni dalla tua applicazione.

Pilastro della sicurezza

La sicurezza del cloud è la massima priorità in. AWS In qualità di AWS cliente, puoi beneficiare di un data center e di un'architettura di rete progettati per soddisfare i requisiti delle organizzazioni più sensibili alla sicurezza. La sicurezza è una responsabilità condivisa tra te e. AWS Il modello di responsabilità condivisa descrive questo come sicurezza del cloud e sicurezza nel cloud:

- Sicurezza del cloud: AWS è responsabile della protezione dell'infrastruttura che gira Servizi AWS su Cloud AWS. AWS fornisce inoltre servizi che è possibile utilizzare in modo sicuro. I revisori esterni testano e verificano regolarmente l'efficacia della AWS sicurezza nell'ambito dei programmi di AWS conformità. Per ulteriori informazioni sui programmi di conformità che si applicano a Neptune, AWS vedere Services in Scope by Compliance Program.
- Sicurezza nel cloud: la tua responsabilità è determinata dall'uso Servizio AWS che utilizzi. L'utente è anche responsabile di altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda e le leggi e le normative applicabili. Per ulteriori informazioni sulla privacy dei dati, consulta la sezione Privacy dei dati FAQs. Per informazioni sulla protezione dei dati in Europa, consulta il modello di responsabilitàAWS condivisa e il post sul blog sul GDPR.

Il pilastro della sicurezza di AWS Well-Architected Framework ti aiuta a capire come applicare il modello di responsabilità condivisa quando usi Neptune Analytics. I seguenti argomenti spiegano come configurare Neptune Analytics per soddisfare gli obiettivi di sicurezza e conformità. Scopri anche come utilizzarne altri Servizi AWS che ti aiutano a monitorare e proteggere le tue risorse di Neptune Analytics. Il pilastro della sicurezza include le seguenti aree di interesse chiave:

- Sicurezza dei dati
- Sicurezza di rete
- Autenticazione e autorizzazione

Implementare la sicurezza dei

La fuga e le violazioni dei dati mettono a rischio i tuoi clienti e possono avere un impatto negativo sostanziale sulla tua azienda. Le seguenti best practice aiutano a proteggere i dati dei clienti dall'esposizione involontaria e dolosa:

- I nomi dei grafici, i tag, i ruoli IAM e altri metadati non devono contenere informazioni riservate o sensibili, poiché tali dati potrebbero apparire nei registri di fatturazione o diagnostica.
- URIs o i collegamenti a server esterni archiviati come dati in Neptune non devono contenere informazioni sulle credenziali per convalidare le richieste.
- Un grafico di Neptune Analytics è crittografato quando è inattivo. Puoi utilizzare la chiave predefinita o una chiave AWS Key Management Service (AWS KMS) a tua scelta per crittografare il grafico. Puoi anche crittografare istantanee e dati esportati in Amazon S3 durante l'importazione in blocco. Puoi rimuovere la crittografia quando l'importazione è completa.
- Quando utilizzate il linguaggio OpenCypher, utilizzate le tecniche di parametrizzazione e convalida degli input appropriate per prevenire l'iniezione di SQL e altre forme di attacchi. Evita di creare query che utilizzano la concatenazione di stringhe con input forniti dall'utente. Utilizzate interrogazioni con parametri o istruzioni preparate per passare in modo sicuro i parametri di input al database grafico. Per ulteriori informazioni, consulta Esempi di query con parametri OpenCypher nella documentazione di Neptune.

Proteggi le tue reti

Puoi abilitare un grafico di Neptune Analytics per la connettività pubblica in modo che possa essere raggiunto dall'esterno di un cloud privato virtuale (VPC). Questa connettività è disabilitata per impostazione predefinita. Il grafico richiede l'autenticazione IAM. Il chiamante deve ottenere un'identità e disporre delle autorizzazioni per utilizzare il grafico. Ad esempio, per eseguire una query OpenCypher, il chiamante deve disporre delle autorizzazioni di lettura, scrittura o eliminazione sul grafico specifico.

Puoi anche creare endpoint privati per il grafico per accedere al grafico dall'interno di un VPC. Quando si crea l'endpoint, si specificano il VPC, le sottoreti e i gruppi di sicurezza per limitare l'accesso alla chiamata al grafico.

Per proteggere i dati in transito, Neptune Analytics applica le connessioni SSL al grafico tramite HTTPS. Per ulteriori informazioni, consulta Protezione dei dati in Neptune Analytics nella documentazione di Neptune Analytics.

Implementa l'autenticazione e l'autorizzazione

Le chiamate a un grafico di Neptune Analytics richiedono l'autenticazione IAM. Il chiamante deve ottenere un'identità e disporre di autorizzazioni sufficienti per eseguire l'azione sul grafico. Per le

descrizioni delle azioni API e delle relative autorizzazioni richieste, consulta la documentazione dell'API Neptune Analytics. Puoi applicare controlli delle condizioni per limitare l'accesso per tag.

L'autenticazione IAM utilizza il protocollo AWS Signature Version 4 (SigV4). Per semplificare l'utilizzo della tua applicazione, ti consigliamo di utilizzare un AWS SDK. Ad esempio, in Python, usa il client Boto3 per Neptune Graph, che astrae SigV4.

Quando carichi i dati nel grafico, il caricamento in batch utilizza le credenziali IAM del chiamante. Il chiamante deve disporre delle autorizzazioni per scaricare dati da Amazon S3 con la relazione di fiducia impostata in modo che Neptune Analytics possa assumere il ruolo di caricare i dati nel grafico dai file Amazon S3.

L'importazione in blocco può essere eseguita durante la creazione del grafico (dal team dell'infrastruttura) o su un grafico vuoto esistente (dal team di progettazione dei dati che dispone delle autorizzazioni per avviare le attività di importazione). In entrambi i casi, Neptune Analytics assume il ruolo IAM fornito dal chiamante come input. Questo ruolo gli consente di leggere ed elencare il contenuto della cartella Amazon S3 in cui sono conservati i dati di input.

Pilastro dell'affidabilità

Il pilastro dell'affidabilità di AWS Well-Architected Framework comprende la capacità di un carico di lavoro di svolgere la funzione prevista in modo corretto e coerente quando previsto. Ciò include la capacità di gestire e testare il carico di lavoro durante l'intero ciclo di vita.

Un carico di lavoro affidabile comincia con decisioni iniziali di progettazione sia per il software sia per l'infrastruttura. Le tue scelte di architettura influiscono sul comportamento del carico di lavoro in tutti i pilastri di Well-Architected. Per quanto riguarda l'affidabilità, è necessario seguire degli schemi specifici, come illustrato in questa sezione.

Il pilastro dell'affidabilità si concentra sulle seguenti aree chiave:

- Architettura del carico di lavoro, incluse quote di servizio e modelli di implementazione
- Gestione delle modifiche
- Gestione dei guasti

Comprendi le quote di servizio di Neptune

Il tuo AWS account ha delle quote predefinite (precedentemente denominate limiti) per ciascuna di esse. Servizio AWS Salvo diversa indicazione, ogni quota si applica a una regione specifica. Puoi richiedere aumenti per alcune quote, ma non per tutte.

Per trovare le quote per Neptune Analytics, apri la console Service Quotas. Nel riquadro di navigazione, scegli Servizi AWS, quindi seleziona Amazon Neptune Analytics. Presta attenzione alle quote relative al numero di grafici e istantanee, alla quantità massima di memoria fornita per un grafico e alle frequenze di richiesta API.

Se la memoria massima fornita non è sufficiente per il set di dati, valuta quali tipi di nodi e edge sono essenziali per l'utilizzo analitico previsto. Carica un sottoinsieme di dati in modo che le analisi siano possibili entro una capacità prevista consentita. Molti carichi di lavoro di analisi, in particolare quelli che eseguono algoritmi grafici, richiedono solo la topologia con un set limitato di proprietà anziché il grafico transazionale completo. (Per una discussione sulle differenze tra i carichi di lavoro transazionali e analitici, consulta la sezione relativa al pilastro dell'efficienza delle prestazioni).

Se il numero massimo di grafici non è sufficiente per l'uso previsto:

- Prendi in considerazione la possibilità di combinare grafici con usi simili.
- Valuta quanti grafici devono essere eseguiti in un determinato momento. Se hai un caso d'uso temporaneo di analisi, crea un'istantanea ed elimina un grafico quando non è più necessario. Ciò riduce il numero di grafici rispetto alla quota.
- Prendi in considerazione la possibilità di fornire grafici in diversi modi. Account AWS

Comprendi i modelli di implementazione di Neptune

Comprendi i seguenti punti decisionali quando prevedi di implementare un grafico di Neptune Analytics:

- Seeding: decidi se creare un grafico vuoto o caricare dati al suo interno al momento della creazione con dati provenienti da Amazon S3, un cluster di database Neptune esistente o uno snapshot del database Neptune esistente.

  Raccomandazione: se l'origine è un cluster o un'istantanea di Neptune, è necessario caricarne i dati al momento della creazione del grafico. Se l'origine è Amazon S3, carica i dati al momento della creazione se lo sforzo di caricamento è significativo e viene eseguito al meglio come attività di provisioning dell'infrastruttura. Se preferisci caricare i dati come attività di ingegneria dei dati o applicativa, crea un grafico vuoto e carica i dati da Amazon S3 in un secondo momento.
- Capacità: stima la capacità fornita richiesta per un grafico, in base alla dimensione dei dati e all'utilizzo previsto dell'applicazione.

  Raccomandazione: al momento della creazione, specificate la quantità massima di memoria disponibile per limitare la dimensione del grafico. Questa impostazione è obbligatoria. È possibile modificare la capacità in un secondo momento, se necessario.
- Disponibilità e tolleranza agli errori: decidi se le repliche sono necessarie per la disponibilità. Una replica funge da standby a caldo per il ripristino in caso di errore del grafico. Un grafico con repliche viene ripristinato più rapidamente di un grafico senza repliche. Considerate anche per quanto tempo è necessario il grafico, se è destinato esclusivamente all'analisi effimera e, in caso affermativo, quando verrà rimosso.

  Consiglio: prima di creare un grafico, stabilisci i requisiti di disponibilità, ad esempio per quanto tempo il grafico può non essere disponibile e quando può essere rimosso.
- Rete e sicurezza: stabilite se avete bisogno di connettività pubblica, privata o entrambe e se desiderate crittografare i dati.

Consiglio: prima di creare un grafico, comprendi i requisiti organizzativi, ad esempio se è consentita la connettività pubblica e dove verranno distribuite le applicazioni client Graph.

- Backup e ripristino: stabilisci se creare le istantanee e, in caso affermativo, quando o in quali condizioni. Valuta se la tua organizzazione ha requisiti di disaster recovery (DR).

  Raccomandazione: la creazione di istantanee è un'attività manuale. Decidi quando creare le istantanee e considera i requisiti di DR prima di creare un grafico.

Gestisci e ridimensiona i cluster Neptune

Un grafico di Neptune Analytics è costituito da un'unica istanza ottimizzata per la memoria. La capacità (m-NCU) dell'istanza viene impostata al momento della creazione. L'istanza può essere scalata verticalmente aumentando la capacità fornita tramite un'azione amministrativa; la capacità fornita può anche essere ridotta. Le repliche sono obiettivi di failover passivi, quindi non aumentano la scala di un grafico. A questo proposito, una replica grafica differisce da una replica di lettura del database Neptune, che è un'istanza attiva in un cluster Neptune in grado di elaborare le operazioni di lettura dalle applicazioni.

Le repliche comportano dei costi. Il prezzo della replica è pari alla tariffa m-NCU del grafico. Ad esempio, se un grafico è predisposto per 128 m-NCU e ha una sola replica, il costo è il doppio di quello di un grafo equivalente senza repliche.

Nell'analisi, ci sono due ragioni principali per aumentare la scalabilità:

- Per fornire più memoria e CPU per le query e gli algoritmi analitici, poiché la singola query è costosa, l'algoritmo grafico da eseguire è intrinsecamente complesso e richiede più risorse in base all'input, oppure la frequenza di richieste simultanee è elevata. Se tali query presentano out-of-memory errori, la scalabilità è un rimedio ragionevole.
- Per supportare un grafico di dimensioni maggiori di quelle previste. Ad esempio, se la capacità attualmente fornita è di 128 m-NCU per supportare 60 GB di dati di origine e sono necessari altri 40 GB di dati di origine, è garantito un aumento a 256 m-NCU.

Monitora le CloudWatch metriche per Neptune Analytics, `NumQueuedRequestsPerSec` come,,, `CPUUtilization` e `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, per determinare se è necessario il ridimensionamento. Puoi aggiornare la configurazione di un grafico tramite la console, oppure.

AWS CLI SDKs (Per esempi e best practice, consulta la sezione relativa al pilastro dell'eccellenza operativa).

Gestisci i backup e gli eventi di failover

Utilizza le repliche per garantire che un grafico rimanga disponibile in caso di errore. Un grafico utilizza la persistenza basata su log per confermare le modifiche tra le zone di disponibilità in un. Regione AWS La replica funge da standby a caldo e ha accesso a questi dati. In caso di errore, il grafico riprende le operazioni sulla replica. L'applicazione continua a utilizzare lo stesso endpoint per connettersi al grafico. Le richieste in corso durante l'errore generano errori con un'eccezione del servizio non disponibile. Prendi in considerazione l'utilizzo di un pattern retry with backoff nel codice dell'applicazione per catturare l'errore e riprova dopo un breve intervallo. Le nuove richieste effettuate durante il failover vengono messe in coda e potrebbero presentare una latenza maggiore.

Se non è configurata alcuna replica e il grafico fallisce, Neptune Analytics esegue il ripristino da uno storage durevole, ma il ripristino richiede più tempo perché Neptune deve reinizializzare le risorse.

Crea istantanee del grafico. (Neptune Analytics non scatta istantanee automatiche). Se il grafico viene modificato regolarmente dopo la creazione, scatta istantanee frequenti per acquisire lo stato corrente. Eliminate le vecchie istantanee se non è necessario il ripristino a un punto temporale precedente.

Puoi condividere le istantanee con altri account e tra di loro. Regioni AWS Se hai requisiti di disaster recovery, valuta se il ripristino del grafico in un'altra regione da un'istantanea soddisfa i requisiti del Recovery Time Objective (RTO) e del Recovery Point Objective (RPO).

Pilastro dell'efficienza delle prestazioni

Il pilastro dell'efficienza delle prestazioni del AWS Well-Architected Framework si concentra su come ottimizzare le prestazioni durante l'acquisizione o l'interrogazione dei dati. L'ottimizzazione delle prestazioni è un processo incrementale e continuo che prevede quanto segue:

- Conferma dei requisiti aziendali
- Misurazione delle prestazioni del carico di lavoro
- Identificazione dei componenti poco performanti
- Ottimizzazione dei componenti per soddisfare le esigenze aziendali

Il pilastro relativo all'efficienza delle prestazioni fornisce linee guida specifiche per ogni caso d'uso che possono aiutarti a identificare il modello di dati grafico e i linguaggi di interrogazione corretti da utilizzare. Include anche le migliori pratiche da seguire per l'inserimento e l'utilizzo di dati da Neptune Analytics.

Il pilastro dell'efficienza delle prestazioni si concentra sulle seguenti aree chiave:

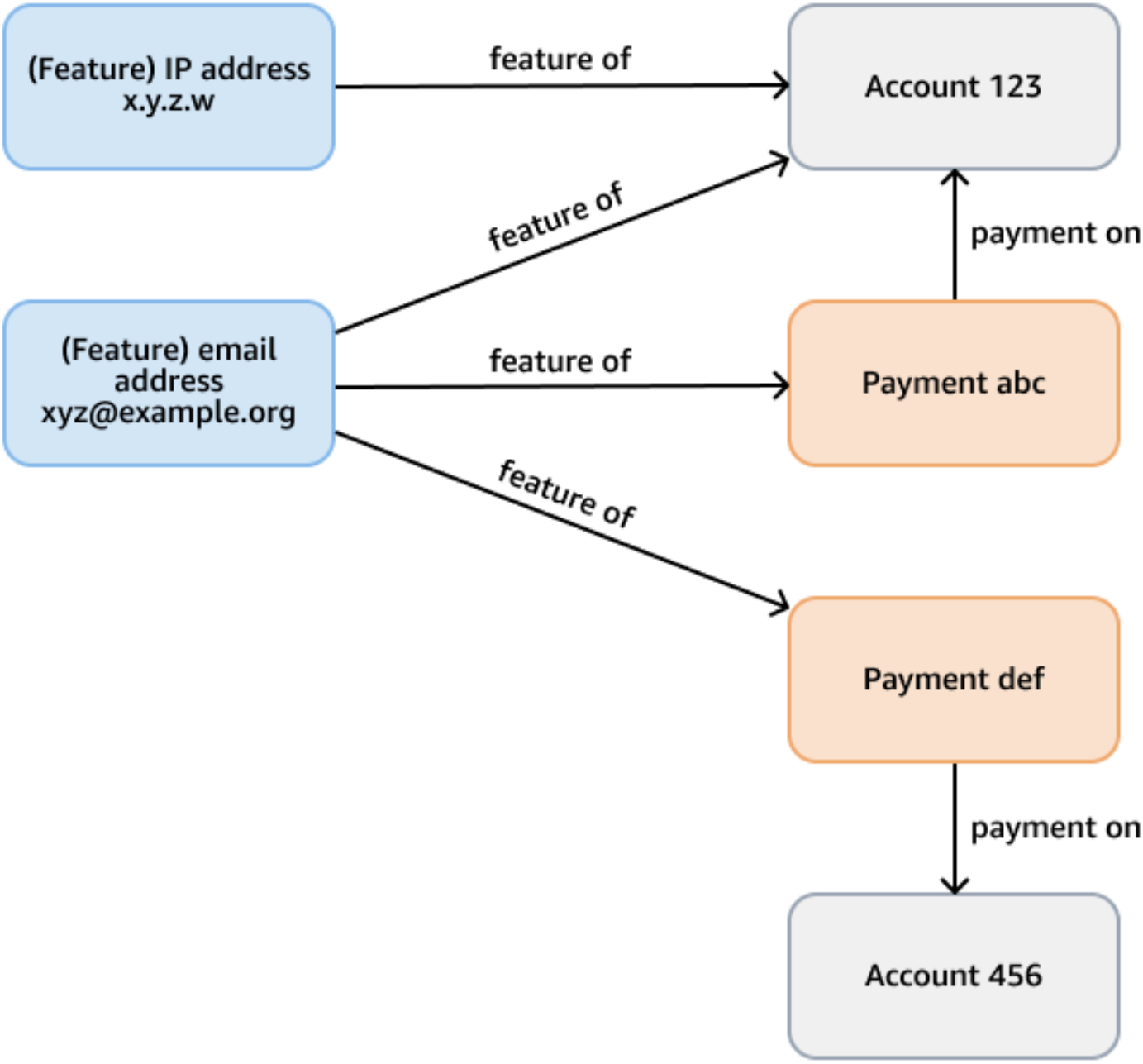
- Modellazione grafica
- Ottimizzazione delle query
- Ridimensionamento corretto del grafico
- Ottimizzazione della scrittura

Comprendi la modellazione grafica per l'analisi

La guida Appliing the AWS Well-Architected Framework for Amazon Neptune illustra la modellazione grafica per l'efficienza delle prestazioni. Le decisioni di modellazione che influiscono sulle prestazioni includono la scelta dei nodi e degli spigoli necessari, delle relative etichette e proprietà IDs, della direzione degli spigoli, se le etichette devono essere generiche o specifiche e, in generale, dell'efficienza con cui il motore di query può navigare nel grafico per elaborare le query più comuni.

Queste considerazioni si applicano anche a Neptune Analytics; tuttavia, è importante distinguere tra modelli di utilizzo transazionali e analitici. Potrebbe essere necessario rimodellare un modello grafico efficiente per le query in un database transazionale come un database Neptune per l'analisi.

Ad esempio, si consideri un grafico delle frodi in un database di Neptune il cui scopo è verificare la presenza di schemi fraudolenti nei pagamenti con carta di credito. Questo grafico potrebbe contenere nodi che rappresentano account, pagamenti e funzionalità (come indirizzo e-mail, indirizzo IP, numero di telefono) sia dell'account che del pagamento. Questo grafico connesso supporta interrogazioni come l'attraversamento di un percorso a lunghezza variabile che parte da un determinato pagamento e richiede diversi passaggi per trovare funzionalità e account correlati. La figura seguente mostra un grafico di questo tipo.



Il requisito analitico potrebbe essere più specifico, ad esempio trovare comunità di account collegate da una funzionalità. A tale scopo è possibile utilizzare l'algoritmo WCC (Weakly Connected Components). Eseguirlo sul modello dell'esempio precedente è inefficiente, perché deve attraversare diversi tipi di nodi e spigoli. Il modello nel diagramma seguente è più efficiente. Collega `account` i nodi a un `shares feature` margine se gli account stessi, o i pagamenti dagli account, condividono una funzionalità. Ad esempio, `Account 123` dispone della funzionalità di posta elettronica e `Account 456` utilizza `xyz@example.org` la stessa e-mail per un pagamento (). `Payment def`

La complessità computazionale del WCC è`O(|E|logD)`: `|E|` dov'è il numero di spigoli nel grafico e qual `D` è il diametro (la lunghezza del percorso più lungo) che collega i nodi. Poiché il modello transazionale omette nodi e spigoli non essenziali, ottimizza sia il numero di spigoli che il diametro e riduce la complessità dell'algoritmo WCC.

Quando usi Neptune Analytics, recupera gli algoritmi e le query analitiche richiesti. Se necessario, rimodella il modello per ottimizzare queste query. È possibile rimodellare il modello prima di caricare i dati nel grafico o scrivere query che modificano i dati esistenti nel grafico.

Ottimizza le interrogazioni

Segui questi consigli per ottimizzare le query di Neptune Analytics:

- Utilizza le interrogazioni parametrizzate e la cache del piano di interrogazione, che è abilitata di default. Quando si utilizza la cache del piano, il motore prepara la query per un uso successivo, a condizione che la query venga completata in 100 millisecondi o meno, il che consente di risparmiare tempo nelle chiamate successive.
- Per le query lente, esegui un piano esplicativo per individuare le strozzature e apportare i miglioramenti di conseguenza.
- Se utilizzi la ricerca per somiglianza vettoriale, decidi se incorporamenti più piccoli producono risultati di somiglianza accurati. Puoi creare, archiviare e cercare incorporamenti più piccoli in modo più efficiente.

- Segui le best practice documentate per l'utilizzo di OpenCypher in Neptune Analytics. Ad esempio, utilizza mappe appiattite in una clausola UNWIND e specifica le etichette dei bordi ove possibile.
- Quando utilizzate un algoritmo grafico, comprendete gli input e gli output dell'algoritmo, la sua complessità computazionale e, in generale, come funziona.
  - Prima di richiamare un algoritmo grafico, utilizzate una `MATCH` clausola per ridurre al minimo il set di nodi di input. Ad esempio, per limitare i nodi da cui eseguire la ricerca breadth-first (BFS), segui gli esempi forniti nella documentazione di Neptune Analytics.
  - Se possibile, filtra sulle etichette dei nodi e dei bordi. Ad esempio, BFS dispone di parametri di input per filtrare l'attraversamento verso una specifica etichetta di nodo (`vertexLabel`) o etichette di bordo specifiche (). `edgeLabels`
  - Utilizzate parametri di delimitazione, ad esempio per limitare i risultati`maxDepth`.
  - Sperimentate con il `concurrency` parametro. Provalo con il valore 0, che utilizza tutti i thread dell'algoritmo disponibili per parallelizzare l'elaborazione. Confrontalo con l'esecuzione a thread singolo impostando il parametro su 1. Un algoritmo può completare più velocemente in un singolo thread, specialmente su input più piccoli, come le ricerche breadth-first poco profonde, in cui il parallelismo non offre una riduzione misurabile dei tempi di esecuzione e potrebbe comportare un sovraccarico.
  - Scegli tra tipi di algoritmi simili. Ad esempio, Bellman-Ford e delta-stepping sono entrambi algoritmi con percorso più breve a fonte singola. Quando esegui il test con il tuo set di dati, prova entrambi gli algoritmi e confronta i risultati. Il delta-stepping è spesso più veloce di Bellman-Ford a causa della sua minore complessità computazionale. Tuttavia, le prestazioni dipendono dal set di dati e dai parametri di input, in particolare dal parametro. `delta`

Ottimizza le scritture

Segui queste pratiche per ottimizzare le operazioni di scrittura in Neptune Analytics:

- Cerca il modo più efficiente per caricare i dati in un grafico. Quando carichi dati in Amazon S3, utilizza l'importazione in blocco se le dimensioni dei dati superano i 50 GB. Per dati più piccoli, usa il caricamento in batch. Se out-of-memory riscontri errori durante l'esecuzione del caricamento in batch, valuta la possibilità di aumentare il valore m-NCU o di suddividere il carico in più richieste. Un modo per farlo consiste nel dividere i file tra più prefissi nel bucket S3. In tal caso, chiamate batch load separatamente per ogni prefisso.

- Utilizzate l'importazione in blocco o il caricatore batch per popolare il set iniziale di dati del grafico. Usa le operazioni transazionali di creazione, aggiornamento ed eliminazione di OpenCypher solo per piccole modifiche.
- Utilizzate l'importazione in blocco o il caricatore batch con una concorrenza di 1 (thread singolo) per inserire gli incorporamenti nel grafico. Prova a caricare gli incorporamenti in anticipo utilizzando uno di questi metodi.
- Valuta la dimensione degli incorporamenti vettoriali necessari per una ricerca di similarità accurata negli algoritmi di ricerca per somiglianza vettoriale. Se possibile, usa una dimensione più piccola. Ciò si traduce in una maggiore velocità di caricamento per gli incorporamenti.
- Se necessario, utilizzate algoritmi di mutazione per ricordare i risultati algoritmici. Ad esempio, l'algoritmo degree mutate centrality trova il grado di ogni nodo di input e scrive quel valore come proprietà del nodo. Se le connessioni che circondano quei nodi non cambiano successivamente, la proprietà mantiene il risultato corretto. Non è necessario eseguire nuovamente l'algoritmo.
- Usa l'azione amministrativa di reimpostazione del grafico per cancellare tutti i nodi, gli edge e gli incorporamenti se devi ricominciare da capo. Eliminare tutti i nodi, gli spigoli e gli incorporamenti utilizzando una query OpenCypher non è possibile se il grafico è di grandi dimensioni. Una singola query drop su un set di dati di grandi dimensioni può scadere. All'aumentare delle dimensioni, la rimozione del set di dati richiede più tempo e la dimensione della transazione aumenta. Al contrario, il tempo necessario per completare la reimpostazione del grafico è all'incirca costante e l'azione offre la possibilità di creare un'istantanea prima di eseguirla.

Grafici di dimensioni corrette

Le prestazioni complessive dipendono dalla capacità fornita di un grafico di Neptune Analytics. La capacità viene misurata in unità denominate Neptune Capacity Units (m-) ottimizzate per la memoria. NCUs Assicurati che le dimensioni del grafico siano sufficienti per supportare le dimensioni del grafico e le query. Tieni presente che una maggiore capacità non migliora necessariamente le prestazioni di una singola query.

Se possibile, crea il grafico importando dati da una fonte esistente come Amazon S3 o un cluster o uno snapshot Neptune esistente. Puoi porre limiti alla capacità minima e massima. È inoltre possibile modificare la capacità assegnata su un grafico esistente.

Monitora CloudWatch metriche come`NumQueuedRequestsPerSec`,, `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, e `CPUUtilization` valuta se il grafico è delle dimensioni corrette. Determina se è necessaria una

maggiore capacità per supportare le dimensioni e il carico del grafico. Per ulteriori informazioni su come interpretare alcune di queste metriche, consulta la sezione relativa al pilastro dell'eccellenza operativa.

Pilastro dell'ottimizzazione dei costi

Il pilastro dell'ottimizzazione dei costi del AWS Well-Architected Framework si concentra sull'evitare costi inutili. I seguenti consigli possono aiutarti a soddisfare i principi di progettazione per l'ottimizzazione dei costi e le migliori pratiche architettoniche per Neptune Analytics.

Il pilastro dell'ottimizzazione dei costi si concentra sulle seguenti aree chiave:

- Comprendere la spesa nel tempo e controllare l'allocazione dei fondi
- Selezione delle risorse del tipo e della quantità corretti
- Scalabilità per soddisfare le esigenze aziendali senza spese eccessive

Comprendi i modelli di utilizzo e i servizi necessari

Prima di adottare Neptune Analytics, valuta se il tuo caso d'uso è adatto all'analisi dei grafici.

- Database a grafo: un database a grafo come Neptune è ideale per il tuo carico di lavoro se il tuo modello di dati ha una struttura a grafi riconoscibile e le tue query devono esplorare le relazioni e attraversare più hop. Un database grafico non è adatto ai seguenti modelli:
  - Principalmente interrogazioni a hop singolo. In questo caso d'uso, valuta se i tuoi dati potrebbero essere meglio rappresentati come attributi di un oggetto.
  - Dati JSON o Binary Large Object (blob) archiviati come proprietà.

- Analisi dei grafici: Neptune Analytics è un motore di database di analisi dei grafici in grado di analizzare rapidamente grandi quantità di dati grafici in memoria per ottenere informazioni e trovare tendenze. È possibile archiviare e interrogare i dati del grafico sia in un database Neptune che in un grafico di Neptune Analytics. Un database Neptune è più adatto per le esigenze di elaborazione transazionale online scalabile (OLTP). Neptune Analytics è ideale per carichi di lavoro di analisi effimeri. Puoi utilizzare i due in combinazione caricando i dati dal tuo database Neptune orientato alle transazioni su un grafico di Neptune Analytics per eseguire l'analisi di tali dati. Una volta completata l'analisi, puoi rimuovere il grafico di Neptune Analytics. Per un confronto più dettagliato, consulta Quando usare Neptune Analytics e quando usare Neptune Database nella documentazione di Neptune Analytics.

Determina, con attenzione ai costi, il modo migliore per compilare il tuo grafico di Neptune Analytics.

- Importa in blocco i dati grafici archiviati in un bucket S3. Consigliamo questa opzione se i dati sono stati precedentemente caricati in blocco su un database Neptune o se disponi già, o puoi produrre prontamente, i dati da analizzare in CSV o in altri formati supportati richiesti dall'importazione in blocco. È possibile eseguire l'importazione in blocco come parte della procedura di creazione del grafico. È possibile porre limiti alla capacità minima e massima. È inoltre possibile eseguire l'importazione su un grafico vuoto creato in precedenza e monitorare l'attività di importazione durante l'esecuzione.
- È possibile creare un grafico vuoto e quindi popolarlo tramite una query OpenCypher utilizzando il caricamento in batch. Questa opzione è ideale se i dati da caricare sono archiviati in Amazon S3 e sono inferiori a 50 GB.
- È possibile popolare il grafico con i dati del cluster di database Neptune (supportato nella versione 1.3.0 o successiva del database Neptune). L'intento di questo modello è eseguire analisi sui dati attualmente presenti nel database grafico. Anche se il database è stato inizialmente popolato tramite caricamento in blocco, da allora potrebbe essere cambiato in modo significativo. Per importare dal database, Neptune Analytics clona il database ed esporta i dati dal clone in un bucket S3. Questa procedura comporta dei costi: in particolare i costi del database Neptune per l'esecuzione del clone e i costi di Amazon S3 per l'archiviazione e il consumo dei dati esportati. Il clone viene rimosso quando l'esportazione è completa. Puoi eliminare i dati esportati in Amazon S3.
- È possibile popolare il grafico dall'istantanea di un cluster di database Neptune. Questa opzione è simile all'opzione precedente, tranne per il fatto che l'origine è un'istantanea del database. Per importare da un'istantanea, Neptune Analytics ripristina prima l'istantanea in un nuovo cluster di database, quindi esporta i dati in un bucket S3. Questa procedura comporta dei costi: in particolare i costi del database Neptune per l'esecuzione del cluster ripristinato e i costi di Amazon S3 per l'archiviazione e il consumo dei dati esportati.
- Puoi anche eseguire query OpenCypher per creare, aggiornare o eliminare dati utilizzando transazioni conformi ad atomicità, coerenza, isolamento, durabilità (ACID) sul grafico. Consigliamo questo approccio come metodo per apportare piccoli aggiornamenti ma non come modo per seminare il grafico.

Se i dati necessari per l'analisi sono già archiviati in Amazon S3, consigliamo l'importazione in blocco o il caricamento in batch. Sono più convenienti rispetto alla compilazione del grafico da un cluster di database Neptune o da un'istantanea.

Seleziona le risorse prestando attenzione ai costi

I prezzi di Neptune Analytics utilizzano un'unità nota come Neptune Capacity Unit (m-NCU) ottimizzata per la memoria. Esiste un costo orario fisso per l'esecuzione di un grafico con un determinato m-NCU. Un grafico può avere delle repliche per il failover e queste repliche comportano anche un costo orario m-NCU.

Consigliamo le seguenti best practice per stimare la capacità, limitare i costi e monitorare i costi rispetto alle prestazioni:

- Se possibile, crea il grafico importando dati da una fonte esistente: dati archiviati in Amazon S3 o un cluster o uno snapshot Neptune esistente. Ciò consente di risparmiare fatica perché Neptune Analytics esegue il lavoro pesante di seminare il grafico ed è possibile specificare una capacità massima limitata.
- È possibile modificare la capacità assegnata su un grafico esistente.
- Quando il grafico non è più necessario, è possibile creare un'istantanea ed eliminare il grafico. Se è necessario riutilizzarlo, è possibile ripristinare il grafico dall'istantanea.
- È possibile scegliere il numero di repliche quando si crea il grafico. Imposta il valore in base ai tuoi requisiti di disponibilità dell'analisi. Risparmia sui costi riducendo al minimo questa impostazione. Il valore massimo di 2 consente la replica di due istanze in zone di disponibilità separate. Il valore minimo 0 indica che Neptune Analytics non eseguirà una replica. Tuttavia, il ripristino è più rapido quando è disponibile una replica. Per una spiegazione dell'errore e del ripristino del grafico, consulta la sezione Reliability Pillar.
- Monitora le spese di Neptune Analytics per i periodi di fatturazione attuali e passati utilizzando. AWS Billing and Cost Management
- Monitora le metriche di Neptune Analytics, `NumQueuedRequestsPerSec` in particolare`NumOpenCypherRequestsPerSec`, `GraphStorageUsagePercentGraphSizeBytes`, `CPUUtlization` e CloudWatch per valutare se la capacità fornita è di dimensioni adeguate per il grafico. Determina se una capacità inferiore è in grado di soddisfare la frequenza di richiesta, l'utilizzo della CPU e le dimensioni del grafico osservati.
- Se hai bisogno di un endpoint privato per il tuo grafico, presta attenzione ai costi degli endpoint elastici del cloud privato virtuale (VPC) IPs, dei gateway NAT o di altri costi relativi al VPC. Per ulteriori informazioni, consulta i prezzi di Amazon VPC e i prezzi di Amazon EC2.

- Potresti voler eseguire una o più istanze di notebook Neptune per fornire un'interfaccia client che aiuti sviluppatori e analisti a interrogare e visualizzare il grafico (vedi i prezzi di Neptune workbench). Per ridurre al minimo i costi, condividi l'istanza tra gli utenti e crea cartelle di notebook separate per ogni utente. Chiudi l'istanza quando non è in uso. Per un approccio all'automazione dello spegnimento, consulta il post AWS sul blog Automatizza l'arresto e l'avvio delle risorse dell'ambiente Amazon Neptune utilizzando i tag di risorsa.

Pilastro della sostenibilità

Il pilastro della sostenibilità del AWS Well-Architected Framework si concentra sulla riduzione al minimo degli impatti ambientali dell'esecuzione di carichi di lavoro cloud. Gli argomenti chiave includono un modello di responsabilità condivisa per la sostenibilità, la comprensione dell'impatto e la massimizzazione dell'uso per ridurre al minimo le risorse necessarie e ridurre gli impatti a valle.

Il pilastro della sostenibilità contiene le seguenti aree di interesse chiave:

- Il tuo impatto
- Obiettivi di sostenibilità
- Utilizzo massimizzato
- Anticipazione e adozione di nuove offerte software più efficienti
- Utilizzo di servizi gestiti
- Riduzione dell'impatto a valle

Questa guida si concentra sulla comprensione del tuo impatto. Per ulteriori informazioni sugli altri principi di progettazione della sostenibilità, vedere il AWS Well-Architected Framework.

Le tue scelte e i tuoi requisiti hanno un impatto sull'ambiente. Se è possibile scegliere soluzioni Regioni AWS con un'intensità di carbonio inferiore e se i requisiti riflettono le esigenze effettive del carico di lavoro anziché limitarsi a massimizzare l'operatività e la durata, la sostenibilità del carico di lavoro aumenta. Nelle sezioni successive vengono illustrate le best practice e le considerazioni che avranno un impatto ambientale positivo se adottate nella progettazione del carico di lavoro e nelle operazioni in corso

Considerate la vostra selezione Regione AWS

Alcuni Regioni AWS sono vicini a progetti di energia rinnovabile di Amazon o si trovano dove la rete ha un'intensità di carbonio pubblicata inferiore rispetto ad altre. Considera l'impatto sulla sostenibilità per le regioni che potrebbe essere fattibile per il tuo carico di lavoro e confronta il tuo elenco con le regioni in cui è disponibile Neptune Analytics.

Ottimizza i consumi

Riduci al minimo il consumo di Neptune Analytics praticando quanto segue:

- L'analisi è spesso effimera. Il grafico è necessario solo per il tempo necessario per eseguire gli algoritmi e registrare i risultati. In tal caso, crea un'istantanea ed elimina il grafico quando non è più necessario. È possibile ripristinarlo da un'istantanea in un secondo momento, se necessario.
- Se il carico di lavoro è effimero e hai la flessibilità necessaria per decidere quando eseguire le analisi, prendi in considerazione day-to-day le tendenze del consumo energetico. La domanda di elettricità è più elevata in determinati periodi. Se ti trovi negli Stati Uniti, consulta le metriche sul consumo giornaliero di elettricità sul sito web della U.S. Energy Information Administration (EIA). Se possibile, esegui i carichi di lavoro durante i periodi non di punta per la tua regione.
- Se il carico di lavoro non è effimero ma deve essere disponibile solo per periodi limitati, elimina il grafico e ripristinalo da un'istantanea quando necessario. Se la sua disponibilità segue una pianificazione, automatizza il processo di ripristino tramite script in modo che il grafico sia pronto all'ora pianificata.
- Se i dati sono di sola lettura o non sono stati modificati dall'ultima istantanea, non copiateli nuovamente prima dell'eliminazione.
- Ferma i notebook Neptune quando non sono in uso.
- Monitora CloudWatch metriche come`NumQueuedRequestsPerSec`,, `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, e valuta se il grafico è `CPUUtilization` sovradimensionato. Determina se una capacità dell'istanza inferiore è in grado di soddisfare la frequenza di richiesta, l'utilizzo della CPU e le dimensioni del grafico osservati.

Ottimizza lo sviluppo del software e i modelli di architettura

Per prevenire gli sprechi, ottimizza i modelli e le query e condividi le risorse di calcolo in modo da utilizzare tutte le risorse disponibili nelle istanze e nei cluster Neptune. Le migliori pratiche specifiche includono:

- Ottimizza le interrogazioni e le invocazioni degli algoritmi grafici. Utilizza le query parametrizzate e la cache del piano di query, che è abilitata per impostazione predefinita. Per le query lente, esegui un piano esplicativo per apportare miglioramenti. Se utilizzi la ricerca per somiglianza vettoriale, decidi se incorporamenti più piccoli producono risultati di somiglianza accurati, poiché gli incorporamenti più piccoli possono essere creati, archiviati e ricercati in modo più efficiente. Prima di richiamare un algoritmo grafico, utilizzate una `MATCH` clausola per ridurre al minimo il set di nodi di input. Se possibile, filtrate sulle etichette dei nodi e dei bordi.

- Cerca il modo più efficiente per caricare i dati nel grafico. Se carichi da dati in Amazon S3, utilizza l'importazione in blocco se le dimensioni dei dati superano i 50 GB. Usa il caricamento in batch per dati più piccoli.
- Chiedi agli sviluppatori di condividere le istanze dei notebook Neptune invece di creare ciascuno la propria istanza. Crea cartelle di notebook separate per ogni sviluppatore su una singola istanza di Jupyter. Chiudi l'istanza quando non è in uso.

Risorse

Riferimenti

- AWS Well-Architected
- AWS Documentazione Well-Architected Framework
- Applicazione del AWS Well-Architected Framework per Amazon Neptune
- Le migliori pratiche di Neptune Analytics

Post e video del blog

- Post AWS sul blog di Neptune (Database Blog)
- Automatizza l'arresto e l'avvio delle risorse dell'ambiente Amazon Neptune utilizzando i tag di risorsa (Database Blog)AWS
- Prodotti snack Amazon Neptune (brevi video) YouTube

Addestramento

- Guida introduttiva ad Amazon Neptune
- Crea con Amazon Neptune
- Modellazione dei dati per Amazon Neptune
- Workshop sull'analisi di Amazon Neptune

Cronologia dei documenti

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

Modifica	Descrizione	Data
Pubblicazione iniziale	—	20 dicembre 2024

AWS Glossario delle linee guida prescrittive

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

Numeri

7 R

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

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

A

ABAC

Vedi controllo degli accessi basato sugli attributi.

servizi astratti

Vedi servizi gestiti.

ACIDO

Vedi atomicità, consistenza, isolamento, durata.

migrazione attiva-attiva

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

migrazione attiva-passiva

Un metodo di migrazione 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 AIOps viene utilizzato nella strategia di AWS migrazione, consulta la guida all'integrazione delle operazioni.

crittografia asimmetrica

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

atomicità, consistenza, isolamento, durabilità (ACID)

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

Controllo degli accessi basato su attributi (ABAC)

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

fonte di dati autorevole

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

Zona di disponibilità

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

AWS Cloud Adoption Framework (CAF)AWS

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

AWS Workload Qualification Framework (WQF)AWS

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

B

bot difettoso

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

BCP

Vedi la pianificazione della continuità operativa.

grafico comportamentale

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

sistema big-endian

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

Classificazione binaria

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

filtro Bloom

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

distribuzioni blu/verdi

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

bot

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

botnet

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

ramo

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

accesso break-glass

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

strategia brownfield

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

cache del buffer

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

capacità di business

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

pianificazione della continuità operativa (BCP)

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

C

CAF

Vedi AWS Cloud Adoption Framework.

implementazione canaria

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

CCoE

Vedi Cloud Center of Excellence.

CDC

Vedi Change Data Capture.

Change Data Capture (CDC)

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

ingegneria del caos

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

CI/CD

Vedi integrazione continua e distribuzione continua.

classificazione

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

crittografia lato client

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

Centro di eccellenza cloud (CCoE)

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

cloud computing

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

modello operativo cloud

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

fasi di adozione del cloud

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

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

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

CMDB

Vedi database di gestione della configurazione.

repository di codice

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

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

cache fredda

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

dati freddi

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

visione artificiale (CV)

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

deriva della configurazione

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

database di gestione della configurazione (CMDB)

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

Pacchetto di conformità

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

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

Il processo di automazione delle fasi di origine, compilazione, test, gestione temporanea e produzione del processo di rilascio del software. CI/CD is commonly described as a pipeline. CI/

CDpuò aiutarvi ad automatizzare i processi, migliorare la produttività, migliorare la qualità del codice e velocizzare le consegne. Per ulteriori informazioni, consulta Vantaggi della distribuzione continua. CD può anche significare continuous deployment (implementazione continua). Per ulteriori informazioni, consulta Distribuzione continua e implementazione continua a confronto.

CV

Vedi visione artificiale.

D

dati a riposo

Dati stazionari nella rete, ad esempio i dati archiviati.

classificazione dei dati

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

deriva dei dati

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

dati in transito

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

rete di dati

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

riduzione al minimo dei dati

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

perimetro dei dati

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

pre-elaborazione dei dati

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

provenienza dei dati

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

soggetto dei dati

Un individuo i cui dati vengono raccolti ed elaborati.

data warehouse

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

linguaggio di definizione del database (DDL)

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

linguaggio di manipolazione del database (DML)

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

DDL

Vedi linguaggio di definizione del database.

deep ensemble

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

deep learning

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

defense-in-depth

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

amministratore delegato

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

implementazione

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

Ambiente di sviluppo

Vedi ambiente.

controllo di rilevamento

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

mappatura del flusso di valore dello sviluppo (DVSM)

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

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

gemello digitale

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

tabella delle dimensioni

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

disastro

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

disaster recovery (DR)

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

DML

Vedi linguaggio di manipolazione del database.

progettazione basata sul dominio

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

DOTT.

Vedi disaster recovery.

rilevamento della deriva

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

DVSM

Vedi la mappatura del flusso di valore dello sviluppo.

E

EDA

Vedi analisi esplorativa dei dati.

MODIFICA

Vedi scambio elettronico di dati.

edge computing

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

scambio elettronico di dati (EDI)

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

crittografia

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

chiave crittografica

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

endianità

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

endpoint

Vedi service endpoint.

servizio endpoint

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

pianificazione delle risorse aziendali (ERP)

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

crittografia envelope

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

ambiente

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

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

- ambiente di produzione: un'istanza di un'applicazione in esecuzione a cui gli utenti finali possono accedere. In una 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 epopee della sicurezza AWS CAF includono la gestione delle identità e degli accessi, i controlli investigativi, la sicurezza dell'infrastruttura, la protezione dei dati e la risposta agli incidenti. Per ulteriori informazioni sulle epiche, consulta la strategia di migrazione AWS , consulta la guida all'implementazione del programma.

ERP

Vedi pianificazione delle risorse aziendali.

analisi esplorativa dei dati (EDA)

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

F

tabella dei fatti

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

fallire velocemente

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

limite di isolamento dei guasti

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

ramo di funzionalità

Vedi filiale.

caratteristiche

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

importanza delle caratteristiche

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

trasformazione delle funzionalità

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

prompt con pochi scatti

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

FGAC

Vedi il controllo granulare degli accessi.

controllo granulare degli accessi (FGAC)

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

migrazione flash-cut

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

FM

Vedi il modello di base.

modello di fondazione (FM)

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

G

AI generativa

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

blocco geografico

Vedi restrizioni geografiche.

limitazioni geografiche (blocco geografico)

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

Flusso di lavoro di GitFlow

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

immagine dorata

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

strategia greenfield

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

guardrail

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

H

AH

Vedi disponibilità elevata.

migrazione di database eterogenea

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

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

alta disponibilità (HA)

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

modernizzazione storica

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

dati di esclusione

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

migrazione di database omogenea

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

dati caldi

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

hotfix

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

periodo di hypercare

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

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

I

IaC

Considera l'infrastruttura come codice.

Policy basata su identità

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

applicazione inattiva

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

IIoT

Vedi Industrial Internet of Things.

infrastruttura immutabile

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

VPC in ingresso (ingress)

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

migrazione incrementale

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

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

Industria 4.0

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

infrastruttura

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

infrastruttura come codice (IaC)

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

IIoInternet delle cose industriale (T)

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

VPC di ispezione

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

Internet of Things (IoT)

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

interpretabilità

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

IoT

Vedi Internet of Things.

libreria di informazioni IT (ITIL)

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

gestione dei servizi IT (ITSM)

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

ITIL

Vedi la libreria di informazioni IT.

ITSM

Vedi Gestione dei servizi IT.

L

controllo degli accessi basato su etichette (LBAC)

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

zona di destinazione

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

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

modello linguistico di grandi dimensioni (LLM)

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

migrazione su larga scala

Una migrazione di 300 o più server.

BIANCO

Vedi controllo degli accessi basato su etichette.

Privilegio minimo

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

eseguire il rehosting (lift and shift)

Vedi 7 R.

sistema little-endian

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

LLM

Vedi modello linguistico di grandi dimensioni.

ambienti inferiori

Vedi ambiente.

M

machine learning (ML)

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

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

ramo principale

Vedi filiale.

malware

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

servizi gestiti

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

sistema di esecuzione della produzione (MES)

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

MAP

Vedi Migration Acceleration Program.

meccanismo

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

account membro

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

MEH.

Vedi sistema di esecuzione della produzione.

Message Queuing Telemetry Transport (MQTT)

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

microservizio

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

architettura di microservizi

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

Programma di accelerazione della migrazione (MAP)

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

migrazione su larga scala

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

fabbrica di migrazione

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

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

metadati di migrazione

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

modello di migrazione

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

Valutazione del portfolio di migrazione (MPA)

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

valutazione della preparazione alla migrazione (MRA)

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

strategia di migrazione

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

ML

Vedi machine learning.

modernizzazione

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

valutazione della preparazione alla modernizzazione

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

applicazioni monolitiche (monoliti)

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

MAPPA

Vedi Migration Portfolio Assessment.

MQTT

Vedi Message Queuing Telemetry Transport.

classificazione multiclasse

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

infrastruttura mutabile

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

O

OAC

Vedi Origin Access Control.

QUERCIA

Vedi Origin Access Identity.

OCM

Vedi gestione delle modifiche organizzative.

migrazione offline

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

OI

Vedi l'integrazione delle operazioni.

OLA

Vedi accordo a livello operativo.

migrazione online

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

OPC-UA

Vedi Open Process Communications - Unified Architecture.

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

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

accordo a livello operativo (OLA)

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

revisione della prontezza operativa (ORR)

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

tecnologia operativa (OT)

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

integrazione delle operazioni (OI)

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

trail organizzativo

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

gestione del cambiamento organizzativo (OCM)

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

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

controllo dell'accesso all'origine (OAC)

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

identità di accesso origine (OAI)

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

ORR

Vedi la revisione della prontezza operativa.

- NON

Vedi la tecnologia operativa.

VPC in uscita (egress)

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

P

limite delle autorizzazioni

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

informazioni di identificazione personale (PII)

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

Informazioni che consentono l'identificazione personale degli utenti

Visualizza le informazioni di identificazione personale.

playbook

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

PLC

Vedi controllore logico programmabile.

PLM

Vedi la gestione del ciclo di vita del prodotto.

policy

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

persistenza poliglotta

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

valutazione del portfolio

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

predicate

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

predicato pushdown

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

controllo preventivo

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

principale

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

privacy fin dalla progettazione

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

zone ospitate private

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

controllo proattivo

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

gestione del ciclo di vita del prodotto (PLM)

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

Ambiente di produzione

Vedi ambiente.

controllore logico programmabile (PLC)

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

concatenamento rapido

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

pseudonimizzazione

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

publish/subscribe (pub/sub)

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

Q

Piano di query

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

regressione del piano di query

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

R

Matrice RACI

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

STRACCIO

Vedi Retrieval Augmented Generation.

ransomware

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

Matrice RASCI

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

RCAC

Vedi controllo dell'accesso a righe e colonne.

replica di lettura

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

riprogettare

Vedi 7 Rs.

obiettivo del punto di ripristino (RPO)

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

obiettivo del tempo di ripristino (RTO)

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

rifattorizzare

Vedi 7 R.

Regione

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

regressione

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

riospitare

Vedi 7 R.

rilascio

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

trasferisco

Vedi 7 Rs.

ripiattaforma

Vedi 7 Rs.

riacquisto

Vedi 7 Rs.

resilienza

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

policy basata su risorse

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

matrice di assegnazione di responsabilità (RACI)

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

controllo reattivo

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

retain

Vedi 7 R.

andare in pensione

Vedi 7 Rs.

Retrieval Augmented Generation (RAG)

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

rotazione

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

controllo dell'accesso a righe e colonne (RCAC)

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

RPO

Vedi l'obiettivo del punto di ripristino.

RTO

Vedi l'obiettivo del tempo di ripristino.

runbook

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

S

SAML 2.0

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

SCADA

Vedi controllo di supervisione e acquisizione dati.

SCP

Vedi la politica di controllo del servizio.

Secret

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

sicurezza fin dalla progettazione

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

controllo di sicurezza

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

rafforzamento della sicurezza

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

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

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

automazione della risposta alla sicurezza

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

Crittografia lato server

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

Policy di controllo dei servizi (SCP)

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

endpoint del servizio

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

accordo sul livello di servizio (SLA)

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

indicatore del livello di servizio (SLI)

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

obiettivo a livello di servizio (SLO)

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

Modello di responsabilità condivisa

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

SIEM

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

punto di errore singolo (SPOF)

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

SLAM

Vedi il contratto sul livello di servizio.

SLI

Vedi l'indicatore del livello di servizio.

LENTA

Vedi obiettivo del livello di servizio.

split-and-seed modello

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

SPOF

Vedi punto di errore singolo.

schema a stella

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

modello del fico strangolatore

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

sottorete

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

controllo di supervisione e acquisizione dati (SCADA)

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

crittografia simmetrica

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

test sintetici

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

prompt di sistema

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

T

tags

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

variabile di destinazione

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

elenco di attività

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

Ambiente di test

Vedi ambiente.

training

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

Transit Gateway

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

flusso di lavoro basato su trunk

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

Accesso attendibile

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

regolazione

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

team da due pizze

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

U

incertezza

Un concetto che si riferisce a informazioni imprecise, incomplete o sconosciute che possono minare l'affidabilità dei modelli di machine learning predittivi. Esistono due tipi di incertezza: l'incertezza epistemica, che è causata da dati limitati e incompleti, mentre l'incertezza aleatoria è causata dal rumore e dalla casualità insiti nei dati. 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 VPCs che consente di indirizzare il traffico utilizzando indirizzi IP privati. Per ulteriori informazioni, consulta Che cos'è il peering VPC? nella documentazione di Amazon VPC.

vulnerabilità

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

W

cache calda

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

dati caldi

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

funzione finestra

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

Carico di lavoro

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

flusso di lavoro

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

VERME

Vedi scrivere una volta, leggere molti.

WQF

Vedi AWS Workload Qualification Framework.

scrivi una volta, leggi molte (WORM)

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

Z

exploit zero-day

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

vulnerabilità zero-day

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

prompt zero-shot

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

applicazione zombie

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

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